

Institute of Mathematics and Computer Sciences

ISSN - 0103-2569

Benchmarking Text Collections for Classification and Clustering Tasks

Rafael Geraldeli Rossi
Ricardo Marcondes Marcacini
Solange Oliveira Rezende

Nº 395

ICMC TECHNICAL REPORT

São Carlos, SP, Brazil
November/2013

Benchmarking Text Collections for Classification and Clustering Tasks*

Rafael Geraldeli Rossi¹
Ricardo Marcondes Marcacini^{1,2}
Solange Oliveira Rezende¹

¹University of São Paulo
Institute of Mathematics and Computer Sciences
Laboratory of Computational Intelligence
São Carlos – SP – Brazil

²Federal University of Mato Grosso do Sul
Três Lagoas – MS – Brazil

November/2013

*Grant 2011/12823-6, 2011/19850-9, and 2010/20564-8, Sao Paulo Research Foundation (FAPESP).

Abstract

Several text mining techniques have been proposed to deal with the huge number of textual documents that are available and that have been published nowadays. Mainly classification techniques, which assign pre-defined labels to new documents, and clustering techniques, which separates texts into clusters. The techniques proposed in literature are usually applied to few textual collections, which are not sufficient to indicate how good a technique is or which characteristics of the collections make a technique obtain better results than others. Besides, new techniques are compared with traditional algorithms considering a small range of parameters, which make the comparison unfair. This technical report solve this lack by i) providing a characterization of 45 text collections; ii) providing classification results using traditional algorithms as Naïve Bayes, Multinomial Naïve Bayes, C4.5, k -Nearest Neighbors, and Support Vector Machines using a big range of values for the parameters; and iii) providing clustering results using traditional hierarchical algorithms as Bisecting k -Means and Unweighted Pair Group Method with Arithmetic Mean. We also make all the collections used in this technical report available in an on-line repository.

Contents

Contents	iii
List of Tables	x
List of Figures	xi
List of Abbreviations	xiii
1 Introduction	1
2 Characteristics of the Benchmarking Text Collections	3
2.1 20Ng Collection	5
2.2 ACM Collection	6
2.3 Classic4 Collection	8
2.4 CSTR Collection	9
2.5 Dmoz-Business-500 Collection	10
2.6 Dmoz-Computers-500 Collection	12
2.7 Dmoz-Health-500 Collection	13
2.8 Dmoz-Science-500 Collection	14
2.9 Dmoz-Sports-500 Collection	15
2.10 Enron-Top-20 Collection	16
2.11 FBIS Collection	18
2.12 Hitech Collection	19
2.13 Industry-Sector Collection	20
2.14 Irish-Sentiment Collection	21
2.15 La1s Collection	22
2.16 La2s Collection	23
2.17 LATimes Collection	24
2.18 Multi-Domain-Sentiment Collection	25
2.19 New3 Collection	26
2.20 NSF Collection	28
2.21 Oh0 Collection	29
2.22 Oh5 Collection	30
2.23 Oh10 Collection	31
2.24 Oh15 Collection	32
2.25 Ohscal Collection	33
2.26 Ohsumed-400 Collection	34
2.27 Opinions Collection	35
2.28 Pubmed-Cancer Collection	37
2.29 Re0 Collection	38
2.30 Re1 Collection	39
2.31 Re8 Collection	40
2.32 Review-Polarity Collection	41
2.33 Reviews Collection	42

2.34	SpamAssassin Collection	43
2.35	SpamTrec-3000 Collection	44
2.36	SyskillWebert Collection	45
2.37	Tr11 Collection	46
2.38	Tr12 Collection	47
2.39	Tr21 Collection	48
2.40	Tr23 Collection	49
2.41	Tr31 Collection	50
2.42	Tr41 Collection	51
2.43	Tr45 Collection	52
2.44	WAP Collection	53
2.45	WebKB Collection	54
2.46	Summary of the characteristics for all the collections	55
3	Classification Results	57
3.1	20ng Collection	60
3.2	ACM Collection	62
3.3	Classic4 Collection	64
3.4	CSTR Collection	66
3.5	Dmoz-Business-500 Collection	68
3.6	Dmoz-Computers-500 Collection	70
3.7	Dmoz-Health-500 Collection	72
3.8	Dmoz-Science-500 Collection	74
3.9	Dmoz-Sports-500 Collection	76
3.10	Enron-Top-20 Collection	78
3.11	FBIS Collection	80
3.12	Hitech Collection	82
3.13	Industry-Sector Collection	84
3.14	Irish-Sentiment Collection	86
3.15	Lals Collection	88
3.16	La2s Collection	90
3.17	LATimes Collection	92
3.18	Multi-Domain-Sentiment Collection	94
3.19	New3 Collection	96
3.20	NSF Collection	98
3.21	Oh0 Collection	100
3.22	Oh5 Collection	102
3.23	Oh10 Collection	104
3.24	Oh15 Collection	106
3.25	Ohscal Collection	108
3.26	Ohsumed-400 Collection	110
3.27	Opinosis Collection	112
3.28	Pubmed-Cancer Collection	114
3.29	Re0 Collection	116
3.30	Re1 Collection	118
3.31	Re8 Collection	120
3.32	Review-Polarity Collection	122
3.33	Reviews Collection	124
3.34	SpamAssassin Collection	126
3.35	SpamTrec-3000 Collection	128

3.36 SyskillWebert Collection	130
3.37 Tr11 Collection	132
3.38 Tr12 Collection	134
3.39 Tr21 Collection	136
3.40 Tr23 Collection	138
3.41 Tr31 Collection	140
3.42 Tr41 Collection	142
3.43 Tr45 Collection	144
3.44 WAP Collection	146
3.45 WebKb Collection	148
3.46 Synthesis of the Classification Results	150
4 Clustering Results	153
5 Final Considerations	157
References	161

List of Tables

2.1	20ng collection characteristics.	5
2.2	Number of documents and S-Index per class for 20ng collection.	5
2.3	ACM collection characteristics.	6
2.4	Number of documents per class for ACM collection.	7
2.5	Classic4 collection characteristics.	8
2.6	Number of documents and S-Index per class for Classic4 collection.	8
2.7	CSTR collection characteristics.	9
2.8	Number of documents and S-Index per class for CSTR collection.	9
2.9	Dmoz-Business-500 collection characteristics.	10
2.10	Number of documents and S-Index per class for Dmoz-Business-500 collection.	11
2.11	Dmoz-Computers-500 collection characteristics.	12
2.12	Number of documents and S-Index per class for Dmoz-Computers-500 collection.	12
2.13	Dmoz-Health-500 collection characteristics.	13
2.14	Number of documents and S-Index per class for Dmoz-Health-500 collection.	13
2.15	Dmoz-Science-500 collection characteristics.	14
2.16	Number of documents and S-Index per class for Dmoz-Science-500 collection.	14
2.17	Dmoz-Sports-500 collection characteristics.	15
2.18	Number of documents and S-Index per class for Dmoz-Sports-500 collection.	15
2.19	Enron-Top-20 collection characteristics.	16
2.20	Number of documents and S-Index per class for Enron-Top-20 collection.	17
2.21	FBIS collection characteristics.	18
2.22	Number of documents and S-Index per class for FBIS collection.	18
2.23	Hitech collection characteristics.	19
2.24	Number of documents and S-Index per class for Hitech collection.	19
2.25	Industry-Sector collection characteristics.	20
2.26	Number of documents and S-Index per class for Industry-Sector collection.	20
2.27	Irish-Sentiment collection characteristics.	21
2.28	Number of documents and S-Index per class for Irish-Sentiment collection.	21
2.29	La1s collection characteristics.	22
2.30	Number of documents and S-Index per class for La1s collection.	22
2.31	La2s collection characteristics.	23
2.32	Number of documents and S-Index per class for La2s collection.	23
2.33	LATimes collection characteristics.	24
2.34	Number of documents and S-Index per class for LATimes collection.	24
2.35	Multi-Domain-Sentiment collection characteristics.	25
2.36	Number of documents and S-Index per class for Multi-Domain-Sentiment collection.	25
2.37	New3 collection characteristics.	26

2.38	Number of documents and S-Index per class for New3 collection.	27
2.39	NSF collection characteristics.	28
2.40	Number of documents and S-Index per class for NSF collection.	28
2.41	Oh0 collection characteristics.	29
2.42	Number of documents and S-Index per class for Oh0 collection.	29
2.43	Oh5 collection characteristics.	30
2.44	Number of documents and S-Index per class for Oh5 collection.	30
2.45	Oh10 collection characteristics.	31
2.46	Number of documents and S-Index per class for Oh10 collection.	31
2.47	Oh15 collection characteristics.	32
2.48	Number of documents and S-Index per class for Oh15 collection.	32
2.49	Ohscal collection characteristics.	33
2.50	Number of documents and S-Index per class for Ohscal collection.	33
2.51	Ohsumed-400 collection characteristics.	34
2.52	Number of documents and S-Index per class for Ohsumed-400 collection.	34
2.53	Opinosis collection characteristics.	35
2.54	Number of documents and S-Index per class for Opinosis collection.	36
2.55	Pubmed-Cancer collection characteristics.	37
2.56	Number of documents and S-Index per class for Pubmed-Cancer collection.	37
2.57	Re0 collection characteristics.	38
2.58	Number of documents and S-Index per class for Re0 collection.	38
2.59	Re1 collection characteristics.	39
2.60	Number of documents and S-Index per class for Re1 collection.	39
2.61	Re8 collection characteristics.	40
2.62	Number of documents and S-Index per class for Re8 collection.	40
2.63	Review-Polarity collection characteristics.	41
2.64	Number of documents and S-Index per class for Review-Polarity collection.	41
2.65	Reviews collection characteristics.	42
2.66	Number of documents and S-Index per class for Reviews collection.	42
2.67	SpamAssassin collection characteristics.	43
2.68	Number of documents and S-Index per class for SpamAssassin collection.	43
2.69	SpamTrec-3000 collection characteristics.	44
2.70	Number of documents and S-Index per class for SpamTrec-3000 collection.	44
2.71	SyskillWebert collection characteristics.	45
2.72	Number of documents and S-Index per class for SyskillWebert collection.	45
2.73	Tr11 collection characteristics.	46
2.74	Number of documents and S-Index per class for Tr11 collection.	46
2.75	Tr12 collection characteristics.	47
2.76	Number of documents and S-Index per class for Tr12 collection.	47
2.77	Tr21 collection characteristics.	48
2.78	Number of documents and S-Index per class for Tr21 collection.	48
2.79	Tr23 collection characteristics.	49
2.80	Number of documents and S-Index per class for Tr23 collection.	49
2.81	Tr31 collection characteristics.	50
2.82	Number of documents and S-Index per class for Tr31 collection.	50
2.83	Tr41 collection characteristics.	51
2.84	Number of documents and S-Index per class for Tr41 collection.	51
2.85	Tr45 collection characteristics.	52
2.86	Number of documents and S-Index per class for Tr45 collection.	52
2.87	Wap collection characteristics.	53

2.88	Number of documents and S-Index per class for Wap collection.	53
2.89	WebKb collection characteristics.	54
2.90	Number of documents and S-Index per class for WebKb collection.	54
2.91	Characteristics of all the text collections analyzed in this technical report	55
3.1	Contingency table for category c_i [Sebastiani, 2002].	58
3.2	Global contingency table [Sebastiani, 2002].	58
3.3	Classification Results for 20ng collection.	60
3.4	Classification Results for 20ng collection.	61
3.5	Classification Results for ACM collection.	62
3.6	Classification Results for ACM collection.	63
3.7	Classification Results for Classic4 collection.	64
3.8	Classification Results for Classic4 collection.	65
3.9	Classification Results for CSTR collection.	66
3.10	Classification Results for CSTR collection.	67
3.11	Classification Results for Dmoz-Business-500 collection.	68
3.12	Classification Results for Dmoz-Business-500 collection.	69
3.13	Classification Results for Dmoz-Computers-500 collection.	70
3.14	Classification Results for Dmoz-Computers-500 collection.	71
3.15	Classification Results for Dmoz-Health-500 collection.	72
3.16	Classification Results for Dmoz-Health-500 collection.	73
3.17	Classification Results for Dmoz-Science-500 collection.	74
3.18	Classification Results for Dmoz-Science-500 collection.	75
3.19	Classification Results for Dmoz-Sports-500 collection.	76
3.20	Classification Results for Dmoz-Sports-500 collection.	77
3.21	Classification Results for Enron-Top-20 collection.	78
3.22	Classification Results for Enron-Top-20 collection.	79
3.23	Classification Results for FBIS collection.	80
3.24	Classification Results for FBIS collection.	81
3.25	Classification Results for Hitech collection.	82
3.26	Classification Results for Hitech collection.	83
3.27	Classification Results for Industry-Sector collection.	84
3.28	Classification Results for Industry-Sector collection.	85
3.29	Classification Results for Irish-Sentiment collection.	86
3.30	Classification Results for Irish-Sentiment collection.	87
3.31	Classification Results for La1s collection.	88
3.32	Classification Results for La1s collection.	89
3.33	Classification Results for La2s collection.	90
3.34	Classification Results for La2s collection.	91
3.35	Classification Results for LATimes collection.	92
3.36	Classification Results for LATimes collection.	93
3.37	Classification Results for Multi-Domain-Sentiment collection.	94
3.38	Classification Results for Multi-Domain-Sentiment collection.	95
3.39	Classification Results for New3 collection.	96
3.40	Classification Results for New3 collection.	97
3.41	Classification Results for NSF collection.	98
3.42	Classification Results for NSF collection.	99
3.43	Classification Results for Oh0 collection.	100
3.44	Classification Results for Oh0 collection.	101
3.45	Classification Results for Oh5 collection.	102

3.46	Classification Results for Oh5 collection.	103
3.47	Classification Results for Oh10 collection.	104
3.48	Classification Results for Oh10 collection.	105
3.49	Classification Results for Oh15 collection.	106
3.50	Classification Results for Oh15 collection.	107
3.51	Classification Results for Ohscal collection.	108
3.52	Classification Results for Ohscal collection.	109
3.53	Classification Results for Ohsumed-400 collection.	110
3.54	Classification Results for Ohsumed-400 collection.	111
3.55	Classification Results for Opinosis collection.	112
3.56	Classification Results for Opinosis collection.	113
3.57	Classification Results for Pubmed-Cancer collection.	114
3.58	Classification Results for Pubmed-Cancer collection.	115
3.59	Classification Results for Re0 collection.	116
3.60	Classification Results for Re0 collection.	117
3.61	Classification Results for Re1 collection.	118
3.62	Classification Results for Re1 collection.	119
3.63	Classification Results for Re8 collection.	120
3.64	Classification Results for Re8 collection.	121
3.65	Classification Results for Review-Polatility collection.	122
3.66	Classification Results for Review-Polarity collection.	123
3.67	Classification Results for Reviews collection.	124
3.68	Classification Results for Reviews collection.	125
3.69	Classification Results for SpamAssassin collection.	126
3.70	Classification Results for SpamAssassin collection.	127
3.71	Classification Results for SpamTrec-3000 collection.	128
3.72	Classification Results for SpamTrec-3000 collection.	129
3.73	Classification Results for SyskillWebert collection.	130
3.74	Classification Results for SyskillWebert collection.	131
3.75	Classification Results for Tr11 collection.	132
3.76	Classification Results for Tr11 collection.	133
3.77	Classification Results for Tr12 collection.	134
3.78	Classification Results for Tr12 collection.	135
3.79	Classification Results for Tr21 collection.	136
3.80	Classification Results for Tr21 collection.	137
3.81	Classification Results for Tr23 collection.	138
3.82	Classification Results for Tr23 collection.	139
3.83	Classification Results for Tr31 collection.	140
3.84	Classification Results for Tr31 collection.	141
3.85	Classification Results for Tr41 collection.	142
3.86	Classification Results for Tr41 collection.	143
3.87	Classification Results for Tr45 collection.	144
3.88	Classification Results for Tr45 collection.	145
3.89	Classification Results for WAP collection.	146
3.90	Classification Results for WAP collection.	147
3.91	Classification Results for WebKb collection.	148
3.92	Classification Results for WebKb collection.	149
3.93	Best accuracies for the classification algorithms used in this technical report.	151
4.1	F_{SCORE} for the clustering algorithms used in this technical report.	155

List of Figures

3.1	Average ranking diagrams for classification results.	152
4.1	Average ranking diagrams for clustering results.	156

List of Abbreviations

20Ng	<i>20 Newsgroups</i>
ACM	<i>Association for Computing Machinery</i>
CCR	<i>Correct Classification Rate</i>
CSTR	<i>Computer Science Technical Reports</i>
EM	<i>E-mail</i>
FBIS	<i>Foreign Broadcast Information Service</i>
FN	<i>False Negatives</i>
FP	<i>False Positives</i>
ICR	<i>Incorrect Classification Rate</i>
ICT	<i>Inductive Classification Tool</i>
kNN	<i>k Nearest Neighbors</i>
MD	<i>Medical documents</i>
MNB	<i>Multinomial Naïve Bayes</i>
NA	<i>News articles</i>
NB	<i>Naïve Bayes</i>
NSF	<i>National Science Foundation</i>
Pr	<i>Precision</i>
Re	<i>Recall</i>
SA	<i>Sentiment analysis</i>
SD	<i>Scientific documents</i>
SMO	<i>Sequential Minimal Optimization</i>
SVM	<i>Support Vector Machines</i>
TN	<i>True Negatives</i>
TP	<i>True Positives</i>
TPT	<i>Text Preprocessing Tool</i>
TREC	<i>Text REtrieval Conference</i>
WAP	<i>WebACE Project</i>
WP	<i>Web pages</i>
WV	<i>Weighted Vote</i>

Chapter 1

Introduction

Due to the huge number of textual documents that are available and that have been published nowadays, text mining techniques to manage, retrieve, and extract knowledge from these collections have become useful. Consequently, several techniques have been developed and applied to text collections.

The text mining techniques are usually evaluated using benchmarking collections. Although, for most of the the articles found in literature, the techniques are applied to few text collections. This kind of evaluation does not indicate how good a technique is compared to others, or even provide the characteristics of the collection that makes one technique obtain better results than other techniques. The main reason of this lack is the huge computational effort to build benchmark collections, characterize them, and run several algorithms with different parameter values to generate results for comparison.

In this technical report we present the characteristics of 45 text collections from different domains: e-mails, medical, news, sentiment analysis, scientific documents, and TREC (Text REtrieval Conference [TREC, 2013]) documents. We also run traditional/state-of-the art algorithms for classification (Naïve Bayes, Multinomial Naïve Bayes, C4.5, Support Vector Machines, and k Nearest Neighbors) and clustering (k -Means and Bisecting- k -Means), the main used text mining techniques [Weiss et al., 2010, Berry et al., 2008, Berry, 2004], to provide results for comparison with new/existent text mining techniques. We also make all the collections used in this technical report available in an on-line repository: http://sites.labic.icmc.usp.br/text_collections/.

The remainder of this technical report is organized as follows. Section 2 presents the characteristics of the benchmarking text collections. Section 3 presents the classification results and Section 4 presents the clustering results.

Chapter 2

Characteristics of the Benchmarking Text Collections

In this chapter we present the characteristics of the textual documents. To do so, the collections need to be represented in a structured format. The collections analyzed in this technical report were structured using the bag-of-words representation [Salton, 1989]. This representation is based on a vector space model, in which the documents are represented by vectors and each single word (feature/term) corresponds to a dimension of the vectors. To extract the terms we remove the stopwords¹ using the stoplist² of the Pre-text tool [Soares et al., 2008], and the remaining words were stemmed using the Porter's algorithm [Porter, 1980]. The bag-of-words representation were generate in `arff` format [Witten and Frank, 2005] using the Text Preprocessing Tool³ (TPT). The representations are available at http://sites.labic.icmc.usp.br/text_collections/.

For each generated bag-of-words representation we extracted the following characteristics:

- **# Documents:** number of documents contained in the collection.
- **Abs. # Docs.:** number of documents contained in each class of the collection.
- **Rel. # Docs.:** relative number of documents contained in each class of the collection.
- **# Terms:** number of term extracted after the preprocessing step.
- **$\overline{\#}$ Terms:** average occurrence of the terms in the documents.
- **Matrix Sparsity:** relative number of zeros in the bag-of-words representation.
- **# Classes:** number of classes in the text collection.

¹Non-important words as articles, prepositions and some verbs.

²List of stopwords

³TPT is available at <http://sites.labic.icmc.usp.br/tpt/>

- **Class S.D.:** standard deviation considering the percentage of documents that belong to each class in the collection
- **Majority Class:** percentage of documents that belong to the majority class.
- **S-Index:** is an adaptation of the silhouette coefficient [Kogan, 2007] which intend to measure the overlap among the classes. Let a_i be the average dissimilarity of a document i and all other documents of the same class. Let b_i be the average dissimilarity of a document i and all the documents of the closest class. Then, the silhouette coefficient for an document i is give by:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} . \quad (2.1)$$

In this technical report we consider the S-Index for the collection, i.e., the mean of the $s(i)$ for all documents of the collection, and the S-Index for each class, i.e, the mean of the $s(i)$ for documents of a class. The values of $s(i)$ ranges from $]1, 1]$. When the mean of $s(i)$ is close to 1 means that the documents from one class are well separated to the documents of the other classes, i.e, there is no overlap among the classes.

The next sections present the characteristics described above for the text collections analyzed in this technical report.

2.1 20Ng Collection

20Ng (20 Newsgroups) collection is composed by 20000 taken messages from 20 Usenet groups [Rennie, 2008]. Table 2.1 presents the characteristics and Table 2.2 presents the number of documents and the S-Index for each class of the 20ng collection.

Table 2.1: 20ng collection characteristics.

Domain	E-mails
# Documents	18808
# Terms	45434
# Terms	76.47
Matrix Sparsity	99.83%
# Classes	20
Class S.D.	0.52%
Majority Class	5.31%
S-Index	0.795

Table 2.2: Number of documents and S-Index per class for 20ng collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
talk_religion_misc	628	3.34%	0.540
talk_politics_misc	775	4.12%	0.808
alt_atheism	779	4.14%	0.738
talk_politic_guns	910	4.84%	0.869
talk_politics_mideast	940	5.00%	0.853
comp_sys_mac_hardware	961	5.11%	0.743
misc_forsale	972	5.17%	0.758
comp_graphics	973	5.17%	0.764
comp_windows_x	980	5.21%	0.681
sci_electronics	981	5.22%	0.784
comp_sys_ibm_pc_hardware	982	5.22%	0.657
comp_os_ms-windows_misc	985	5.24%	0.723
sci_space	987	5.25%	0.862
rec_autos	990	5.26%	0.786
sci_med	990	5.26%	0.900
sci_crypt	991	5.27%	0.803
rec_motorcycles	994	5.28%	0.874
rec_sport_baseball	994	5.28%	0.875
soc_religion_christian	997	5.30%	0.860
rec_sport_hockey	999	5.31%	0.914

2.2 ACM Collection

ACM (Association for Computing Machinery) collection are composed by proceedings from the ACM digital library⁴. The proceedings are from 40 conferences of different computer science areas. Table 2.3 presents the characteristics and Table 2.4 presents the number of documents and the S-Index for each class of the ACM collection.

Table 2.3: ACM collection characteristics.

Domain	Scientific
# Documents	3493
# Terms	60768
# Terms	720.30
Matrix Sparsity	98.81%
# Classes	40
Class S.D.	0.37%
Majority Class	3.03%
S-Index	0.752

⁴<http://dl.acm.org/>

Table 2.4: Number of documents per class for ACM collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
EmbeddedNetworked	50	1.43%	0.900
MolecularBiology	69	1.98%	0.928
IEEEVis	70	2.00%	0.614
ArchitectureEducation	71	2.03%	0.873
InformationRetrieval	71	2.03%	0.803
SystemSupportGames	71	2.03%	0.873
WebAccessibility	71	2.03%	0.901
SoftwareReusability	72	2.06%	0.875
PhysicalModeling	74	2.12%	0.892
CommunicationSystems	75	2.15%	0.907
EletronicSociety	80	2.29%	0.763
EmbeddedInteraction	81	2.32%	0.864
MobileMultimedia	82	2.35%	0.561
SoftwareEng	82	2.35%	0.768
VirtualReality	83	2.38%	0.759
SimulationConference	84	2.40%	0.821
RationalityKnowledge	85	2.43%	0.894
WebIntelligence	86	2.46%	0.605
TechnologyEducation	87	2.49%	0.851
ComputationalGeometry	89	2.55%	0.888
ModelsTechnologies	90	2.58%	0.933
3DTech	91	2.61%	0.780
AdHoc	91	2.61%	0.736
SoftwarePerformance	92	2.63%	0.783
HypertextHypermedia	93	2.66%	0.645
SystemDesign	93	2.66%	0.720
MobileSystems	95	2.72%	0.663
ManagementData	96	2.75%	0.656
ParallelProgramming	96	2.75%	0.698
DataManagement	98	2.81%	0.429
DistributedSimulation	98	2.81%	0.755
ParallelAlgorithms	98	2.81%	0.378
DeclarativeProgramming	101	2.89%	0.822
EmbeddedSystems	102	2.92%	0.657
TheoryComputing	103	2.95%	0.786
DatabaseSystems	104	2.98%	0.721
SoftwareTechnology	104	2.98%	0.683
VolumeVisualization	104	2.98%	0.846
Microarchitecture	105	3.01%	0.667
DataMining	106	3.03%	0.717

2.3 Classic4 Collection

Classic4 collection [Research, 2010] are composed by 4 distinct collections: CACM (titles and abstracts from the journal Communications of the ACM), CISI (information retrieval papers), CRANFIELD (aeronautical system papers), and MEDLINE (medical journals). Table 2.5 presents the characteristics of the Table 2.6 presents the number of documents and the S-Index for each class of the Classic4 collection.

Table 2.5: Classic4 collection characteristics.

Domain	Abstracts
# Documents	7095
# Terms	7749
# Terms	35.28
Matrix Sparsity	99.54%
# Classes	4
Class S.D.	13.70%
Majority Class	45.16%
S-Index	0.889

Table 2.6: Number of documents and S-Index per class for Classic4 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
med	1033	14.56%	0.971
cran	1398	19.70%	0.981
cisi	1460	20.58%	0.958
cacm	3204	45.16%	0.791

2.4 CSTR Collection

CSTR (Computer Science Technical Reports) collection are composed by abstracts and technical reports published in the Department of Computer Science at University of Rochester from 1991 to 2007. The documents belong to 4 areas: Natural Language Processing, Robotics/Vision, Systems, and Theory. We select documents from 16 categories. Table 2.7 presents the characteristics of the Table 2.8 presents the number of documents and the S-Index for each class of the CSTR collection.

Table 2.7: CSTR collection characteristics.

Domain	Scientific
# Documents	299
# Terms	1726
# Terms	54.27
Matrix Sparsity	96.86%
# Classes	4
Class S.D.	15.89%
Majority Class	42.81%
S-Index	0.896

Table 2.8: Number of documents and S-Index per class for CSTR collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Systems	25	8.36%	1.000
Theory	46	15.38%	0.935
Robotics	100	33.44%	0.920
ArtificialIntelligence	128	42.81%	0.844

2.5 Dmoz-Business-500 Collection

Dmoz-Business-500 are composed by web pages of the business category extracted from *DMOZ - Open Directory Project* [Netscape, 2013]. The document class are the subcategories of the business category. We consider 500 document for each class. Table 2.9 presents the characteristics and Table 2.10 presents the number of documents and the S-Index for each class of the Dmoz-Business-500 collection.

Table 2.9: Dmoz-Business-500 collection characteristics.

Domain	Web Pages
# Documents	18500
# Terms	8303
# Terms	11.93
Matrix Sparsity	99.86%
# Classes	37
Class S.D.	0.00%
Majority Class	2.70%
S-Index	0.545

Table 2.10: Number of documents and S-Index per class for Domoz-Business-500 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Accounting	500	2.70%	0.852
Aerospace_and_Defense	500	2.70%	0.436
Agriculture_and_Forestry	500	2.70%	0.550
Arts_and_Entertainment	500	2.70%	0.592
Automotive	500	2.70%	0.474
Biotechnology_and_Pharmaceuticals	500	2.70%	0.584
Business_Services	500	2.70%	0.352
Chemicals	500	2.70%	0.568
Construction_and_Maintenance	500	2.70%	0.356
Consumer_Goods_and_Services	500	2.70%	0.518
E-Commerce	500	2.70%	0.508
Education_and_Training	500	2.70%	0.770
Electronics_and_Electrical	500	2.70%	0.632
Employment	500	2.70%	0.742
Energy	500	2.70%	0.504
Environment	500	2.70%	0.516
Financial_Services	500	2.70%	0.596
Food_and_Related_Products	500	2.70%	0.630
Healthcare	500	2.70%	0.334
Hospitality	500	2.70%	0.454
Human_Resources	500	2.70%	0.480
Industrial_Goods_and_Services	500	2.70%	0.356
Information_Technology	500	2.70%	0.428
International_Business_and_Trade	500	2.70%	0.370
Investing	500	2.70%	0.650
Management	500	2.70%	0.554
Marketing_and_Advertising	500	2.70%	0.512
Materials	500	2.70%	0.550
Mining_and_Drilling	500	2.70%	0.632
Opportunities	500	2.70%	0.480
Publishing_and_Printing	500	2.70%	0.590
Real_Estate	500	2.70%	0.638
Retail_Trade	500	2.70%	0.556
Small_Business	500	2.70%	0.572
Telecommunications	500	2.70%	0.496
Textiles_and_Nonwovens	500	2.70%	0.780
Transportation_and_Logistics	500	2.70%	0.554

2.6 Dmoz-Computers-500 Collection

Dmoz-Computers-500 are composed by web pages of the computers category extracted from *DMOZ - Open Directory Project* [Netscape, 2013]. The document class are the subcategories of the computers category. We consider 500 document for each class. Table 2.11 presents the characteristics and Table 2.12 presents the number of documents and the S-Index for each class of the Dmoz-Computers-500 collection.

Table 2.11: Dmoz-Computers-500 collection characteristics.

Domain	Web Pages
# Documents	9500
# Terms	5011
# Terms	10.83
Matrix Sparsity	99.78%
# Classes	19
Class S.D.	0.00%
Majority Class	5.26%
S-Index	

Table 2.12: Number of documents and S-Index per class for Dmoz-Computers-500 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Artificial_Intelligence	500	5.26%	0.544
CAD_and_CAM	500	5.26%	0.570
Companies	500	5.26%	0.562
Computer_Science	500	5.26%	0.714
Consultants	500	5.26%	0.582
Data_Communications	500	5.26%	0.632
Data_Formats	500	5.26%	0.664
Education	500	5.26%	0.806
Graphics	500	5.26%	0.832
Hardware	500	5.26%	0.578
Internet	500	5.26%	0.626
Mobile_Computing	500	5.26%	0.598
Multimedia	500	5.26%	0.512
Open_Source	500	5.26%	0.596
Programming	500	5.26%	0.524
Robotics	500	5.26%	0.796
Security	500	5.26%	0.506
Software	500	5.26%	0.320
Systems	500	5.26%	0.508

2.7 Dmoz-Health-500 Collection

Dmoz-Health-500 are composed by web pages of the health category extracted from *DMOZ - Open Directory Project* [Netscape, 2013]. The document class are the subcategories of the health category. We consider 500 document for each class. Table 2.13 presents the characteristics and Table 2.14 presents the number of documents and the S-Index for each class of the Dmoz-Health-500 collection.

Table 2.13: Dmoz-Health-500 collection characteristics.

Domain	Web Pages
# Documents	6500
# Terms	4217
# Terms	12.40
Matrix Sparsity	99.71%
# Classes	13
Class S.D.	0.00%
Majority Class	7.69%
S-Index	0.718

Table 2.14: Number of documents and S-Index per class for Dmoz-Health-500 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Addictions	500	7.69%	0.786
Alternative	500	7.69%	0.718
Animal	500	7.69%	0.682
Conditions_and_Diseases	500	7.69%	0.688
Medicine	500	7.69%	0.644
Mental_Health	500	7.69%	0.606
Nursing	500	7.69%	0.840
Nutrition	500	7.69%	0.834
Pharmacy	500	7.69%	0.738
Professions	500	7.69%	0.714
Public_Health_and_Safety	500	7.69%	0.692
Reproductive_Health	500	7.69%	0.574
Senior_Health	500	7.69%	0.812

2.8 Dmoz-Science-500 Collection

Dmoz-Science-500 are composed by web pages of the science category extracted from *DMOZ - Open Directory Project* [Netscape, 2013]. The document class are the subcategories of the science category. We consider 500 document for each class. Table 2.15 presents the characteristics and Table 2.16 presents the number of documents and the S-Index for each class of the Dmoz-Science-500 collection.

Table 2.15: Dmoz-Science-500 collection characteristics.

Domain	Web Pages
# Documents	6000
# Terms	4821
# Terms	11.52
Matrix Sparsity	99.76%
# Classes	12
Class S.D.	0.00%
Majority Class	8.33%
S-Index	0.664

Table 2.16: Number of documents and S-Index per class for Dmoz-Science-500 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Agriculture	500	8.33%	0.722
Astronomy	500	8.33%	0.730
Biology	500	8.33%	0.594
Chemistry	500	8.33%	0.610
Earth_Sciences	500	8.33%	0.548
Environment	500	8.33%	0.628
Instruments_and_Supplies	500	8.33%	0.830
Math	500	8.33%	0.712
Physics	500	8.33%	0.588
Science_in_Society	500	8.33%	0.730
Social_Sciences	500	8.33%	0.692
Technology	500	8.33%	0.578

2.9 Dmoz-Sports-500 Collection

Dmoz-Sports-500 are composed by web pages of the sports category extracted from *DMOZ - Open Directory Project* [Netscape, 2013]. The document class are the subcategories of the sports category. We consider 500 document for each class. Table 2.17 presents the characteristics and Table 2.18 presents the number of documents and the S-Index for each class of the Dmoz-Sports-500 collection.

Table 2.17: Dmoz-Sports-500 collection characteristics.

Domain	Web Pages
# Documents	13500
# Terms	5682
# Terms	11.87
Matrix Sparsity	99.79%
# Classes	27
Class S.D.	0.00%
Majority Class	3.70%
S-Index	0.767

Table 2.18: Number of documents and S-Index per class for Dmoz-Sports-500 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Baseball	500	3.70%	0.658
Basketball	500	3.70%	0.804
Bowling	500	3.70%	0.864
Cricket	500	3.70%	0.824
Cycling	500	3.70%	0.764
Equestrian	500	3.70%	0.774
Fencing	500	3.70%	0.852
Flying_Discs	500	3.70%	0.820
Football	500	3.70%	0.558
Golf	500	3.70%	0.816
Gymnastics	500	3.70%	0.824
Hockey	500	3.70%	0.562
Lacrosse	500	3.70%	0.868
Martial_Arts	500	3.70%	0.858
Motorsports	500	3.70%	0.694
Paintball	500	3.70%	0.730
Running	500	3.70%	0.684
Skating	500	3.70%	0.736
Soccer	500	3.70%	0.696
Softball	500	3.70%	0.852
Strength_Sports	500	3.70%	0.786
Tennis	500	3.70%	0.740
Track_and_Field	500	3.70%	0.886
Volleyball	500	3.70%	0.838
Water_Sports	500	3.70%	0.584
Winter_Sports	500	3.70%	0.786
Wrestling	500	3.70%	0.860

2.10 Enron-Top-20 Collection

Enron-Top-20 collection are derived from Enron dataset [Cohen, 2009]. Enron dataset contain e-mails from 150 users, mostly senior managers of the Enron Corporation, organized into folders. To make this data suitable for classification, we merge all the e-mails from different users that were contained into a folder with the same label. Then we remove folds that could contain e-mails that were not specifically about a subject. The removed folders were: "contact", "sent", "inbox", "e-mail", "save", "archive", "archiving", "attachments", "personal", "personnel", "to_do", "myfriends", "list", "misc", "miscellaneous", "carol_stclair", "general_stuff", "mark", "read", "resumes" e "schedule_crawler".

Folders that contained the same label were also merged into one single folder. All the folders that contain the label "california" were merged into one single folder. All the folders that contain the label "eci" were merged into one single folder. All the folder that contain the label "eol" were merged into one single folder. All the folders that contain the label "rto" were merged into one single folder. All the folder that contain the label "gas" were merged into one single folder. The folders "federal_legis" and "federal_legislation" were merged into one single folder.

The 20 remaining folders with with a higher number of documents were considered to compose the dataset. Table 2.19 presents the characteristics and Table 2.20 presents the number of documents and the S-Index for each class of the Enron-Top-20 collection.

Table 2.19: Enron-Top-20 collection characteristics.

Domain	E-mails
# Documents	13199
# Terms	18194
# Terms	50.69
Matrix Sparsity	99.72%
# Classes	20
Class S.D.	2.37%
Majority Class	9.03%
S-Index	0.656

Table 2.20: Number of documents and S-Index per class for Enron-Top-20 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
_americas_mrha_ooc	309	2.34%	0.394
wellhead	339	2.57%	0.808
universities	367	2.78%	0.785
online_trading	370	2.80%	0.700
c_technote_mail_projects	379	2.87%	0.612
ces	397	3.01%	0.818
corporate	407	3.08%	0.826
iso_pricecaps	420	3.18%	0.640
genco_jv_ipo	489	3.70%	0.759
c_mangmt_group_management	526	3.99%	0.625
tufco	609	4.61%	0.949
c	657	4.98%	0.213
_americas	715	5.42%	0.267
_americas_esvl	715	5.42%	0.511
deal_communication	897	6.80%	0.923
bill_williams_iii	1022	7.74%	0.575
california	1108	8.39%	0.712
meetings	1122	8.50%	0.964
tw_commercial_group	1159	8.78%	0.653
logistics	1192	9.03%	0.484

2.11 FBIS Collection

FBIS (Foreign Broadcast Information Service) collection [Forman, 2006] is part of the TREC collection [TREC, 2013]. This collection is composed by newspaper articles from around the world. Table 2.21 presents the characteristics and Table 2.22 presents the number of documents and the S-Index for each class of the FBIS collection.

Table 2.21: FBIS collection characteristics.

Domain	News Articles
# Documents	2463
# Terms	2001
# Terms	159.24
Matrix Sparsity	92.04%
# Classes	17
Class S.D.	5.66%
Majority Class	20.54%
S-Index	0.809

Table 2.22: Number of documents and S-Index per class for FBIS collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
95	38	1.54%	0.737
4	43	1.75%	0.953
11	46	1.87%	0.978
119	46	1.87%	0.609
240	46	1.87%	0.870
3	48	1.95%	0.917
161	65	2.64%	0.908
100	92	3.74%	0.891
108	94	3.82%	0.979
221	119	4.83%	0.899
12	121	4.91%	0.793
187	125	5.08%	0.872
118	139	5.64%	0.791
202	190	7.71%	0.647
189	358	14.54%	0.754
111	387	15.71%	0.708
142	506	20.54%	0.877

2.12 Hitech Collection

Hitech collection [Karypis, 2006] is derived from the San Jose Mercury newspaper⁵. The news are about computers, electronics, health, medical, research, and technology. This collection is part of the TREC collection [TREC, 2013]. Table 2.23 presents the characteristics and Table 2.24 presents the number of documents and the S-Index for each class of the Hitech collection.

Table 2.23: Hitech collection characteristics.

Domain	News Articles
# Documents	2301
# Terms	12942
# Terms	141.93
Matrix Sparsity	98.90%
# Classes	6
Class S.D.	8.25%
Majority Class	26.21%
S-Index	0.775

Table 2.24: Number of documents and S-Index per class for Hitech collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
electronics	116	5.04%	0.862
technology	187	8.13%	0.722
medical	429	18.64%	0.746
research	481	20.90%	0.823
computer	485	21.08%	0.847
health	603	26.21%	0.698

⁵<http://www.mercurynews.com/>

2.13 Industry-Sector Collection

Industry-Sector collection [Nigam, 2000] is composed by web pages of companies from various economic sectors. Table 2.25 presents the characteristics and Table 2.26 presents the number of documents and the S-Index for each class of the Industry Sector collection.

Table 2.25: Industry-Sector collection characteristics.

Domain	Web Pages
# Documents	8817
# Terms	21490
# Terms	88.49
Matrix Sparsity	99.59%
# Classes	12
Class S.D.	7.37%
Majority Class	29.43%
S-Index	0.526

Table 2.26: Number of documents and S-Index per class for Industry-Sector collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
conglomerates_industry	100	1.13%	0.750
utilities	282	3.20%	0.670
energy	354	4.01%	0.605
healthcare	397	4.50%	0.620
technology	495	5.61%	0.626
transportation	505	5.73%	0.657
consumer_non-cyclical	557	6.32%	0.513
capital_goods	635	7.20%	0.543
basic_materials	949	10.76%	0.551
financial	959	10.88%	0.623
consumer_cyclical	991	11.24%	0.428
services	2595	29.43%	0.423

2.14 Irish-Sentiment Collection

Irish-Sentiment collection [Group, 2009] composed by articles labeled by volunteers as positive, negative or irrelevant. The articles were extracted from Irish online sources: RTE News, The Irish Times, and the Irish Independent. Table 2.27 presents the characteristics and Table 2.28 presents the number of documents and the S-Index for each class of the Irish-Sentiment collection.

Table 2.27: Irish-Sentiment collection characteristics.

Domain	Sentiment Analysis
# Documents	1660
# Terms	8659
# Terms	112.65
Matrix Sparsity	98.70%
# Classes	3
Class S.D.	0.0683
Majority Class	39.46%
S-Index	0.683

Table 2.28: Number of documents and S-Index per class for Irish-Sentiment collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
positive	431	25.96%	0.635
irrelevant	574	34.58%	0.787
negative	655	39.46%	0.617

2.15 La1s Collection

La1s collection [Forman, 2006] is composed by Los Angeles Times news articles extracted from TREC-5 [TREC, 2013]. Table 2.29 presents the characteristics and Table 2.30 presents the number of documents and the S-Index for each class of the La1s collection.

Table 2.29: La1s collection characteristics.

Domain	News Articles
# Documents	3204
# Terms	13196
# Terms	144.64
Matrix Sparsity	98.90%
# Classes	6
Class S.D.	8.22%
Majority Class	29.43%
S-Index	0.846

Table 2.30: Number of documents and S-Index per class for La1s collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
National	273	8.52%	0.780
Foreign	341	10.64%	0.930
Entertainment	354	11.05%	0.915
Financial	555	17.32%	0.854
Sports	738	23.03%	0.818
Metro	943	29.43%	0.825

2.16 La2s Collection

La2s collection [Forman, 2006] is composed by Los Angeles Times news articles extracted from TREC-5 [TREC, 2013]. Table 2.31 presents the characteristics and Table 2.32 presents the number of documents and the S-Index for each class of the La2s collection.

Table 2.31: La2s collection characteristics.

Domain	News Articles
# Documents	3075
# Terms	12433
# Terms	144.83
Matrix Sparsity	98.84%
# Classes	6
Class S.D.	8.59%
Majority Class	29.43%
S-Index	0.854

Table 2.32: Number of documents and S-Index per class for La2s collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
National	248	8.07%	0.754
Foreign	301	9.79%	0.953
Entertainment	375	12.20%	0.941
Financial	487	15.84%	0.850
Sports	759	24.68%	0.835
Metro	905	29.43%	0.831

2.17 LATimes Collection

LATimes collections is the union of the La1s and La2s collectinos. Table 2.33 presents the characteristics and Table 2.34 presents the number of documents and the S-Index for each class of the LATimes collection.

Table 2.33: LATimes collection characteristics.

Domain	News Articles
# Documents	6279
# Terms	10020
# Terms	42.19
Matrix Sparsity	99.58%
# Classes	6
Class S.D.	8.38%
Majority Class	29.43%
S-Index	0.771

Table 2.34: Number of documents and S-Index per class for LATimes collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
national	521	8.30%	0.635
foreign	642	10.22%	0.893
entertainment	729	11.61%	0.871
financial	1042	16.59%	0.798
sports	1497	23.84%	0.720
metro	1848	29.43%	0.752

2.18 Multi-Domain-Sentiment Collection

Multi-Domain-Sentiment collection [Blitzer et al., 2009] contains product reviews taken from Amazon⁶ from different product types. Table 2.35 presents the characteristics and Table 2.36 presents the number of documents and the S-Index for each class of the Multi-Domain-Sentiment collection.

Table 2.35: Multi-Domain-Sentiment collection characteristics.

Domain	Sentiment Analysis
# Documents	8000
# Terms	13360
# Terms	42.36
Matrix Sparsity	99.68%
# Classes	2
Class S.D.	0
Majority Class	50.00%
S-Index	0.748

Table 2.36: Number of documents and S-Index per class for Multi-Domain-Sentiment collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
negative	4000	50.00%	0.817
positive	4000	50.00%	0.680

⁶<http://www.amazon.com/>

2.19 New3 Collection

New 3 collection [Forman, 2006] is composed by news articles from San Jose Mercury⁷. Table 2.37 presents the characteristics and Table 2.38 presents the number of documents and the S-Index for each class of the New3 collection.

Table 2.37: New3 collection characteristics.

Domain	News Articles
# Documents	9558
# Terms	26833
# Terms	234.53
Matrix Sparsity	99.13%
# Classes	44
Class S.D.	1.32%
Majority Class	7.28%
S-Index	0.818

⁷<http://www.mercurynews.com/>

Table 2.38: Number of documents and S-Index per class for New3 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
318	104	1.09%	0.798
400	105	1.10%	0.857
346	106	1.11%	0.943
286	110	1.15%	0.991
324	110	1.15%	0.918
273	115	1.20%	0.913
383	116	1.21%	0.612
390	120	1.26%	0.800
319	123	1.29%	0.911
265	124	1.30%	0.968
100	126	1.32%	0.897
108	130	1.36%	0.931
285	136	1.42%	0.801
360	139	1.45%	0.899
398	139	1.45%	0.964
221	141	1.48%	0.745
391	153	1.60%	0.941
12	159	1.66%	0.843
304	161	1.68%	0.894
187	171	1.79%	0.865
311	171	1.79%	0.801
367	174	1.82%	0.655
389	179	1.87%	0.771
395	181	1.89%	0.785
307	187	1.96%	0.893
331	196	2.05%	0.929
374	198	2.07%	0.788
118	211	2.21%	0.858
321	218	2.28%	0.881
357	238	2.49%	0.882
352	243	2.54%	0.922
343	253	2.65%	0.775
332	270	2.82%	0.852
291	276	2.89%	0.935
202	278	2.91%	0.633
251	281	2.94%	0.744
370	306	3.20%	0.951
269	326	3.41%	0.788
354	328	3.43%	0.701
306	330	3.45%	0.903
301	369	3.86%	0.721
189	493	5.16%	0.688
111	568	5.94%	0.702
142	696	7.28%	0.806

2.20 NSF Collection

NSF (National Science Foundation) collection [Pazzani and Meyers, 2003] are composed by abstracts of grants awarded by the National Science Foundation⁸ between 1999 and August 2003. Table 2.39 presents the characteristics and Table 2.40 presents the number of documents and the S-Index for each class of the NSF collection.

Table 2.39: NSF collection characteristics.

Domain	Scientific
# Documents	10524
# Terms	3888
# Terms	6.56
Matrix Sparsity	99.83%
# Classes	16
Class S.D.	3.82%
Majority Class	13.39%
S-Index	0.735

Table 2.40: Number of documents and S-Index per class for NSF collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
gravitational_theory	130	1.24%	0.915
metals	201	1.91%	0.896
neuroscience	307	2.92%	0.915
networking	345	3.28%	0.829
hydro	355	3.37%	0.808
data_management	402	3.82%	0.764
theory_computing	442	4.20%	0.778
software_engineering	524	4.98%	0.819
politic	603	5.73%	0.610
statistics	647	6.15%	0.685
sociology	739	7.02%	0.648
ecology	889	8.45%	0.848
oceanography	990	9.41%	0.721
geophysics	1202	11.42%	0.785
math	1339	12.72%	0.683
economics	1409	13.39%	0.630

⁸www.nsf.gov

2.21 Oh0 Collection

Oh0 collection [Forman, 2006] is part of the OHSUMED [Hersh et al., 1994] collection. Table 2.41 presents the characteristics and Table 2.42 presents the number of documents and the S-Index for each class of the Oh0 collection.

Table 2.41: Oh0 collection characteristics.

Domain	Medical Documents
# Documents	1003
# Terms	3183
# Terms	52.50
Matrix Sparsity	98.35%
# Classes	10
Class S.D.	5.33%
Majority Class	19.34%
S-Index	0.928

Table 2.42: Number of documents and S-Index per class for Oh0 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Mexico	51	5.08%	0.882
Uric-Acid	56	5.58%	0.875
6-Ketoprostaglandin-F1-alpha	57	5.68%	0.965
Larynx	66	6.58%	0.924
Brain-Chemistry	71	7.08%	0.958
Creatine-Kinase	76	7.58%	0.882
Ethics	115	11.47%	0.930
Fundus-Oculi	136	13.56%	0.978
England	181	18.05%	0.917
Heart-Valve-Prosthesis	194	19.34%	0.928

2.22 Oh5 Collection

Oh5 collection [Forman, 2006] is part of the OHSUMED [Hersh et al., 1994] collection. Table 2.43 presents the characteristics and Table 2.44 presents the number of documents and the S-Index for each class of the Oh0 collection.

Table 2.43: Oh5 collection characteristics.

Domain	Medical Documents
# Documents	918
# Terms	3013
# Terms	54.43
Matrix Sparsity	98.19%
# Classes	10
Class S.D.	3.72%
Majority Class	16.23%
S-Index	0.861

Table 2.44: Number of documents and S-Index per class for Oh5 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Solutions	59	6.43%	0.814
Child-Development	61	6.64%	0.951
Radiation-Dosage	61	6.64%	0.869
Audiometry	72	7.84%	0.917
Anticoagulants	74	8.06%	0.878
Microsomes	85	9.26%	0.847
Nitrogen	93	10.13%	0.871
Neck	120	13.07%	0.917
Phospholipids	144	15.69%	0.750
Graft-Survival	149	16.23%	0.866

2.23 Oh10 Collection

Oh10 collection [Forman, 2006] is part of the OHSUMED [Hersh et al., 1994] collection. Table 2.45 presents the characteristics and Table 2.46 presents the number of documents and the S-Index for each class of the Oh10 collection.

Table 2.45: Oh10 collection characteristics.

Domain	Medical Documents
# Documents	1050
# Terms	3239
# Terms	55.64
Matrix Sparsity	98.28%
# Classes	10
Class S.D.	4.25%
Majority Class	15.71%
S-Index	0.848

Table 2.46: Number of documents and S-Index per class for Oh10 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Sampling-Studies	52	4.95%	0.827
Italy	60	5.71%	0.833
Retinal-Diseases	61	5.81%	0.951
Nutrition	70	6.67%	0.671
Delivery	87	8.29%	0.943
Spermatozoa	116	11.05%	0.888
Arachidonic-Acids	126	12.00%	0.921
Hematocrit	148	14.10%	0.804
Costs-and-Cost-Analysis	165	15.71%	0.848
Eating	165	15.71%	0.800

2.24 Oh15 Collection

Oh15 collection [Forman, 2006] is part of the OHSUMED [Hersh et al., 1994] collection. Table 2.47 presents the characteristics and Table 2.48 presents the number of documents and the S-Index for each class of the Oh15 collection.

Table 2.47: Oh15 collection characteristics.

Domain	Medical Documents
# Documents	3101
# Terms	54142
# Terms	17.46
Matrix Sparsity	99.97%
# Classes	10
Class S.D.	1.26%
Majority Class	5.06%
S-Index	0.857

Table 2.48: Number of documents and S-Index per class for Oh15 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Aluminum	53	1.71%	0.774
Adenosine-Diphosphate	56	1.81%	0.786
Leucine	56	1.81%	0.857
Uremia	66	2.13%	0.879
Blood-Coagulation-Factors	69	2.23%	0.884
Blood-Vessels	98	3.16%	0.806
Memory	98	3.16%	0.898
Cell-Movement	106	3.42%	0.811
Enzyme-Activation	154	4.97%	0.864
Staphylococcal-Infections	157	5.06%	0.917

2.25 Ohscal Collection

Ohscal collection [Forman, 2006] is part of the OHSUMED [Hersh et al., 1994] collection. Table 2.49 presents the characteristics and Table 2.50 presents the number of documents and the S-Index for each class of the Oh15 collection.

Table 2.49: Ohscal collection characteristics.

Domain	Medical Documents
# Documents	11162
# Terms	11466
# Terms	60.38
Matrix Sparsity	99.47%
# Classes	10
Class S.D.	2.66%
Majority Class	14.52%
S-Index	0.697

Table 2.50: Number of documents and S-Index per class for Ohscal collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Carcinoma	709	6.35%	0.673
DNA	764	6.84%	0.560
Molecular-Sequence-Data	864	7.74%	0.766
In-Vitro	1001	8.97%	0.793
Prognosis	1037	9.29%	0.740
Antibodies	1159	10.38%	0.635
Tomography	1260	11.29%	0.713
Receptors	1297	11.62%	0.677
Risk-Factors	1450	12.99%	0.677
Pregnancy	1621	14.52%	0.717

2.26 Ohsumed-400 Collection

Ohsumed-400 collection [Moschitti, 1997] is derived from OHSUMED [Hersh et al., 1994] is composed by medical abstracts about 23 cardiovascular diseases. We selected 400 documents from each category. Table 2.51 presents the characteristics and Table 2.52 presents the number of documents and the S-Index for each class of the Ohsumed-400 collection.

Table 2.51: Ohsumed-400 collection characteristics.

Domain	Medical Documents
# Documents	9200
# Terms	13512
# Terms	55.15
Matrix Sparsity	99.59%
# Classes	12
Class S.D.	0.00%
Majority Class	4.35%
S-Index	0.431

Table 2.52: Number of documents and S-Index per class for Ohsumed-400 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
C01 - Bacterial Infections and Mycoses	400	4.35%	0.337
C02 - Virus Diseases	400	4.35%	0.545
C03 - Parasitic Diseases	400	4.35%	0.520
C04 - Neoplasms	400	4.35%	0.470
C05 - Musculoskeletal Diseases	400	4.35%	0.392
C06 - Digestive System Diseases	400	4.35%	0.532
C07 - Stomatognathic Diseases	400	4.35%	0.457
C08 - Respiratory Tract Diseases	400	4.35%	0.355
C09 - Otorhinolaryngologic Diseases	400	4.35%	0.475
C10 - Nervous System Diseases	400	4.35%	0.335
C11 - Eye Diseases	400	4.35%	0.557
C12 - Urologic and Male Genital Diseases	400	4.35%	0.415
C13 - Female Genital Diseases and Pregnancy Complications	400	4.35%	0.447
C14 - Cardiovascular Diseases	400	4.35%	0.552
C15 - Hemic and Lymphatic Diseases	400	4.35%	0.385
C16 - Neonatal Diseases and Abnormalities	400	4.35%	0.447
C17 - Skin and Connective Tissue Diseases	400	4.35%	0.417
C18 - Nutritional and Metabolic Diseases	400	4.35%	0.520
C19 - Endocrine Diseases	400	4.35%	0.317
C20 - Immunologic Diseases	400	4.35%	0.152
C21 - Disorders of Environmental Origin	400	4.35%	0.472
C22 - Animal Diseases	400	4.35%	0.685
C23 - Pathological Conditions. Signs and Symptoms	400	4.35%	0.140

2.27 Opinois Collection

Opinois collection [Ganesan, 2012] contains user opinions from user for different topics as “Performance of Honda Accord 2008” or “Features of the Windows 7”. The sources of the reviews were Tripadvisor⁹ (hotels), Edmunds.com¹⁰ (cars) and Amazon.com¹¹ (electronics). Table 2.53 presents the characteristics and Table 2.54 presents the number of documents and the S-Index for each class of the Opinois.

Table 2.53: Opinois collection characteristics.

Domain	Sentiment Analysis
# Documents	6457
# Terms	2693
# Terms	7.56
Matrix Sparsity	99.72%
# Classes	51
Class S.D.	1.42%
Majority Class	8.18%
S-Index	0.727

⁹<http://www.tripadvisor.com/>

¹⁰<http://www.edmunds.com/>

¹¹<http://www.amazon.com>

Table 2.54: Number of documents and S-Index per class for Opinosis collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
performance_honda_accord_2008	45	0.70%	0.867
performance_netbook_1005ha	45	0.70%	0.867
display_garmin_nuvi_255W_gps	48	0.74%	0.729
food_swissotel_chicago	51	0.79%	0.765
screen_ipod_nano_8gb	52	0.81%	0.808
features_windows7	56	0.87%	1.000
satellite_garmin_nuvi_255W_gps	57	0.88%	0.947
fonts_amazon_kindle	59	0.91%	0.881
navigation_amazon_kindle	59	0.91%	0.915
battery-life_ipod_nano_8gb	60	0.93%	0.800
updates_garmin_nuvi_255W_gps	62	0.96%	0.903
accuracy_garmin_nuvi_255W_gps	64	0.99%	0.766
speed_garmin_nuvi_255W_gps	64	0.99%	0.906
quality_toyota_camry_2007	67	1.04%	0.866
eyesight-issues_amazon_kindle	73	1.13%	0.822
seats_honda_accord_2008	76	1.18%	0.829
sound_ipod_nano_8gb	79	1.22%	0.924
size_asus_netbook_1005ha	80	1.24%	0.825
battery-life_amazon_kindle	82	1.27%	0.744
bathroom_bestwestern_hotel_sfo	86	1.33%	0.872
voice_garmin_nuvi_255W_gps	87	1.35%	0.874
parking_bestwestern_hotel_sfo	88	1.36%	0.909
interior_honda_accord_2008	89	1.38%	0.764
screen_garmin_nuvi_255W_gps	90	1.39%	0.767
price_amazon_kindle	92	1.42%	0.957
directions_garmin_nuvi_255W_gps	96	1.49%	0.740
food_holiday_inn_london	103	1.60%	0.835
interior_toyota_camry_2007	104	1.61%	0.760
gas_mileage_toyota_camry_2007	110	1.70%	0.727
speed_windows7	111	1.72%	0.919
free_bestwestern_hotel_sfo	112	1.73%	0.911
comfort_toyota_camry_2007	116	1.80%	0.543
keyboard_netbook_1005ha	119	1.84%	0.798
service_bestwestern_hotel_sfo	125	1.94%	0.632
price_holiday_inn_london	130	2.01%	0.862
transmission_toyota_camry_2007	142	2.20%	0.972
rooms_swissotel_chicago	145	2.25%	0.697
mileage_honda_accord_2008	153	2.37%	0.765
service_holiday_inn_london	155	2.40%	0.303
buttons_amazon_kindle	157	2.43%	0.904
comfort_honda_accord_2008	159	2.46%	0.509
video_ipod_nano_8gb	166	2.57%	0.916
screen_netbook_1005ha	171	2.65%	0.737
service_swissotel_hotel_chicago	181	2.80%	0.448
staff_swissotel_chicago	191	2.96%	0.597
rooms_bestwestern_hotel_sfo	241	3.73%	0.552
location_bestwestern_hotel_sfo	266	4.12%	0.613
staff_bestwestern_hotel_sfo	284	4.40%	0.606
battery-life_netbook_1005ha	304	4.71%	0.737
location_holiday_inn_london	377	5.84%	0.833
room_holiday_inn_london	528	8.18%	0.500

2.28 Pubmed-Cancer Collection

Pubmed-Cancer collection are composed by abstracts of scientific articles about 12 types of cancer extracted from Pubmed ¹². Table 2.55 presents the characteristics and Table 2.56 presents the number of documents and the S-Index for each class of the Pubmed-Cancer.

Table 2.55: Pubmed-Cancer collection characteristics.

Domain	Medical Documents
# Documents	65991
# Terms	28329
# Terms	71.20
Matrix Sparsity	99.75%
# Classes	12
Class S.D.	9.09%
Majority Class	26.75%
S-Index	0.841

Table 2.56: Number of documents and S-Index per class for Pubmed-Cancer collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
kidney	104	0.16%	0.817
non-hodgkin lymphoma	433	0.66%	0.875
endometrial	557	0.84%	0.790
thyroid	926	1.40%	0.868
pancreatic	2220	3.36%	0.813
colon	2407	3.65%	0.806
bladder	2848	4.32%	0.777
melanoma	6676	10.12%	0.874
prostate	7390	11.20%	0.825
lung	9177	13.91%	0.842
leukemia	15603	23.64%	0.916
breast	17650	26.75%	0.785

¹²<http://www.ncbi.nlm.nih.gov/pubmed>

2.29 Re0 Collection

Re0 collection [Forman, 2006] is composed by articles from Reuters-21578 collection [Lewis, 2004]. Table 2.57 presents the characteristics and Table 2.58 presents the number of documents and the S-Index for each class of the Re0 collection.

Table 2.57: Re0 collection characteristics.

Domain	News Articles
# Documents	1504
# Terms	2887
# Terms	51.73
Matrix Sparsity	98.21%
# Classes	13
Class S.D.	11.56%
Majority Class	40.43%
S-Index	0.756

Table 2.58: Number of documents and S-Index per class for Re0 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
lei	11	1.00%	1.000
wpi	15	1.00%	1.000
housing	16	1.06%	0.938
retail	20	1.33%	0.850
ipi	37	2.46%	0.946
bop	38	2.53%	0.947
jobs	39	2.59%	0.872
reserves	42	2.79%	1.000
cpi	60	3.99%	0.800
gnp	80	5.32%	0.888
interest	219	14.56%	0.776
trade	319	21.21%	0.781
money	608	40.43%	0.648

2.30 Re1 Collection

Re1 collection [Forman, 2006] is composed by articles from Reuters-21578 collection [Lewis, 2004]. Table 2.59 presents the characteristics and Table 2.60 presents the number of documents and the S-Index for each class of the Re1 collection.

Table 2.59: Re1 collection characteristics.

Domain	News Articles
# Documents	1657
# Terms	3759
# Terms	52.70
Matrix Sparsity	98.60%
# Classes	25
Class S.D.	0.055
Majority Class	22.39%
S-Index	0.830

Table 2.60: Number of documents and S-Index per class for Re1 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
zinc	10	0.60%	1.000
orange	13	0.78%	1.000
meal	15	0.91%	1.000
cotton	17	1.03%	0.882
pet	18	1.09%	0.944
tin	18	1.09%	0.889
gas	19	1.15%	1.000
wheat	19	1.15%	0.895
carcass	20	1.21%	0.800
dlr	20	1.21%	1.000
nat	27	1.63%	0.926
alum	31	1.87%	0.968
iron	31	1.87%	0.903
rubber	32	1.93%	0.906
copper	37	2.23%	0.946
livestock	42	2.53%	0.952
cocoa	48	2.90%	0.896
oilseed	50	3.02%	0.660
veg	60	3.62%	0.800
gold	87	5.25%	0.977
coffee	99	5.97%	0.949
sugar	106	6.40%	0.943
ship	137	8.27%	0.869
crude	330	19.92%	0.812
grain	371	22.39%	0.650

2.31 Re8 Collection

Re8 collection [Forman, 2006] is composed by articles from Reuters-21578 collection [Lewis, 2004]. Table 2.61 presents the characteristics and Table 2.62 presents the number of documents and the S-Index for each class of the Re8 collection.

Table 2.61: Re8 collection characteristics.

Domain	News Articles
# Documents	7674
# Terms	8901
# Terms	35.31
Matrix Sparsity	99.60%
# Classes	8
Class S.D.	18.24%
Majority Class	51.12%
S-Index	0.886

Table 2.62: Number of documents and S-Index per class for Re8 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
grain	51	0.66%	0.961
ship	144	1.88%	0.896
interest	271	3.53%	0.878
money	293	3.82%	0.857
trade	326	4.25%	0.954
crude	374	4.87%	0.917
acq	2292	29.87%	0.940
earn	3923	51.12%	0.847

2.32 Review-Polarity Collection

Review-Polarity collection [Pang and Lee, 2004] are composed by 1000 positive reviews and 1000 negative reviews about movies. Table 2.63 presents the characteristics and Table 2.64 presents the number of documents and the S-Index for each class of the Review-Polarity collection.

Table 2.63: Review-Polarity collection characteristics.

Domain	Sentiment Analysis
# Documents	2000
# Terms	15698
# Terms	205.06
Matrix Sparsity	98.69%
# Classes	2
Class S.D.	0.00%
Majority Class	50.00%
S-Index	0.731

Table 2.64: Number of documents and S-Index per class for Review-Polarity collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
neg	1000	50.00%	0.704
pos	1000	50.00%	0.758

2.33 Reviews Collection

Reviews collection [Karypis, 2006] is derived from the San Jose Mercury newspaper¹³. The news are about food, movies, music, radio, and restaurants. Table 2.65 presents the characteristics and Table 2.66 presents the number of documents and the S-Index for each class of the Reviews collection.

Table 2.65: Reviews collection characteristics.

Domain	News Articles
# Documents	4069
# Terms	22927
# Terms	183.10
Matrix Sparsity	99.20%
# Classes	5
Class S.D.	12.80%
Majority Class	34.11%
S-Index	0.912

Table 2.66: Number of documents and S-Index per class for Reviews collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
radio	137	3.37%	0.891
restaurant	412	10.13%	0.920
food	999	24.55%	0.863
movie	1133	27.84%	0.915
music	1388	34.11%	0.944

¹³<http://www.mercurynews.com/>

2.34 SpamAssassin Collection

SpamAssassin collection [Apache, 2006] is a public e-mail collection for testing spam filtering systems. This collection is composed by spams and non-spam (ham) e-mails. Table 2.67 presents the characteristics and Table 2.68 presents the number of documents and the S-Index for each class of the SpamAssassin collection.

Table 2.67: SpamAssassin collection characteristics.

Domain	E-mails
# Documents	9348
# Terms	97851
# Terms	108.02
Matrix Sparsity	99.89%
# Classes	2
Class S.D.	34.45%
Majority Class	74.36%
S-Index	0.888

Table 2.68: Number of documents and S-Index per class for SpamAssassin collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Spam	2397	25.64%	0.859
Ham	6951	74.36%	0.899

2.35 SpamTrec-3000 Collection

SpamTrec-3000 collection [Cormack and Lynam, 2007] is composed by spam and non-spam (ham) e-mails. We selected 3000 e-mail from each type of e-mail. Table 2.69 presents the characteristics and Table 2.70 presents the number of documents and the S-Index for each class of the SpamTrec-3000.

Table 2.69: SpamTrec-3000 collection characteristics.

Domain	E-mails
# Documents	3000
# Terms	100464
# Terms	488.17
Matrix Sparsity	99.51%
# Classes	2
Class S.D.	0.00%
Majority Class	100.00%
S-Index	0.926

Table 2.70: Number of documents and S-Index per class for SpamTrec-3000 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
ham	3000	100.00%	0.961
spam	3000	100.00%	0.890

2.36 SyskillWebert Collection

SyskillWebert collection [Pazzani, 1998] is composed by web pages about bands, sheeps, goats, and biomedical. Table 2.71 presents the characteristics and Table 2.72 presents the number of documents and the S-Index for each class of the SyskillWebert collection.

Table 2.71: SyskillWebert collection characteristics.

Domain	Web Pages
# Documents	334
# Terms	4340
# Terms	93.16
Matrix Sparsity	97.85%
# Classes	4
Class S.D.	10.75%
Majority Class	41.02%
S-Index	0.964

Table 2.72: Number of documents and S-Index per class for SyskillWebert collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Bands	61	18.26%	1.000
Sheep	65	19.46%	0.892
Goats	71	21.26%	0.930
Biomedical	137	41.02%	1.000

2.37 Tr11 Collection

Tr11 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to documents that were judged relevant to particular queries. Table 2.73 presents the characteristics and Table 2.74 presents the number of documents and the S-Index for each class of the Tr11 collection.

Table 2.73: Tr11 collection characteristics.

Domain	TREC Documents
# Documents	414
# Terms	6430
# Terms	281.66
Matrix Sparsity	95.62%
# Classes	9
Class S.D.	9.80%
Majority Class	31.88%
S-Index	0.937

Table 2.74: Number of documents and S-Index per class for Tr11 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
44	6	1.45%	1.000
5	11	2.66%	1.000
24	20	4.83%	1.000
1	21	5.07%	0.952
6	29	7.00%	0.931
4	52	12.56%	0.846
3	69	16.67%	0.884
11	74	17.87%	1.000
12	132	31.88%	0.946

2.38 Tr12 Collection

Tr12 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to the documents that were judged relevant to particular queries. Table 2.75 presents the characteristics and Table 2.76 presents the number of documents and the S-Index for each class of the Tr12 collection.

Table 2.75: Tr12 collection characteristics.

Domain	TREC Documents
# Documents	313
# Terms	5805
# Terms	273.60
Matrix Sparsity	95.29%
# Classes	8
Class S.D.	7.98%
Majority Class	29.71%
S-Index	0.955

Table 2.76: Number of documents and S-Index per class for Tr12 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
77	9	2.88%	1.000
58	29	9.27%	1.000
94	29	9.27%	0.897
78	30	9.58%	1.000
54	34	10.86%	1.000
82	35	11.18%	0.971
95	54	17.25%	0.889
100	93	29.71%	0.956

2.39 Tr21 Collection

Tr21 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to the documents that were judged relevant to particular queries. Table 2.77 presents the characteristics and Table 2.78 presents the number of documents and the S-Index for each class of the Tr21 collection.

Table 2.77: Tr21 collection characteristics.

Domain	TREC Documents
# Documents	336
# Terms	7903
# Terms	469.86
Matrix Sparsity	94.05%
# Classes	6
Class S.D.	0.2587882762
Majority Class	68.75%
S-Index	0.842

Table 2.78: Number of documents and S-Index per class for Tr21 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
259	4	1.19%	1.000
254	9	2.68%	1.000
252	16	4.76%	1.000
257	35	10.42%	0.971
255	41	12.20%	0.878
251	231	68.75%	0.796

2.40 Tr23 Collection

Tr23 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to the documents that were judged relevant to particular queries. Table 2.79 presents the characteristics and Table 2.80 presents the number of documents and the S-Index for each class of the Tr23 collection.

Table 2.79: Tr23 collection characteristics.

Domain	TREC Documents
# Documents	204
# Terms	5833
# Terms	385.29
Matrix Sparsity	93.39%
# Classes	6
Class S.D.	15.58%
Majority Class	44.61%
S-Index	0.843

Table 2.80: Number of documents and S-Index per class for Tr23 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
280	6	2.94%	1.000
272	11	5.39%	0.818
277	15	7.35%	0.933
271	36	17.65%	0.806
274	45	22.06%	0.733
273	91	44.61%	0.890

2.41 Tr31 Collection

Tr31 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to the documents that were judged relevant to particular queries. Table 2.81 presents the characteristics and Table 2.82 presents the number of documents and the S-Index for each class of the Tr31 collection.

Table 2.81: Tr31 collection characteristics.

Domain	TREC Documents
# Documents	927
# Terms	10129
# Terms	268.50
Matrix Sparsity	97.35%
# Classes	7
Class S.D.	13.37%
Majority Class	37.97%
S-Index	0.929

Table 2.82: Number of documents and S-Index per class for Tr31 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
310	2	0.22%	1.000
305	21	2.27%	1.000
302	63	6.80%	0.888
307	111	11.97%	0.819
304	151	16.29%	0.867
306	227	24.49%	0.991
301	352	37.97%	0.954

2.42 Tr41 Collection

Tr41 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to the documents that were judged relevant to particular queries. Table 2.83 presents the characteristics and Table 2.84 presents the number of documents and the S-Index for each class of the Tr41 collection.

Table 2.83: Tr41 collection characteristics.

Domain	TREC Documents
# Documents	8778
# Terms	7455
# Terms	19.54
Matrix Sparsity	99.74%
# Classes	10
Class S.D.	0.91%
Majority Class	2.77%
S-Index	0.933

Table 2.84: Number of documents and S-Index per class for Tr41 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
356	9	0.10%	1.000
359	18	0.21%	1.000
351	26	0.30%	0.962
358	33	0.38%	1.000
355	35	0.40%	0.971
360	83	0.95%	0.976
353	95	1.08%	0.842
357	162	1.85%	0.914
352	174	1.98%	0.994
354	243	2.77%	0.897

2.43 Tr45 Collection

Tr45 collection [Forman, 2006] are derived from Trec-5, Trec-6 and Trec-7 collections [TREC, 2013]. The classes correspond to the documents that were judged relevant to particular queries. Table 2.85 presents the characteristics and Table 2.86 presents the number of documents and the S-Index for each class of the Tr45 collection.

Table 2.85: Tr45 collection characteristics.

Domain	TREC Documents
# Documents	690
# Terms	8262
# Terms	280.58
Matrix Sparsity	96.60%
# Classes	10
Class S.D.	6.69%
Majority Class	23.19%
S-Index	0.924

Table 2.86: Number of documents and S-Index per class for Tr45 collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
394	14	2.03%	0.857
397	18	2.61%	0.944
396	36	5.22%	0.861
393	47	6.81%	0.893
399	63	9.13%	0.809
400	67	9.71%	0.910
392	75	10.87%	0.840
398	82	11.88%	0.963
391	128	18.55%	0.984
395	160	23.19%	0.975

2.44 WAP Collection

WAP (WebACE Project) collection [Forman, 2006] is composed by web pages from de WebACE Projec [Han et al., 1998]. The web pages belong to the subject hierarchy of Yahoo¹⁴. Table 2.87 presents the characteristics and Table 2.88 presents the number of documents and the S-Index for each class of the WAP collection.

Table 2.87: Wap collection characteristics.

Domain	Web Pages
# Documents	1560
# Terms	8461
# Terms	141.33
Matrix Sparsity	98.33%
# Classes	20
Class S.D.	5.20%
Majority Class	21.86%
S-Index	0.908

Table 2.88: Number of documents and S-Index per class for Wap collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
Entertainment	5	0.32%	1.000
Multimedia	11	0.71%	0.909
Stage	13	0.83%	1.000
Art	15	0.96%	0.867
Media	18	1.15%	1.000
Cable	33	2.12%	0.788
Variety	35	2.24%	1.000
Technology	37	2.37%	0.946
Online	40	2.56%	0.900
Industry	44	2.82%	0.841
Culture	54	3.46%	0.630
Politics	65	4.17%	0.877
Business	76	4.87%	0.868
Music	91	5.83%	0.956
Review	91	5.83%	0.901
Sports	97	6.22%	0.990
Television	130	8.33%	0.908
People	168	10.77%	0.756
Film	196	12.56%	0.939
Health	341	21.86%	0.988

¹⁴<http://dir.yahoo.com/>

2.45 WebKB Collection

WebKB collection is composed by web pages collected from computer science departments of various universities in January 1997 by the World Wide Knowledge Base¹⁵ (WebKb) project of the CMU Text Learning Group¹⁶. Table 2.89 presents the characteristics and Table 2.90 presents the number of documents and the S-Index for each class of the WebKB collection.

Table 2.89: WebKb collection characteristics.

Domain	Web Pages
# Documents	8282
# Terms	22892
# Terms	89.78
Matrix Sparsity	99.61%
# Classes	7
Class S.D.	15.19%
Majority Class	45.45%
S-Index	0.566

Table 2.90: Number of documents and S-Index per class for WebKb collection.

Class Labels	Abs. # Docs.	Rel. # Docs.	S-Index
staff	137	1.65%	0.438
department	182	2.20%	0.901
project	504	6.09%	0.724
course	930	11.23%	0.728
faculty	1124	13.57%	0.654
student	1641	19.81%	0.567
other	3764	45.45%	0.466

¹⁵<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wskb/index.html>

¹⁶<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-4/text-learning/www/index.html>

2.46 Summary of the characteristics for all the collections

All the characteristics of the collections analyzed in this technical report are summarized in Table 2.91.

Table 2.91: Characteristics of all the text collections analyzed in this technical report

Collection	$ \mathcal{D} $	$ \mathcal{T} $	$ \overline{\mathcal{T}} $	M.S.	$ \mathcal{C} $	$\sigma(\mathcal{C})$	$\max(\mathcal{C})$	$S - Index$	Domain
20Ng	18808	45434	76.47	99.83%	20	0.52%	5.31%	0.80	EM
ACM	3493	60768	720.3	98.81%	40	0.37%	3.03%	0.75	SD
Classic4	7095	7749	35.28	99.54%	4	13.70%	45.16%	0.89	SD
CSTR	299	1726	54.27	96.86%	4	15.89%	42.81%	0.90	SD
Dmoz-Business-500	18500	8303	11.92	99.86%	37	0.00%	2.70%	0.55	WP
Dmoz-Computers-500	9500	5011	10.83	99.78%	19	0.00%	5.26%	0.00	WP
Dmoz-Health-500	6500	4217	12.39	99.71%	13	0.00%	7.69%	0.72	WP
Dmoz-Science-500	6000	4821	11.52	99.76%	12	0.00%	8.33%	0.66	WP
Dmoz-Sports-500	13500	5682	11.87	99.79%	27	0.00%	3.70%	0.58	WP
Enron-Top-20	13199	18194	50.69	99.72%	20	2.37%	9.03%	0.66	EM
FBIS	2463	2001	159.24	92.04%	17	5.66%	20.54%	0.81	NA
Hitech	2301	12942	141.93	98.90%	6	8.25%	26.21%	0.78	NA
Industry Sector	8817	21490	88.48	99.59%	12	7.37%	29.43%	0.53	WP
Irish-Sentiment	1660	8659	112.64	98.70%	3	6.83%	39.46%	0.68	SA
La1s	3204	13196	144.63	98.90%	6	8.22%	29.43%	0.85	NA
La2s	3075	12433	144.82	98.84%	6	8.59%	29.43%	0.85	NA
LATimes	6279	10020	42.19	99.58%	6	8.38%	29.43%	0.77	NA
Multi-Domain-Sentiment	8000	13360	42.36	99.68%	2	0.00%	50.00%	0.75	SA
New3	9558	26833	234.53	99.13%	44	1.32%	7.28%	0.82	NA
NSF	10524	3888	6.55	99.83%	16	3.82%	13.39%	0.74	SD
Oh0	1003	3183	52.50	98.35%	10	5.33%	19.34%	0.93	MD
Oh5	918	3013	54.43	98.19%	10	3.72%	16.23%	0.86	MD
Oh10	1050	3239	55.63	98.28%	10	4.25%	15.71%	0.85	MD
Oh15	3101	54142	17.46	99.97%	10	1.26%	5.06%	0.86	MD
Ohscal	11162	11466	60.38	99.47%	10	2.66%	14.52%	0.70	MD
Ohsumed-400	9200	13512	55.14	99.59%	12	0.00%	4.35%	0.00	MD
Opinosis	6457	2693	7.55	99.72%	51	1.42%	8.18%	0.73	SA
Pubmed-Cancer	65991	28329	71.19	99.75%	12	9.09%	26.75%	0.84	MD
Re0	1504	2887	51.72	98.21%	13	11.56%	40.43%	0.76	NA
Re1	1657	3759	52.69	98.60%	25	5.54%	22.39%	0.83	NA
Re8	7674	8901	35.30	99.60%	8	18.24%	51.12%	0.89	NA
Review Polarity	2000	15698	205.06	98.69%	2	0.00%	50.00%	0.73	SA
Reviews	4069	22927	183.10	99.20%	5	12.80%	34.11%	0.91	NA
SpamAssassin	9348	97851	108.02	99.89%	2	34.45%	74.36%	0.89	EM
SpamTrec-3000	6000	100464	244.08	99.76%	2	0.00%	50.00%	0.93	EM
SyskillWebert	334	4340	93.15	97.85%	4	10.75%	41.02%	0.96	WP
Tr11	414	6430	281.66	95.62%	9	9.80%	31.88%	0.00	TD
Tr12	313	5805	273.59	95.29%	8	7.98%	29.71%	0.96	TD
Tr21	336	7903	469.86	94.05%	6	25.88%	68.75%	0.00	TD
Tr23	204	5833	385.29	93.39%	6	15.58%	44.61%	0.84	TD
Tr31	927	10129	268.49	97.35%	7	13.37%	37.97%	0.00	TD
Tr41	8778	7455	19.53	99.74%	10	0.91%	2.77%	0.93	TD
Tr45	690	8262	280.58	96.60%	10	6.69%	23.19%	0.93	TD
WAP	1560	8461	141.33	98.33%	20	5.20%	21.86%	0.91	WP
WebKB	8282	22892	89.77	99.61%	7	15.19%	45.45%	0.57	WP

Chapter 3

Classification Results

We developed the Inductive Classification Tool¹ (ICT) to generate the classification results. This tool was developed in Java language and uses the Weka library² to read the representations of the text collections in **arff** format and run the classification algorithms.

We run traditional classification algorithms from different paradigms: Naïve Bayes and Multinomial Naïve Bayes (probabilistic paradigms), J48³ (symbolic paradigm), SMO⁴ (statistical paradigm), and IBk⁵ (instance-based paradigm) [Witten and Frank, 2005].

For SMO algorithm we used 3 types of kernel: linear, polynomial (exponent = 2) and RBF (radial bases function). The values considered for C were: $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$ [Caruana and Niculescu-Mizil, 2006].

For IBk algorithm we used $k = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 25, 29, 35, 41, 49, 57, 73, 89\}$. Some studies evaluating the k -NN algorithm selected a set of values ranging from 1 to $|number_of_documents|$. We did not select this variation because we tested the k -NN without and with a weighted vote (WV), which gives for each of the nearest neighbors a weight vote equal to $1/(1 - s)$, where s is a similarity measure between neighbors. Without the weighting scheme, as the number of neighbors becomes closer to $|number_of_documents|$, the documents are classified as the majority class. The cosine distance was used as similarity measure.

Let $\mathcal{C} = \{c_1, c_2, \dots, c_l\}$ be the set of label classes of a text collection. Let c_i be the interest class of a classifier. We can have four possibilities during the evaluation of a classifier:

- **True Positives (TP)**: examples that belong to class c_i and are classified as c_i .
- **True Negatives (TN)**: examples that do not belong to class c_i and are not classified as c_i .

¹ICT is available at <http://sites.labic.icmc.usp.br/ict/>

²Weka library: <http://www.cs.waikato.ac.nz/ml/weka/>

³J48 is the implementation of the C4.5 algorithm.

⁴SMO (Sequential Minimal Optimization) is used in the optimization process of the Support Vector Machines

⁵IBk is the implementation of the k -Nearest Neighbors algorithm

- **False Positives (FP)**: examples that do not belong to class c_i and are classified as c_i .
- **False Negatives (FN)**: examples that belong to class c_i and are no classified as c_i .

All these four possibilities are summarized in a local contingency table as presented in Table 3.1. When we want to consider all the classes of a collection we can summarize these possibilities in a global contingency table as presented in Table 3.2.

Table 3.1: Contingency table for category c_i [Sebastiani, 2002].

Category c_i		Real Class	
		P	N
Predicted Class	P	TP_i	FP_i
	N	FN_i	TN_i

Table 3.2: Global contingency table [Sebastiani, 2002].

Category c_i		Real Class	
		P	N
Predicted Class	P	$TP = \sum_{i=1}^{ C } TP_i$	$FP = \sum_{i=1}^{ C } FP_i$
	N	$FN = \sum_{i=1}^{ C } FN_i$	$TN = \sum_{i=1}^{ C } TN_i$

The items of the Tables 3.1 and 3.2 composes the evaluation measures used in this technical report. The evaluation measures used here were:

- **CCR (Correct Classification Rate)**: is the percentage of correctly classified documents.

$$CCR = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

- **ICR (Incorrect Classification Rate)**: is the percentage of incorrectly classified documents.

$$IRC = \frac{FP + FN}{TP + TN + FP + FN} = 1 - CCR \quad (3.2)$$

- $\sigma(\text{CCR})$: is the standard deviation of the CCR when a X -fold cross-validation procedure is used [Tan et al., 2005].
- **Pr (Precision)**: is the rate of TP considering all the positive predictions of a classifier. The precision of a class c_i is obtained by the Equation 3.3.

$$Pr_i = \frac{TP_i}{TP_i + FP_i} \quad (3.3)$$

- **Re (Recall)**: is the rate of TP considering all the real positive examples. The recall of a class c_i is obtained by the Equation 3.4.

$$Re_i = \frac{TP_i}{TP_i + FN_i} \quad (3.4)$$

- **F₁**: is the harmonic mean of precision and recall. F_1 measure is obtained by the Equation 3.5

$$F_{1_i} = \frac{2 * Pr_i * Re_i}{Pr_i + Re_i} \quad (3.5)$$

- **Microaveraging**: summarizes the precision, recall, and F_1 for all the classes of a collection through the sum of the individual parts of the equations. Micro-precision (Pr^μ), micro-recall (Re^μ), and micro- F_1 (F_1^μ) are given by Equations 3.6, 3.7 and 3.8 respectively.

$$Pr^\mu = \frac{\sum_{i=1}^c TP_i}{\sum_{i=1}^{|C|} (TP_i + FP_i)} \quad (3.6)$$

$$Re^\mu = \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} (TP_i + FN_i)} \quad (3.7)$$

$$F_1^\mu = \frac{2 * Pr^\mu * Re^\mu}{Pr^\mu + Re^\mu} \quad (3.8)$$

- **Macroaveraging**: summarizes the precision, recall, and F_1 for all the classes of a collection through the arithmetic mean of the evaluation measures for each class. Macro-precision (Pr^M), macro-recall (Re^M), and macro- F_1 are given by Equations 3.9, 3.10, and 3.11 respectively.

$$Pr^M = \frac{\sum_{i=1}^{|C|} Pr_i}{|C|} \quad (3.9)$$

$$Re^M = \frac{\sum_{i=1}^{|C|} Re_i}{|C|} \quad (3.10)$$

$$F_1 = \frac{2 * Pr^M * Re^M}{Pr^M + Re^M} \quad (3.11)$$

All the values of the evaluation measures were obtained by an arithmetic mean of the results produced in each step of a 10-fold cross-validation procedure. The next subsections presents the classification results for the collections analyzed in this technical report.

3.1 20ng Collection

Table 3.3 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.4 presents the classification results obtained by IBk for 20ng collection.

Table 3.3: Classification Results for 20ng collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	64.606	35.394	1.156	0.646	0.646	0.662	0.646	0.646	0.654
Multinomial NB	90.084	9.916	0.539	0.901	0.901	0.902	0.896	0.901	0.899
J48	73.959	26.041	1.033	0.740	0.740	0.738	0.736	0.740	0.737
SMO-Linear-0	5.205	94.795	0.644	0.052	0.052	0.003	0.050	0.052	0.005
SMO-Linear- 10^{-5}	4.605	95.395	0.217	0.046	0.046	0.002	0.050	0.046	0.004
SMO-Linear- 10^{-4}	4.621	95.379	0.232	0.046	0.046	0.010	0.050	0.046	0.016
SMO-Linear- 10^{-3}	4.801	95.199	0.247	0.048	0.048	0.131	0.052	0.048	0.074
SMO-Linear- 10^{-2}	41.662	58.338	3.472	0.417	0.417	0.884	0.417	0.417	0.567
SMO-Linear- 10^{-1}	82.308	17.692	0.868	0.823	0.823	0.876	0.821	0.823	0.848
SMO-Linear- 10^0	84.810	15.190	0.586	0.848	0.848	0.860	0.847	0.848	0.854
SMO-Linear- 10^1	83.211	16.789	0.619	0.832	0.832	0.841	0.832	0.832	0.837
SMO-Linear- 10^2	81.894	18.106	0.516	0.819	0.819	0.834	0.819	0.819	0.826
SMO-Linear- 10^3	81.193	18.807	0.642	0.812	0.812	0.837	0.812	0.812	0.824
SMO-Linear- 10^4	81.193	18.807	0.642	0.812	0.812	0.837	0.812	0.812	0.824
SMO-Linear- 10^5	81.193	18.807	0.642	0.812	0.812	0.837	0.812	0.812	0.824
SMO-Poly-0	5.205	94.795	0.644	0.052	0.052	0.003	0.050	0.052	0.005
SMO-Poly- 10^{-5}	5.444	94.556	1.556	0.054	0.054	0.082	0.058	0.054	0.068
SMO-Poly- 10^{-4}	5.072	94.928	0.429	0.051	0.051	0.262	0.054	0.051	0.090
SMO-Poly- 10^{-3}	11.669	88.331	2.048	0.117	0.117	0.844	0.120	0.117	0.210
SMO-Poly- 10^{-2}	47.217	52.783	2.603	0.472	0.472	0.875	0.470	0.472	0.612
SMO-Poly- 10^{-1}	65.371	34.629	1.314	0.654	0.654	0.852	0.651	0.654	0.738
SMO-Poly- 10^0	68.233	31.767	1.525	0.682	0.682	0.812	0.680	0.682	0.740
SMO-Poly- 10^1	68.032	31.968	1.522	0.680	0.680	0.793	0.679	0.680	0.731
SMO-Poly- 10^2	64.978	35.022	1.159	0.650	0.650	0.780	0.649	0.650	0.709
SMO-Poly- 10^3	61.393	38.607	1.623	0.614	0.614	0.772	0.613	0.614	0.684
SMO-Poly- 10^4	59.103	40.897	1.130	0.591	0.591	0.787	0.591	0.591	0.675
SMO-Poly- 10^5	58.838	41.162	1.278	0.588	0.588	0.789	0.589	0.588	0.674
SMO-RBF-0	5.205	94.795	0.644	0.052	0.052	0.003	0.050	0.052	0.005
SMO-RBF- 10^{-5}	4.605	95.395	0.217	0.046	0.046	0.002	0.050	0.046	0.004
SMO-RBF- 10^{-4}	4.605	95.395	0.217	0.046	0.046	0.002	0.050	0.046	0.004
SMO-RBF- 10^{-3}	4.605	95.395	0.217	0.046	0.046	0.002	0.050	0.046	0.004
SMO-RBF- 10^{-2}	4.626	95.374	0.247	0.046	0.046	0.003	0.050	0.046	0.005
SMO-RBF- 10^{-1}	6.124	93.876	1.565	0.061	0.061	0.070	0.066	0.061	0.068
SMO-RBF- 10^0	54.717	45.283	2.197	0.547	0.547	0.808	0.547	0.547	0.652
SMO-RBF- 10^1	82.329	17.671	0.738	0.823	0.823	0.851	0.823	0.823	0.836
SMO-RBF- 10^2	83.301	16.699	0.769	0.833	0.833	0.840	0.833	0.833	0.837
SMO-RBF- 10^3	82.271	17.729	0.713	0.823	0.823	0.830	0.823	0.823	0.826
SMO-RBF- 10^4	80.486	19.514	0.808	0.805	0.805	0.821	0.806	0.805	0.813
SMO-RBF- 10^5	80.476	19.524	0.809	0.805	0.805	0.822	0.806	0.805	0.813

Table 3.4: Classification Results for 20ng collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	86.706	13.294	0.620	0.867	0.867	0.866	0.867	0.867	0.867
IB3	83.880	16.120	0.902	0.839	0.839	0.841	0.839	0.839	0.840
IB5	83.556	16.444	0.866	0.836	0.836	0.836	0.834	0.836	0.835
IB7	82.707	17.293	0.731	0.827	0.827	0.826	0.825	0.827	0.826
IB9	82.207	17.793	0.821	0.822	0.822	0.821	0.820	0.822	0.821
IB11	81.798	18.202	0.903	0.818	0.818	0.816	0.816	0.818	0.816
IB13	81.400	18.600	0.884	0.814	0.814	0.813	0.812	0.814	0.812
IB15	81.182	18.818	0.744	0.812	0.812	0.810	0.809	0.812	0.810
IB17	80.848	19.152	0.712	0.808	0.808	0.807	0.805	0.808	0.806
IB19	80.587	19.413	0.590	0.806	0.806	0.804	0.802	0.806	0.803
IB21	80.364	19.636	0.879	0.804	0.804	0.803	0.800	0.804	0.801
IB25	80.088	19.912	0.742	0.801	0.801	0.801	0.797	0.801	0.799
IB29	79.621	20.379	0.618	0.796	0.796	0.795	0.792	0.796	0.794
IB35	79.010	20.990	0.723	0.790	0.790	0.789	0.785	0.790	0.787
IB41	78.601	21.399	0.794	0.786	0.786	0.787	0.781	0.786	0.784
IB49	78.118	21.882	0.718	0.781	0.781	0.784	0.776	0.781	0.780
IB57	-	-	-	-	-	-	-	-	-
IB73	77.278	22.722	0.634	0.773	0.773	0.778	0.766	0.773	0.772
IB89	76.737	23.263	0.541	0.767	0.767	0.775	0.760	0.767	0.767
IB1-WV	86.706	13.294	0.620	0.867	0.867	0.866	0.867	0.867	0.867
IB3-WV	86.265	13.735	0.729	0.863	0.863	0.862	0.862	0.867	0.862
IB5-WV	85.628	14.372	0.861	0.856	0.856	0.856	0.855	0.867	0.855
IB7-WV	84.890	15.110	0.774	0.849	0.849	0.848	0.847	0.849	0.848
IB9-WV	84.465	15.535	0.722	0.845	0.845	0.844	0.843	0.845	0.843
IB11-WV	84.162	15.838	0.820	0.842	0.842	0.840	0.839	0.842	0.840
IB13-WV	83.726	16.274	0.698	0.837	0.837	0.836	0.834	0.837	0.835
IB15-WV	83.243	16.757	0.655	0.832	0.832	0.831	0.829	0.832	0.830
IB17-WV	82.940	17.060	0.737	0.829	0.829	0.828	0.826	0.829	0.827
IB19-WV	82.579	17.421	0.635	0.826	0.826	0.825	0.822	0.826	0.824
IB21-WV	82.229	17.771	0.764	0.822	0.822	0.822	0.819	0.822	0.821
IB25-WV	81.862	18.138	0.588	0.819	0.819	0.819	0.815	0.819	0.817
IB29-WV	81.427	18.573	0.583	0.814	0.814	0.815	0.810	0.814	0.812
IB35-WV	80.949	19.051	0.684	0.809	0.809	0.811	0.805	0.809	0.808
IB41-WV	80.518	19.482	0.684	0.805	0.805	0.807	0.800	0.805	0.804
IB49-WV	79.966	20.034	0.699	0.800	0.800	0.803	0.794	0.800	0.798
IB57-WV	-	-	-	-	-	-	-	-	-
IB73-WV	78.798	21.202	0.610	0.788	0.788	0.794	0.781	0.788	0.787
IB89-WV	78.266	21.734	0.592	0.783	0.783	0.790	0.775	0.783	0.783

3.2 ACM Collection

Table 3.5 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.6 presents the classification results obtained by IBk for ACM collection.

Table 3.5: Classification Results for ACM collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	73.891	26.109	2.820	0.739	0.739	0.758	0.737	0.739	0.747
Multinomial NB	76.038	23.962	2.393	0.760	0.760	0.778	0.760	0.760	0.769
J48	66.360	33.640	2.072	0.664	0.664	0.673	0.664	0.664	0.668
SMO-Linear-0	3.035	96.965	1.090	0.030	0.030	0.001	0.025	0.030	0.001
SMO-Linear- 10^{-5}	2.090	97.910	0.725	0.021	0.021	0.003	0.017	0.021	0.005
SMO-Linear- 10^{-4}	5.784	94.216	3.195	0.058	0.058	0.057	0.049	0.058	0.053
SMO-Linear- 10^{-3}	2.090	97.910	0.725	0.021	0.021	0.003	0.017	0.021	0.005
SMO-Linear- 10^{-2}	67.479	32.521	3.055	0.675	0.675	0.776	0.656	0.675	0.711
SMO-Linear- 10^{-1}	77.384	22.616	2.344	0.774	0.774	0.804	0.773	0.774	0.788
SMO-Linear- 10^0	77.356	22.645	2.294	0.774	0.774	0.802	0.773	0.774	0.787
SMO-Linear- 10^1	76.926	23.074	2.331	0.769	0.769	0.791	0.769	0.769	0.780
SMO-Linear- 10^2	76.926	23.074	2.331	0.769	0.769	0.791	0.769	0.769	0.780
SMO-Linear- 10^3	76.926	23.074	2.331	0.769	0.769	0.791	0.769	0.769	0.780
SMO-Linear- 10^4	76.926	23.074	2.331	0.769	0.769	0.791	0.769	0.769	0.780
SMO-Linear- 10^5	76.926	23.074	2.331	0.769	0.769	0.791	0.769	0.769	0.780
SMO-Poly-0	3.035	96.965	1.090	0.030	0.030	0.001	0.025	0.030	0.001
SMO-Poly- 10^{-5}	4.495	95.505	2.384	0.045	0.045	0.039	0.038	0.045	0.038
SMO-Poly- 10^{-4}	5.498	94.502	3.055	0.055	0.055	0.127	0.057	0.055	0.079
SMO-Poly- 10^{-3}	23.305	76.695	1.881	0.233	0.233	0.586	0.233	0.233	0.333
SMO-Poly- 10^{-2}	28.172	71.828	2.470	0.282	0.282	0.626	0.287	0.282	0.394
SMO-Poly- 10^{-1}	27.199	72.801	3.481	0.272	0.272	0.605	0.279	0.272	0.382
SMO-Poly- 10^0	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-Poly- 10^1	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-Poly- 10^2	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-Poly- 10^3	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-Poly- 10^4	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-Poly- 10^5	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-RBF-0	3.035	96.965	1.090	0.030	0.030	0.001	0.025	0.030	0.001
SMO-RBF- 10^{-5}	2.004	97.996	0.512	0.020	0.020	0.003	0.017	0.020	0.005
SMO-RBF- 10^{-4}	2.004	97.996	0.512	0.020	0.020	0.003	0.017	0.020	0.005
SMO-RBF- 10^{-3}	2.004	97.996	0.512	0.020	0.020	0.003	0.017	0.020	0.005
SMO-RBF- 10^{-2}	4.179	96.336	2.074	0.042	0.042	0.020	0.044	0.037	0.021
SMO-RBF- 10^{-1}	3.664	96.336	1.377	0.037	0.037	0.015	0.040	0.037	0.021
SMO-RBF- 10^0	20.842	79.158	1.102	0.208	0.208	0.554	0.214	0.208	0.308
SMO-RBF- 10^1	28.314	71.687	1.092	0.283	0.283	0.661	0.289	0.283	0.402
SMO-RBF- 10^2	27.827	72.174	1.194	0.278	0.278	0.671	0.283	0.278	0.398
SMO-RBF- 10^3	27.827	72.174	1.194	0.278	0.278	0.671	0.283	0.278	0.398
SMO-RBF- 10^4	27.827	72.174	1.194	0.278	0.278	0.671	0.283	0.278	0.398
SMO-RBF- 10^5	27.827	72.174	1.194	0.278	0.278	0.671	0.283	0.278	0.398

Table 3.6: Classification Results for ACM collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	57.800	42.201	1.702	0.578	0.578	0.581	0.578	0.578	0.579
IB3	58.057	41.943	2.599	0.581	0.581	0.593	0.580	0.581	0.587
IB5	60.518	39.482	3.347	0.605	0.605	0.610	0.607	0.605	0.608
IB7	62.237	37.763	3.310	0.622	0.622	0.625	0.623	0.622	0.624
IB9	62.495	37.505	3.347	0.625	0.625	0.630	0.626	0.625	0.628
IB11	62.868	37.132	3.201	0.629	0.629	0.634	0.629	0.629	0.631
IB13	63.470	36.531	2.642	0.635	0.635	0.640	0.635	0.635	0.638
IB15	63.211	36.789	3.249	0.632	0.632	0.639	0.632	0.632	0.635
IB17	63.497	36.503	2.897	0.635	0.635	0.644	0.635	0.635	0.640
IB19	63.497	36.503	6.673	0.635	0.635	0.643	0.636	0.635	0.640
IB21	63.555	36.445	2.611	0.636	0.636	0.642	0.637	0.636	0.640
IB25	63.211	36.789	2.377	0.632	0.632	0.641	0.633	0.632	0.637
IB29	63.812	36.188	3.123	0.638	0.638	0.646	0.640	0.638	0.643
IB35	63.497	36.503	2.730	0.635	0.635	0.650	0.638	0.635	0.644
IB41	63.068	36.932	2.534	0.631	0.631	0.646	0.633	0.631	0.639
IB49	62.696	37.304	2.295	0.627	0.627	0.652	0.628	0.627	0.639
IB57	62.324	37.676	2.488	0.623	0.623	0.652	0.624	0.623	0.637
IB73	62.352	37.648	2.910	0.624	0.624	0.660	0.623	0.624	0.641
IB89	61.465	38.535	2.746	0.615	0.615	0.648	0.614	0.615	0.630
IB1-WV	57.800	42.201	1.702	0.578	0.578	0.581	0.579	0.578	0.580
IB3-WV	60.662	39.338	1.701	0.607	0.607	0.606	0.608	0.607	0.607
IB5-WV	62.580	37.420	2.692	0.626	0.626	0.629	0.628	0.626	0.628
IB7-WV	62.552	37.448	2.820	0.626	0.626	0.629	0.627	0.626	0.628
IB9-WV	63.325	36.675	2.911	0.633	0.633	0.640	0.634	0.633	0.637
IB11-WV	63.440	36.560	3.059	0.634	0.634	0.640	0.635	0.634	0.637
IB13-WV	64.070	35.930	2.813	0.641	0.641	0.649	0.641	0.641	0.645
IB15-WV	63.812	36.188	3.268	0.638	0.638	0.647	0.639	0.638	0.643
IB17-WV	63.898	36.102	2.977	0.639	0.639	0.650	0.640	0.639	0.645
IB19-WV	63.984	36.016	2.930	0.640	0.640	0.650	0.640	0.640	0.645
IB21-WV	64.356	35.644	2.780	0.644	0.644	0.651	0.644	0.644	0.648
IB25-WV	64.156	35.844	2.760	0.642	0.642	0.648	0.642	0.642	0.645
IB29-WV	64.156	35.844	3.312	0.642	0.642	0.651	0.643	0.642	0.647
IB35-WV	63.755	36.245	2.825	0.638	0.638	0.652	0.640	0.638	0.646
IB41-WV	63.469	36.531	2.770	0.635	0.635	0.649	0.636	0.635	0.642
IB49-WV	63.469	36.531	2.430	0.635	0.635	0.660	0.636	0.635	0.647
IB57-WV	63.183	36.817	2.882	0.632	0.632	0.659	0.633	0.632	0.646
IB73-WV	62.782	37.218	2.839	0.628	0.628	0.662	0.628	0.628	0.645
IB89-WV	61.923	38.077	2.865	0.619	0.619	0.651	0.620	0.619	0.635

3.3 Classic4 Collection

Table 3.7 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.8 presents the classification results obtained by IBk for Classic4 collection.

Table 3.7: Classification Results for Classic4 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	88.485	11.515	1.176	0.885	0.885	0.885	0.915	0.885	0.900
Multinomial NB	95.800	4.200	0.962	0.958	0.958	0.958	0.966	0.958	0.962
J48	90.360	9.641	0.728	0.904	0.904	0.907	0.892	0.904	0.899
SMO-Linear-0	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-Linear- 10^{-5}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-Linear- 10^{-4}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-Linear- 10^{-3}	45.215	54.785	1.877	0.452	0.452	0.363	0.251	0.452	0.297
SMO-Linear- 10^{-2}	75.406	24.594	1.289	0.754	0.754	0.899	0.664	0.754	0.764
SMO-Linear- 10^{-1}	93.192	6.808	1.018	0.932	0.932	0.955	0.915	0.932	0.935
SMO-Linear- 10^0	94.531	5.469	0.864	0.945	0.945	0.943	0.955	0.945	0.949
SMO-Linear- 10^1	93.897	6.103	0.855	0.939	0.939	0.949	0.933	0.939	0.941
SMO-Linear- 10^2	93.122	6.878	0.691	0.931	0.931	0.943	0.926	0.931	0.934
SMO-Linear- 10^3	93.051	6.949	0.623	0.931	0.931	0.942	0.925	0.931	0.933
SMO-Linear- 10^4	93.065	6.935	0.622	0.931	0.931	0.942	0.925	0.931	0.934
SMO-Linear- 10^5	93.051	6.949	0.623	0.931	0.931	0.942	0.925	0.931	0.933
SMO-Poly-0	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-Poly- 10^{-5}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-Poly- 10^{-4}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-Poly- 10^{-3}	45.723	54.278	1.903	0.457	0.457	0.364	0.257	0.457	0.301
SMO-Poly- 10^{-2}	60.987	39.013	1.593	0.610	0.610	0.863	0.466	0.610	0.605
SMO-Poly- 10^{-1}	76.942	23.058	1.429	0.769	0.769	0.886	0.686	0.769	0.774
SMO-Poly- 10^0	80.550	19.450	1.427	0.805	0.805	0.883	0.742	0.805	0.806
SMO-Poly- 10^1	81.311	18.689	1.287	0.813	0.813	0.881	0.754	0.813	0.812
SMO-Poly- 10^2	81.283	18.717	1.259	0.813	0.813	0.879	0.755	0.813	0.812
SMO-Poly- 10^3	81.297	18.703	1.298	0.813	0.813	0.879	0.755	0.813	0.812
SMO-Poly- 10^4	81.297	18.703	1.291	0.813	0.813	0.879	0.755	0.813	0.812
SMO-Poly- 10^5	81.311	18.689	1.296	0.813	0.813	0.879	0.755	0.813	0.812
SMO-RBF-0	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-RBF- 10^{-5}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-RBF- 10^{-4}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-RBF- 10^{-3}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-RBF- 10^{-2}	45.159	54.841	1.880	0.452	0.452	0.113	0.250	0.452	0.155
SMO-RBF- 10^{-1}	48.767	51.233	1.795	0.488	0.488	0.634	0.302	0.488	0.409
SMO-RBF- 10^0	84.567	15.433	0.975	0.846	0.846	0.884	0.811	0.846	0.846
SMO-RBF- 10^1	92.488	7.513	0.842	0.925	0.925	0.927	0.919	0.925	0.923
SMO-RBF- 10^2	92.911	7.090	0.809	0.929	0.929	0.926	0.929	0.929	0.927
SMO-RBF- 10^3	92.220	7.780	0.901	0.922	0.922	0.919	0.922	0.921	0.922
SMO-RBF- 10^4	91.938	8.062	0.988	0.919	0.919	0.917	0.919	0.919	0.918
SMO-RBF- 10^5	91.938	8.062	0.988	0.919	0.919	0.917	0.919	0.919	0.918

Table 3.8: Classification Results for Classic4 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	93.221	6.779	0.775	0.932	0.932	0.941	0.927	0.932	0.934
IB3	93.925	6.075	0.617	0.939	0.939	0.950	0.931	0.939	0.940
IB5	93.925	6.075	0.492	0.939	0.939	0.951	0.930	0.939	0.940
IB7	94.151	5.849	0.579	0.942	0.942	0.953	0.932	0.942	0.942
IB9	94.151	5.849	0.825	0.942	0.942	0.952	0.932	0.942	0.942
IB11	94.080	5.920	0.703	0.941	0.941	0.952	0.930	0.941	0.941
IB13	94.137	5.863	0.712	0.941	0.941	0.953	0.931	0.941	0.942
IB15	94.151	5.849	0.733	0.942	0.942	0.952	0.931	0.942	0.942
IB17	94.094	5.906	0.536	0.941	0.941	0.952	0.930	0.941	0.941
IB19	93.939	6.061	0.398	0.939	0.939	0.951	0.928	0.939	0.940
IB21	93.925	6.075	0.462	0.939	0.939	0.951	0.928	0.939	0.940
IB25	93.953	6.047	0.580	0.940	0.940	0.952	0.928	0.940	0.940
IB29	93.827	6.174	0.567	0.938	0.938	0.950	0.927	0.938	0.938
IB35	93.827	6.174	0.640	0.938	0.938	0.951	0.926	0.938	0.938
IB41	93.672	6.329	0.778	0.937	0.937	0.949	0.924	0.937	0.936
IB49	93.432	6.568	0.656	0.934	0.934	0.947	0.921	0.934	0.934
IB57	93.221	6.779	0.763	0.932	0.932	0.946	0.918	0.932	0.931
IB73	93.023	6.977	0.902	0.930	0.930	0.945	0.915	0.930	0.930
IB89	92.840	7.160	0.788	0.928	0.928	0.944	0.912	0.928	0.928
IB1-WV	93.221	6.779	0.775	0.932	0.932	0.941	0.927	0.932	0.934
IB3-WV	94.193	5.807	0.652	0.942	0.942	0.951	0.935	0.942	0.943
IB5-WV	94.151	5.849	0.559	0.942	0.942	0.952	0.934	0.942	0.943
IB7-WV	94.151	5.849	0.622	0.942	0.942	0.952	0.933	0.942	0.942
IB9-WV	94.235	5.765	0.768	0.942	0.942	0.952	0.933	0.942	0.943
IB11-WV	94.151	5.849	0.739	0.942	0.942	0.952	0.932	0.942	0.942
IB13-WV	94.249	5.751	0.740	0.942	0.942	0.953	0.933	0.942	0.943
IB15-WV	94.179	5.821	0.766	0.942	0.942	0.952	0.932	0.942	0.942
IB17-WV	94.193	5.807	0.663	0.942	0.942	0.952	0.932	0.942	0.942
IB19-WV	94.109	5.892	0.382	0.941	0.941	0.952	0.931	0.941	0.941
IB21-WV	94.052	5.948	0.432	0.941	0.941	0.952	0.930	0.941	0.941
IB25-WV	94.094	5.906	0.469	0.941	0.941	0.952	0.930	0.941	0.941
IB29-WV	93.939	6.061	0.485	0.939	0.939	0.951	0.928	0.939	0.939
IB35-WV	93.897	6.103	0.599	0.939	0.939	0.951	0.928	0.939	0.939
IB41-WV	93.784	6.216	0.793	0.938	0.938	0.950	0.926	0.938	0.938
IB49-WV	93.573	6.427	0.585	0.936	0.936	0.948	0.923	0.936	0.935
IB57-WV	93.404	6.596	0.569	0.934	0.934	0.947	0.921	0.934	0.934
IB73-WV	93.164	6.836	0.774	0.932	0.932	0.946	0.917	0.932	0.931
IB89-WV	93.080	6.920	0.814	0.931	0.931	0.946	0.916	0.931	0.930

3.4 CSTR Collection

Table 3.9 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.10 presents the classification results obtained by IBk for CSTR collection.

Table 3.9: Classification Results for CSTR collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	78.598	21.402	5.193	0.786	0.786	0.804	0.817	0.786	0.810
Multinomial NB	84.644	15.356	7.306	0.846	0.846	0.861	0.880	0.846	0.870
J48	66.851	33.150	7.591	0.669	0.669	0.652	0.646	0.669	0.649
SMO-Linear-0	15.391	84.609	6.197	0.154	0.154	0.038	0.250	0.154	0.067
SMO-Linear- 10^{-5}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-Linear- 10^{-4}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-Linear- 10^{-3}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-Linear- 10^{-2}	44.103	55.897	7.534	0.441	0.441	0.608	0.263	0.441	0.367
SMO-Linear- 10^{-1}	74.218	25.782	8.806	0.742	0.742	0.825	0.726	0.742	0.772
SMO-Linear- 10^0	74.920	25.081	8.964	0.749	0.749	0.813	0.739	0.749	0.774
SMO-Linear- 10^1	75.264	24.736	9.059	0.753	0.753	0.815	0.744	0.753	0.778
SMO-Linear- 10^2	75.264	24.736	9.059	0.753	0.753	0.815	0.744	0.753	0.778
SMO-Linear- 10^3	75.264	24.736	9.059	0.753	0.753	0.815	0.744	0.753	0.778
SMO-Linear- 10^4	75.264	24.736	9.059	0.753	0.753	0.815	0.744	0.753	0.778
SMO-Linear- 10^5	75.264	24.736	9.059	0.753	0.753	0.815	0.744	0.753	0.778
SMO-Poly-0	15.391	84.609	6.197	0.154	0.154	0.038	0.250	0.154	0.067
SMO-Poly- 10^{-5}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-Poly- 10^{-4}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-Poly- 10^{-3}	43.092	56.908	8.961	0.431	0.431	0.482	0.257	0.431	0.335
SMO-Poly- 10^{-2}	54.793	45.207	8.194	0.548	0.548	0.807	0.404	0.548	0.539
SMO-Poly- 10^{-1}	55.805	44.196	9.090	0.559	0.559	0.790	0.422	0.559	0.550
SMO-Poly- 10^0	56.471	43.529	9.006	0.565	0.565	0.774	0.439	0.565	0.560
SMO-Poly- 10^1	56.805	43.196	8.945	0.569	0.569	0.775	0.445	0.569	0.565
SMO-Poly- 10^2	56.805	43.196	8.945	0.569	0.569	0.775	0.445	0.569	0.565
SMO-Poly- 10^3	56.805	43.196	8.945	0.569	0.569	0.775	0.445	0.569	0.565
SMO-Poly- 10^4	56.805	43.196	8.945	0.569	0.569	0.775	0.445	0.569	0.565
SMO-Poly- 10^5	56.805	43.196	8.945	0.569	0.569	0.775	0.445	0.569	0.565
SMO-RBF-0	15.391	84.609	6.197	0.154	0.154	0.038	0.250	0.154	0.067
SMO-RBF- 10^{-5}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-RBF- 10^{-4}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-RBF- 10^{-3}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-RBF- 10^{-2}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-RBF- 10^{-1}	42.770	57.230	7.981	0.428	0.428	0.107	0.250	0.428	0.150
SMO-RBF- 10^0	48.793	51.207	7.874	0.488	0.488	0.529	0.308	0.488	0.389
SMO-RBF- 10^1	72.264	27.736	9.729	0.722	0.722	0.806	0.711	0.722	0.755
SMO-RBF- 10^2	72.276	27.724	9.486	0.722	0.722	0.799	0.702	0.722	0.747
SMO-RBF- 10^3	72.276	27.724	9.486	0.544	0.722	0.722	0.799	0.621	0.759
SMO-RBF- 10^4	72.276	27.724	9.486	0.545	0.722	0.722	0.799	0.621	0.759
SMO-RBF- 10^5	72.276	27.724	9.486	0.543	0.722	0.722	0.799	0.620	0.759

Table 3.10: Classification Results for CSTR collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	78.609	21.391	9.064	0.786	0.786	0.782	0.797	0.786	0.790
IB3	80.632	19.368	6.915	0.806	0.806	0.810	0.810	0.806	0.810
IB5	79.621	20.379	7.278	0.796	0.796	0.800	0.794	0.796	0.797
IB7	79.609	20.391	5.632	0.796	0.796	0.820	0.793	0.796	0.806
IB9	82.299	17.701	5.929	0.823	0.823	0.841	0.827	0.823	0.834
IB11	80.954	19.046	5.930	0.809	0.809	0.841	0.806	0.809	0.823
IB13	80.299	19.701	4.750	0.803	0.803	0.835	0.783	0.803	0.808
IB15	80.621	19.379	5.068	0.806	0.806	0.834	0.788	0.806	0.810
IB17	81.954	18.046	5.776	0.819	0.819	0.852	0.806	0.819	0.828
IB19	81.287	18.713	5.575	0.813	0.813	0.847	0.790	0.813	0.817
IB21	79.264	20.736	3.583	0.793	0.793	0.833	0.738	0.793	0.782
IB25	81.287	18.713	5.575	0.813	0.813	0.853	0.771	0.813	0.810
IB29	80.943	19.058	4.201	0.809	0.809	0.845	0.757	0.809	0.799
IB35	79.264	20.736	5.116	0.793	0.793	0.835	0.732	0.793	0.780
IB41	77.586	22.414	5.802	0.776	0.776	0.858	0.690	0.776	0.765
IB49	72.575	27.425	5.922	0.726	0.726	0.840	0.594	0.726	0.696
IB57	71.253	28.747	7.723	0.712	0.712	0.842	0.552	0.712	0.667
IB73	68.218	31.782	8.363	0.682	0.682	0.580	0.507	0.682	0.541
IB89	63.874	36.127	8.292	0.639	0.639	0.565	0.446	0.639	0.498
IB1-WV	78.609	21.391	9.064	0.786	0.786	0.782	0.797	0.786	0.790
IB3-WV	79.621	20.379	7.329	0.796	0.796	0.789	0.800	0.796	0.794
IB5-WV	78.943	21.058	7.278	0.789	0.789	0.798	0.789	0.789	0.793
IB7-WV	78.943	21.058	5.342	0.789	0.789	0.810	0.796	0.789	0.803
IB9-WV	80.954	19.046	5.543	0.809	0.809	0.824	0.817	0.809	0.821
IB11-WV	80.954	19.046	4.428	0.809	0.809	0.845	0.821	0.809	0.833
IB13-WV	79.287	20.713	4.347	0.793	0.793	0.822	0.799	0.793	0.810
IB15-WV	80.621	19.379	5.688	0.806	0.806	0.837	0.803	0.806	0.820
IB17-WV	80.954	19.046	5.930	0.809	0.809	0.848	0.814	0.809	0.830
IB19-WV	79.621	20.379	6.354	0.796	0.796	0.839	0.793	0.796	0.816
IB21-WV	78.943	21.058	5.933	0.789	0.789	0.837	0.774	0.789	0.804
IB25-WV	78.943	21.058	5.933	0.789	0.789	0.838	0.753	0.789	0.793
IB29-WV	79.943	20.058	4.908	0.799	0.799	0.843	0.772	0.799	0.806
IB35-WV	79.276	20.724	5.079	0.793	0.793	0.837	0.753	0.793	0.793
IB41-WV	78.943	21.058	4.676	0.789	0.789	0.848	0.743	0.789	0.792
IB49-WV	73.575	26.425	5.273	0.736	0.736	0.842	0.638	0.736	0.726
IB57-WV	73.931	26.069	6.574	0.739	0.739	0.852	0.620	0.739	0.718
IB73-WV	68.218	31.782	6.908	0.682	0.682	0.577	0.522	0.682	0.548
IB89-WV	64.207	35.793	7.779	0.642	0.642	0.563	0.463	0.642	0.508

3.5 Dmoz-Business-500 Collection

Table 3.11 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.12 presents the classification results obtained by IBk for Dmoz-Business-500 collection.

Table 3.11: Classification Results for Dmoz-Business-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	58.605	41.395	0.693	0.586	0.586	0.599	0.586	0.586	0.593
Multinomial NB	68.276	31.724	0.711	0.683	0.683	0.693	0.683	0.683	0.688
J48	53.103	46.897	0.707	0.531	0.531	0.532	0.531	0.531	0.532
SMO-Linear-0	2.703	97.297	0.461	0.027	0.027	0.001	0.027	0.027	0.001
SMO-Linear- 10^{-5}	2.568	97.433	0.962	0.026	0.026	0.043	0.026	0.026	0.032
SMO-Linear- 10^{-4}	2.568	97.433	0.962	0.026	0.026	0.043	0.026	0.026	0.032
SMO-Linear- 10^{-3}	2.211	97.789	0.671	0.022	0.022	0.022	0.022	0.022	0.022
SMO-Linear- 10^{-2}	17.254	82.746	4.532	0.173	0.173	0.403	0.173	0.173	0.242
SMO-Linear- 10^{-1}	64.351	35.649	0.939	0.644	0.644	0.676	0.644	0.644	0.659
SMO-Linear- 10^0	65.308	34.692	0.901	0.653	0.653	0.664	0.653	0.653	0.658
SMO-Linear- 10^1	61.854	38.146	1.065	0.619	0.619	0.629	0.619	0.619	0.624
SMO-Linear- 10^2	60.708	39.292	1.204	0.607	0.607	0.617	0.607	0.607	0.612
SMO-Linear- 10^3	60.708	39.292	1.204	0.607	0.607	0.617	0.607	0.607	0.612
SMO-Linear- 10^4	60.708	39.292	1.204	0.607	0.607	0.617	0.607	0.607	0.612
SMO-Linear- 10^5	60.708	39.292	1.204	0.607	0.607	0.617	0.607	0.607	0.612
SMO-Poly-0	2.076	97.924	0.602	0.021	0.021	0.036	0.021	0.021	0.026
SMO-Poly- 10^{-5}	2.076	97.924	0.602	0.021	0.021	0.036	0.021	0.021	0.026
SMO-Poly- 10^{-4}	3.281	96.719	1.797	0.033	0.033	0.180	0.033	0.033	0.056
SMO-Poly- 10^{-3}	2.016	97.984	0.313	0.020	0.020	0.042	0.020	0.020	0.027
SMO-Poly- 10^{-2}	13.092	86.908	3.016	0.131	0.131	0.467	0.131	0.131	0.205
SMO-Poly- 10^{-1}	55.476	44.524	0.737	0.555	0.555	0.659	0.555	0.555	0.602
SMO-Poly- 10^0	56.941	43.060	0.701	0.569	0.569	0.621	0.569	0.569	0.594
SMO-Poly- 10^1	54.876	45.124	0.912	0.549	0.549	0.590	0.549	0.549	0.569
SMO-Poly- 10^2	50.514	49.487	0.978	0.505	0.505	0.560	0.505	0.505	0.531
SMO-Poly- 10^3	49.232	50.768	0.961	0.492	0.492	0.566	0.492	0.492	0.527
SMO-Poly- 10^4	49.232	50.768	0.961	0.492	0.492	0.566	0.492	0.492	0.527
SMO-Poly- 10^5	49.232	50.768	0.961	0.492	0.492	0.566	0.492	0.492	0.527
SMO-RBF-0	2.703	97.297	0.461	0.027	0.027	0.001	0.027	0.027	0.001
SMO-RBF- 10^{-5}	2.616	97.384	1.041	0.026	0.026	0.034	0.026	0.026	0.030
SMO-RBF- 10^{-4}	2.616	97.384	1.041	0.026	0.026	0.004	0.036	0.026	0.007
SMO-RBF- 10^{-3}	2.616	97.384	1.041	0.026	0.026	0.004	0.036	0.026	0.007
SMO-RBF- 10^{-2}	2.373	97.627	0.823	0.024	0.024	0.003	0.033	0.024	0.005
SMO-RBF- 10^{-1}	55.476	44.524	0.737	0.555	0.555	0.659	0.555	0.555	0.602
SMO-RBF- 10^0	45.951	54.049	3.481	0.460	0.460	0.600	0.460	0.460	0.520
SMO-RBF- 10^1	65.843	34.157	0.789	0.658	0.658	0.678	0.659	0.658	0.668
SMO-RBF- 10^2	64.243	35.757	0.733	0.642	0.642	0.651	0.643	0.642	0.647
SMO-RBF- 10^3	61.465	38.535	1.319	0.615	0.615	0.623	0.615	0.615	0.619
SMO-RBF- 10^4	61.054	38.946	1.265	0.611	0.611	0.618	0.611	0.611	0.614
SMO-RBF- 10^5	61.054	38.946	1.265	0.611	0.611	0.618	0.611	0.611	0.614

Table 3.12: Classification Results for Dmoz-Business-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	54.422	45.578	1.384	0.544	0.544	0.553	0.544	0.544	0.548
IB3	55.492	44.508	1.024	0.555	0.555	0.600	0.555	0.555	0.577
IB5	58.719	41.281	0.979	0.587	0.587	0.615	0.587	0.587	0.600
IB7	59.595	40.406	0.973	0.596	0.596	0.617	0.596	0.596	0.606
IB9	59.795	40.206	1.122	0.598	0.598	0.620	0.598	0.598	0.609
IB11	60.157	39.843	1.132	0.602	0.602	0.625	0.602	0.602	0.613
IB13	60.108	39.892	1.123	0.601	0.601	0.623	0.601	0.601	0.612
IB15	60.108	39.892	0.950	0.601	0.601	0.625	0.601	0.601	0.613
IB17	60.341	39.660	0.873	0.603	0.603	0.629	0.603	0.603	0.616
IB19	60.086	39.914	0.836	0.601	0.601	0.626	0.601	0.601	0.613
IB21	60.119	39.881	0.865	0.601	0.601	0.627	0.601	0.601	0.614
IB25	59.811	40.189	0.670	0.598	0.598	0.625	0.598	0.598	0.611
IB29	59.686	40.314	0.682	0.597	0.597	0.625	0.597	0.597	0.611
IB35	59.195	40.806	0.773	0.592	0.592	0.621	0.592	0.592	0.606
IB41	58.762	41.238	0.723	0.588	0.588	0.619	0.588	0.588	0.603
IB49	58.497	41.503	0.792	0.585	0.585	0.617	0.585	0.585	0.601
IB57	58.141	41.860	0.755	0.581	0.581	0.614	0.581	0.581	0.597
IB73	57.346	42.654	0.924	0.573	0.573	0.612	0.573	0.573	0.592
IB89	56.643	43.357	0.865	0.566	0.566	0.607	0.566	0.566	0.586
IB1-WV	54.422	45.578	1.384	0.544	0.544	0.553	0.544	0.544	0.548
IB3-WV	57.573	42.427	1.292	0.576	0.576	0.583	0.576	0.576	0.580
IB5-WV	60.173	39.827	0.977	0.602	0.602	0.612	0.602	0.602	0.607
IB7-WV	60.941	39.060	1.067	0.609	0.609	0.622	0.609	0.609	0.615
IB9-WV	61.151	38.849	1.080	0.612	0.612	0.625	0.612	0.612	0.618
IB11-WV	61.449	38.551	1.273	0.614	0.614	0.630	0.614	0.614	0.622
IB13-WV	61.346	38.654	1.215	0.613	0.613	0.630	0.613	0.613	0.622
IB15-WV	61.238	38.762	1.145	0.612	0.612	0.631	0.612	0.612	0.621
IB17-WV	61.486	38.514	1.035	0.615	0.615	0.635	0.615	0.615	0.625
IB19-WV	61.335	38.665	1.179	0.613	0.613	0.635	0.613	0.613	0.624
IB21-WV	61.238	38.762	1.098	0.612	0.612	0.634	0.612	0.612	0.623
IB25-WV	60.730	39.270	0.718	0.607	0.607	0.632	0.607	0.607	0.619
IB29-WV	60.508	39.492	0.691	0.605	0.605	0.631	0.605	0.605	0.618
IB35-WV	60.049	39.951	0.815	0.600	0.600	0.628	0.600	0.600	0.614
IB41-WV	59.654	40.346	0.860	0.597	0.597	0.626	0.597	0.597	0.611
IB49-WV	59.346	40.654	0.813	0.593	0.593	0.624	0.593	0.593	0.608
IB57-WV	59.081	40.919	0.843	0.591	0.591	0.622	0.591	0.591	0.606
IB73-WV	58.314	41.687	0.966	0.583	0.583	0.620	0.583	0.583	0.601
IB89-WV	57.535	42.465	0.960	0.575	0.575	0.614	0.575	0.575	0.594

3.6 Dmoz-Computers-500 Collection

Table 3.13 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.14 presents the classification results obtained by IBk for Dmoz-Computers-500 collection.

Table 3.13: Classification Results for Dmoz-Computers-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	59.674	40.326	1.924	0.597	0.597	0.616	0.597	0.597	0.606
Multinomial NB	70.137	29.863	1.486	0.701	0.701	0.701	0.701	0.701	0.701
J48	54.884	45.116	1.713	0.549	0.549	0.541	0.549	0.549	0.545
SMO-Linear-0	5.263	94.737	0.796	0.053	0.053	0.003	0.053	0.053	0.005
SMO-Linear- 10^{-5}	4.547	95.453	0.812	0.045	0.045	0.051	0.045	0.045	0.048
SMO-Linear- 10^{-4}	4.547	95.453	0.812	0.045	0.045	0.051	0.045	0.045	0.048
SMO-Linear- 10^{-3}	4.368	95.632	0.625	0.044	0.044	0.050	0.044	0.044	0.047
SMO-Linear- 10^{-2}	19.189	80.811	4.900	0.192	0.192	0.442	0.192	0.192	0.268
SMO-Linear- 10^{-1}	63.505	36.495	1.506	0.635	0.635	0.684	0.635	0.635	0.659
SMO-Linear- 10^0	66.011	33.990	1.840	0.660	0.660	0.669	0.660	0.660	0.664
SMO-Linear- 10^1	63.326	36.674	1.863	0.633	0.633	0.637	0.633	0.633	0.635
SMO-Linear- 10^2	60.821	39.179	1.396	0.608	0.608	0.612	0.608	0.608	0.610
SMO-Linear- 10^3	60.632	39.369	1.404	0.606	0.606	0.610	0.606	0.606	0.608
SMO-Linear- 10^4	60.632	39.369	1.404	0.606	0.606	0.610	0.606	0.606	0.608
SMO-Linear- 10^5	60.632	39.369	1.404	0.606	0.606	0.610	0.606	0.606	0.608
SMO-Poly-0	5.263	94.737	0.796	0.053	0.053	0.003	0.053	0.053	0.005
SMO-Poly- 10^{-5}	4.284	95.716	0.380	0.043	0.043	0.056	0.043	0.043	0.048
SMO-Poly- 10^{-4}	9.000	91.000	3.173	0.090	0.090	0.262	0.090	0.090	0.134
SMO-Poly- 10^{-3}	4.947	95.053	1.147	0.049	0.049	0.081	0.049	0.049	0.061
SMO-Poly- 10^{-2}	20.253	79.747	4.880	0.203	0.203	0.429	0.203	0.203	0.275
SMO-Poly- 10^{-1}	55.832	44.169	2.000	0.558	0.558	0.635	0.558	0.558	0.594
SMO-Poly- 10^0	55.832	44.169	2.268	0.558	0.558	0.635	0.558	0.558	0.594
SMO-Poly- 10^1	54.316	45.684	2.049	0.543	0.543	0.584	0.543	0.543	0.563
SMO-Poly- 10^2	50.474	49.526	1.779	0.505	0.505	0.563	0.505	0.505	0.532
SMO-Poly- 10^3	48.495	51.505	1.451	0.485	0.485	0.547	0.485	0.485	0.514
SMO-Poly- 10^4	48.242	51.758	1.388	0.482	0.482	0.544	0.482	0.482	0.511
SMO-Poly- 10^5	48.232	51.769	1.399	0.482	0.482	0.544	0.482	0.482	0.511
SMO-RBF-0	5.263	94.737	0.796	0.053	0.053	0.003	0.053	0.053	0.005
SMO-RBF- 10^{-5}	4.547	95.453	0.814	0.045	0.045	0.051	0.045	0.045	0.048
SMO-RBF- 10^{-4}	4.547	95.453	0.814	0.045	0.045	0.051	0.045	0.045	0.048
SMO-RBF- 10^{-3}	4.547	95.453	0.814	0.045	0.045	0.051	0.045	0.045	0.048
SMO-RBF- 10^{-2}	4.547	95.453	0.814	0.045	0.045	0.051	0.045	0.045	0.048
SMO-RBF- 10^{-1}	5.453	94.547	2.313	0.055	0.055	0.055	0.055	0.055	0.055
SMO-RBF- 10^0	43.737	56.263	4.080	0.437	0.437	0.619	0.437	0.437	0.512
SMO-RBF- 10^1	54.316	45.684	1.975	0.543	0.543	0.584	0.543	0.543	0.563
SMO-RBF- 10^2	50.474	49.526	2.057	0.505	0.505	0.563	0.505	0.505	0.532
SMO-RBF- 10^3	62.653	37.347	1.814	0.627	0.627	0.636	0.627	0.627	0.631
SMO-RBF- 10^4	60.695	39.305	1.701	0.607	0.607	0.617	0.607	0.607	0.612
SMO-RBF- 10^5	60.695	39.305	1.701	0.607	0.607	0.617	0.607	0.607	0.612

Table 3.14: Classification Results for Dmoz-Computers-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	58.611	41.390	1.800	0.586	0.586	0.582	0.586	0.586	0.584
IB3	59.368	40.632	2.066	0.594	0.594	0.611	0.594	0.594	0.602
IB5	61.379	38.621	1.877	0.614	0.614	0.623	0.614	0.614	0.618
IB7	62.189	37.811	2.019	0.622	0.622	0.626	0.622	0.622	0.624
IB9	62.547	37.453	2.062	0.625	0.625	0.633	0.625	0.625	0.629
IB11	62.579	37.421	1.938	0.626	0.626	0.635	0.626	0.626	0.630
IB13	62.589	37.411	2.156	0.626	0.626	0.637	0.626	0.626	0.631
IB15	62.274	37.726	2.295	0.623	0.623	0.634	0.623	0.623	0.628
IB17	62.158	37.842	2.254	0.622	0.622	0.633	0.622	0.622	0.627
IB19	62.074	37.926	2.255	0.621	0.621	0.632	0.621	0.621	0.626
IB21	62.063	37.937	2.255	0.621	0.621	0.632	0.621	0.621	0.626
IB25	61.642	38.358	2.210	0.616	0.616	0.629	0.616	0.616	0.622
IB29	61.316	38.684	1.914	0.613	0.613	0.627	0.613	0.613	0.620
IB35	60.905	39.095	2.323	0.609	0.609	0.625	0.609	0.609	0.617
IB41	60.579	39.421	2.166	0.606	0.606	0.623	0.606	0.606	0.614
IB49	59.811	40.190	2.365	0.598	0.598	0.618	0.598	0.598	0.608
IB57	59.126	40.874	2.016	0.591	0.591	0.611	0.591	0.591	0.601
IB73	58.442	41.558	2.211	0.584	0.584	0.604	0.584	0.584	0.594
IB89	57.400	42.600	2.100	0.574	0.574	0.596	0.574	0.574	0.585
IB1-WV	58.600	41.400	1.804	0.586	0.586	0.582	0.586	0.586	0.584
IB3-WV	61.326	38.674	2.068	0.613	0.613	0.608	0.613	0.613	0.610
IB5-WV	62.642	37.358	1.673	0.626	0.626	0.623	0.626	0.626	0.625
IB7-WV	63.558	36.442	1.716	0.636	0.636	0.632	0.636	0.636	0.634
IB9-WV	63.705	36.295	2.128	0.637	0.637	0.637	0.637	0.637	0.637
IB11-WV	63.453	36.547	2.119	0.635	0.635	0.638	0.635	0.635	0.636
IB13-WV	63.568	36.432	1.934	0.636	0.636	0.641	0.636	0.636	0.638
IB15-WV	63.516	36.484	2.090	0.635	0.635	0.641	0.635	0.635	0.638
IB17-WV	63.358	36.642	2.075	0.634	0.634	0.640	0.634	0.634	0.637
IB19-WV	63.200	36.800	2.294	0.632	0.632	0.639	0.632	0.632	0.635
IB21-WV	63.053	36.947	2.004	0.631	0.631	0.637	0.631	0.631	0.634
IB25-WV	62.600	37.400	2.092	0.626	0.626	0.635	0.626	0.626	0.630
IB29-WV	62.337	37.663	1.952	0.623	0.623	0.635	0.623	0.623	0.629
IB35-WV	61.989	38.011	2.015	0.620	0.620	0.635	0.620	0.620	0.627
IB41-WV	61.432	38.569	2.019	0.614	0.614	0.630	0.614	0.614	0.622
IB49-WV	60.937	39.063	2.309	0.609	0.609	0.629	0.609	0.609	0.619
IB57-WV	60.337	39.663	2.110	0.603	0.603	0.622	0.603	0.603	0.612
IB73-WV	59.432	40.569	2.145	0.594	0.594	0.613	0.594	0.594	0.603
IB89-WV	58.495	41.505	2.249	0.585	0.585	0.606	0.585	0.585	0.595

3.7 Dmoz-Health-500 Collection

Table 3.15 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.16 presents the classification results obtained by IBk for Dmoz-Health-500 collection.

Table 3.15: Classification Results for Dmoz-Health-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	73.154	26.846	1.679	0.732	0.732	0.748	0.732	0.732	0.740
Multinomial NB	82.077	17.923	1.404	0.821	0.821	0.826	0.821	0.821	0.823
J48	73.615	26.385	2.235	0.736	0.736	0.742	0.736	0.736	0.739
SMO-Linear-0	7.692	92.308	0.921	0.077	0.077	0.006	0.077	0.077	0.011
SMO-Linear- 10^{-5}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-Linear- 10^{-4}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-Linear- 10^{-3}	5.877	94.123	0.721	0.059	0.059	0.042	0.059	0.059	0.049
SMO-Linear- 10^{-2}	42.138	57.862	2.700	0.421	0.421	0.670	0.421	0.421	0.517
SMO-Linear- 10^{-1}	78.400	21.600	1.369	0.784	0.784	0.815	0.784	0.784	0.799
SMO-Linear- 10^0	79.492	20.508	1.851	0.795	0.795	0.803	0.795	0.795	0.799
SMO-Linear- 10^1	77.877	22.123	1.991	0.779	0.779	0.786	0.779	0.779	0.783
SMO-Linear- 10^2	77.723	22.277	2.114	0.777	0.777	0.785	0.777	0.777	0.781
SMO-Linear- 10^3	77.723	22.277	2.114	0.777	0.777	0.785	0.777	0.777	0.781
SMO-Linear- 10^4	77.723	22.277	2.114	0.777	0.777	0.785	0.777	0.777	0.781
SMO-Linear- 10^5	77.723	22.277	2.114	0.777	0.777	0.785	0.777	0.777	0.781
SMO-Poly-0	7.692	92.308	0.921	0.077	0.077	0.006	0.077	0.077	0.011
SMO-Poly- 10^{-5}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-Poly- 10^{-4}	8.138	91.862	3.423	0.081	0.081	0.236	0.081	0.081	0.121
SMO-Poly- 10^{-3}	7.892	92.108	1.548	0.079	0.079	0.190	0.079	0.079	0.111
SMO-Poly- 10^{-2}	44.046	55.954	2.430	0.440	0.440	0.743	0.440	0.440	0.553
SMO-Poly- 10^{-1}	69.523	30.477	2.323	0.695	0.695	0.762	0.695	0.695	0.727
SMO-Poly- 10^0	69.492	30.508	1.798	0.695	0.695	0.736	0.695	0.695	0.715
SMO-Poly- 10^1	66.631	33.369	1.932	0.666	0.666	0.734	0.666	0.666	0.698
SMO-Poly- 10^2	66.400	33.600	1.862	0.664	0.664	0.731	0.664	0.664	0.696
SMO-Poly- 10^3	66.400	33.600	1.862	0.664	0.664	0.731	0.664	0.664	0.696
SMO-Poly- 10^4	66.400	33.600	1.862	0.664	0.664	0.731	0.664	0.664	0.696
SMO-Poly- 10^5	66.400	33.600	1.862	0.664	0.664	0.731	0.664	0.664	0.696
SMO-RBF-0	7.692	92.308	0.921	0.077	0.077	0.006	0.077	0.077	0.011
SMO-RBF- 10^{-5}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-RBF- 10^{-4}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-RBF- 10^{-3}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-RBF- 10^{-2}	5.769	94.231	0.547	0.058	0.058	0.039	0.058	0.058	0.047
SMO-RBF- 10^{-1}	6.231	93.769	1.604	0.062	0.062	0.042	0.062	0.062	0.050
SMO-RBF- 10^0	61.292	38.708	1.662	0.613	0.613	0.756	0.613	0.613	0.677
SMO-RBF- 10^1	79.908	20.092	1.295	0.799	0.799	0.816	0.799	0.799	0.807
SMO-RBF- 10^2	78.831	21.169	1.748	0.788	0.788	0.797	0.788	0.788	0.793
SMO-RBF- 10^3	77.985	22.015	1.419	0.780	0.780	0.789	0.780	0.780	0.784
SMO-RBF- 10^4	77.954	22.046	1.403	0.780	0.780	0.789	0.780	0.780	0.784
SMO-RBF- 10^5	77.954	22.046	1.403	0.780	0.780	0.789	0.780	0.780	0.784

Table 3.16: Classification Results for Dmoz-Health-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	73.908	26.092	1.402	0.739	0.739	0.737	0.739	0.739	0.738
IB3	74.769	25.231	1.215	0.748	0.748	0.762	0.748	0.748	0.755
IB5	76.492	23.508	1.197	0.765	0.765	0.772	0.765	0.765	0.768
IB7	76.723	23.277	1.112	0.767	0.767	0.774	0.767	0.767	0.771
IB9	76.385	23.615	1.387	0.764	0.764	0.771	0.764	0.764	0.768
IB11	76.231	23.769	1.394	0.762	0.762	0.769	0.762	0.762	0.766
IB13	76.431	23.569	1.358	0.764	0.764	0.772	0.764	0.764	0.768
IB15	76.338	23.662	1.369	0.763	0.763	0.770	0.763	0.763	0.767
IB17	75.846	24.154	1.515	0.758	0.758	0.765	0.758	0.758	0.762
IB19	75.646	24.354	1.551	0.756	0.756	0.764	0.756	0.756	0.760
IB21	75.677	24.323	1.719	0.757	0.757	0.764	0.757	0.757	0.760
IB25	75.200	24.800	1.779	0.752	0.752	0.758	0.752	0.752	0.755
IB29	74.985	25.015	1.877	0.750	0.750	0.757	0.750	0.750	0.753
IB35	74.400	25.600	2.131	0.744	0.744	0.751	0.744	0.744	0.748
IB41	74.015	25.985	1.768	0.740	0.740	0.749	0.740	0.740	0.744
IB49	73.585	26.415	1.320	0.736	0.736	0.745	0.736	0.736	0.740
IB57	72.831	27.169	1.273	0.728	0.728	0.738	0.728	0.728	0.733
IB73	71.938	28.062	1.724	0.719	0.719	0.729	0.719	0.719	0.724
IB89	70.938	29.062	1.801	0.709	0.709	0.720	0.709	0.709	0.715
IB1-WV	73.908	26.092	1.402	0.739	0.739	0.737	0.739	0.739	0.738
IB3-WV	76.292	23.708	1.663	0.763	0.763	0.761	0.763	0.763	0.762
IB5-WV	77.554	22.446	1.078	0.776	0.776	0.776	0.776	0.776	0.776
IB7-WV	77.831	22.169	1.401	0.778	0.778	0.780	0.778	0.778	0.779
IB9-WV	77.508	22.492	1.493	0.775	0.775	0.777	0.775	0.775	0.776
IB11-WV	77.046	22.954	1.314	0.770	0.770	0.773	0.770	0.770	0.772
IB13-WV	77.185	22.815	1.265	0.772	0.772	0.776	0.772	0.772	0.774
IB15-WV	76.892	23.108	1.502	0.769	0.769	0.773	0.769	0.769	0.771
IB17-WV	76.538	23.462	1.476	0.765	0.765	0.770	0.765	0.765	0.767
IB19-WV	76.369	23.631	1.474	0.764	0.764	0.768	0.764	0.764	0.766
IB21-WV	76.277	23.723	1.515	0.763	0.763	0.767	0.763	0.763	0.765
IB25-WV	75.892	24.108	1.757	0.759	0.759	0.764	0.759	0.759	0.761
IB29-WV	75.523	24.477	1.551	0.755	0.755	0.761	0.755	0.755	0.758
IB35-WV	75.138	24.862	1.963	0.751	0.751	0.757	0.751	0.751	0.754
IB41-WV	74.769	25.231	1.851	0.748	0.748	0.754	0.748	0.748	0.751
IB49-WV	74.246	25.754	1.291	0.742	0.742	0.749	0.742	0.742	0.746
IB57-WV	73.569	26.431	1.509	0.736	0.736	0.744	0.736	0.736	0.740
IB73-WV	72.692	27.308	1.694	0.727	0.727	0.735	0.727	0.727	0.731
IB89-WV	71.785	28.215	1.769	0.718	0.718	0.728	0.718	0.718	0.723

3.8 Dmoz-Science-500 Collection

Table 3.17 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.18 presents the classification results obtained by IBk for Dmoz-Science-500 collection.

Table 3.17: Classification Results for Dmoz-Science-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	62.117	37.883	1.978	0.621	0.621	0.630	0.621	0.621	0.625
Multinomial NB	73.817	26.183	1.129	0.738	0.738	0.741	0.738	0.738	0.740
J48	57.200	42.800	2.009	0.572	0.572	0.577	0.572	0.572	0.574
SMO-Linear-0	8.333	91.667	1.022	0.083	0.083	0.007	0.083	0.083	0.013
SMO-Linear- 10^{-5}	7.733	92.267	3.518	0.077	0.077	0.085	0.077	0.077	0.081
SMO-Linear- 10^{-4}	7.733	92.267	3.518	0.077	0.077	0.085	0.077	0.077	0.081
SMO-Linear- 10^{-3}	7.183	92.817	1.902	0.072	0.072	0.074	0.072	0.072	0.073
SMO-Linear- 10^{-2}	17.417	82.583	5.050	0.174	0.174	0.322	0.174	0.174	0.226
SMO-Linear- 10^{-1}	66.350	33.650	1.298	0.664	0.664	0.676	0.664	0.664	0.670
SMO-Linear- 10^0	65.517	34.483	1.488	0.655	0.655	0.657	0.655	0.655	0.656
SMO-Linear- 10^1	63.650	36.350	2.239	0.637	0.637	0.640	0.637	0.637	0.638
SMO-Linear- 10^2	63.017	36.983	2.376	0.630	0.630	0.633	0.630	0.630	0.632
SMO-Linear- 10^3	63.017	36.983	2.376	0.630	0.630	0.633	0.630	0.630	0.632
SMO-Linear- 10^4	63.017	36.983	2.376	0.630	0.630	0.633	0.630	0.630	0.632
SMO-Linear- 10^5	63.017	36.983	2.376	0.630	0.630	0.633	0.630	0.630	0.632
SMO-Poly-0	8.333	91.667	1.022	0.083	0.083	0.007	0.083	0.083	0.013
SMO-Poly- 10^{-5}	6.933	93.067	1.200	0.069	0.069	0.072	0.069	0.069	0.071
SMO-Poly- 10^{-4}	14.500	85.500	4.692	0.145	0.145	0.207	0.145	0.145	0.170
SMO-Poly- 10^{-3}	7.900	92.100	2.388	0.079	0.079	0.222	0.079	0.079	0.116
SMO-Poly- 10^{-2}	30.300	69.700	7.872	0.303	0.303	0.459	0.303	0.303	0.365
SMO-Poly- 10^{-1}	53.133	46.867	1.651	0.531	0.531	0.613	0.531	0.531	0.569
SMO-Poly- 10^0	49.883	50.117	1.317	0.499	0.499	0.570	0.499	0.499	0.532
SMO-Poly- 10^1	43.950	56.050	1.590	0.440	0.440	0.561	0.440	0.440	0.493
SMO-Poly- 10^2	43.433	56.567	1.272	0.434	0.434	0.537	0.434	0.434	0.480
SMO-Poly- 10^3	43.483	56.517	1.326	0.434	0.434	0.537	0.434	0.434	0.480
SMO-Poly- 10^4	43.483	56.517	1.326	0.434	0.434	0.537	0.434	0.434	0.480
SMO-Poly- 10^5	43.483	56.517	1.326	0.434	0.434	0.537	0.434	0.434	0.480
SMO-RBF-0	8.333	91.667	1.022	0.083	0.083	0.007	0.083	0.083	0.013
SMO-RBF- 10^{-5}	7.667	92.333	3.320	0.077	0.077	0.080	0.077	0.077	0.078
SMO-RBF- 10^{-4}	7.667	92.333	3.320	0.077	0.077	0.080	0.077	0.077	0.078
SMO-RBF- 10^{-3}	7.667	92.333	3.320	0.077	0.077	0.080	0.077	0.077	0.078
SMO-RBF- 10^{-2}	8.167	91.833	3.597	0.082	0.082	0.098	0.082	0.082	0.089
SMO-RBF- 10^{-1}	9.200	90.800	3.215	0.092	0.092	0.158	0.092	0.092	0.116
SMO-RBF- 10^0	44.267	55.733	6.228	0.443	0.443	0.560	0.443	0.443	0.494
SMO-RBF- 10^1	67.383	32.617	1.028	0.674	0.674	0.681	0.674	0.674	0.677
SMO-RBF- 10^2	64.767	35.233	2.014	0.648	0.648	0.653	0.648	0.648	0.650
SMO-RBF- 10^3	63.500	36.500	2.187	0.635	0.635	0.641	0.635	0.635	0.638
SMO-RBF- 10^4	63.450	36.550	2.199	0.635	0.635	0.640	0.635	0.635	0.637
SMO-RBF- 10^5	63.450	36.550	2.199	0.635	0.635	0.640	0.635	0.635	0.637

Table 3.18: Classification Results for Dmoz-Science-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	60.450	39.550	1.665	0.605	0.605	0.602	0.605	0.605	0.603
IB3	61.400	38.600	1.778	0.614	0.614	0.633	0.614	0.614	0.624
IB5	62.867	37.133	1.622	0.629	0.629	0.636	0.629	0.629	0.632
IB7	62.850	37.150	1.654	0.629	0.629	0.635	0.629	0.629	0.632
IB9	62.850	37.150	1.720	0.629	0.629	0.634	0.629	0.629	0.631
IB11	62.850	37.150	1.301	0.629	0.629	0.634	0.629	0.629	0.631
IB13	62.533	37.467	1.841	0.625	0.625	0.631	0.625	0.625	0.628
IB15	62.617	37.383	1.399	0.626	0.626	0.631	0.626	0.626	0.628
IB17	62.633	37.367	1.807	0.626	0.626	0.632	0.626	0.626	0.629
IB19	62.467	37.533	1.628	0.625	0.625	0.630	0.625	0.625	0.627
IB21	62.250	37.750	1.537	0.623	0.623	0.628	0.623	0.623	0.625
IB25	61.750	38.250	1.960	0.618	0.618	0.624	0.617	0.618	0.621
IB29	61.883	38.117	1.860	0.619	0.619	0.626	0.619	0.619	0.623
IB35	61.450	38.550	1.261	0.615	0.615	0.621	0.614	0.615	0.618
IB41	61.183	38.817	1.343	0.612	0.612	0.617	0.612	0.612	0.614
IB49	60.183	39.817	1.618	0.602	0.602	0.608	0.602	0.602	0.605
IB57	59.900	40.100	1.697	0.599	0.599	0.606	0.599	0.599	0.603
IB73	59.167	40.833	1.238	0.592	0.592	0.598	0.592	0.592	0.595
IB89	58.517	41.483	1.336	0.585	0.585	0.593	0.585	0.585	0.589
IB1-WV	60.450	39.550	1.665	0.605	0.605	0.602	0.605	0.605	0.603
IB3-WV	62.717	37.283	1.753	0.627	0.627	0.625	0.627	0.627	0.626
IB5-WV	64.117	35.883	1.487	0.641	0.641	0.641	0.641	0.641	0.641
IB7-WV	64.283	35.717	1.609	0.643	0.643	0.645	0.643	0.643	0.644
IB9-WV	64.383	35.617	1.500	0.644	0.644	0.646	0.644	0.644	0.645
IB11-WV	64.150	35.850	1.385	0.642	0.642	0.645	0.642	0.642	0.643
IB13-WV	63.933	36.067	1.829	0.639	0.639	0.644	0.639	0.639	0.641
IB15-WV	63.417	36.583	1.665	0.634	0.634	0.638	0.634	0.634	0.636
IB17-WV	63.583	36.417	1.703	0.636	0.636	0.640	0.636	0.636	0.638
IB19-WV	63.500	36.500	1.827	0.635	0.635	0.640	0.635	0.635	0.637
IB21-WV	63.350	36.650	1.776	0.634	0.634	0.639	0.634	0.634	0.636
IB25-WV	62.850	37.150	1.848	0.629	0.629	0.634	0.629	0.629	0.631
IB29-WV	62.850	37.150	1.832	0.629	0.629	0.635	0.629	0.629	0.632
IB35-WV	62.383	37.617	1.111	0.624	0.624	0.630	0.624	0.624	0.627
IB41-WV	62.200	37.800	1.473	0.622	0.622	0.628	0.622	0.622	0.625
IB49-WV	61.317	38.683	1.568	0.613	0.613	0.619	0.613	0.613	0.616
IB57-WV	61.000	39.000	1.683	0.610	0.610	0.617	0.610	0.610	0.614
IB73-WV	60.317	39.683	1.505	0.603	0.603	0.610	0.603	0.603	0.607
IB89-WV	59.500	40.500	1.346	0.595	0.595	0.603	0.595	0.595	0.599

3.9 Dmoz-Sports-500 Collection

Table 3.19 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.20 presents the classification results obtained by IBk for Dmoz-Sports-500 collection.

Table 3.19: Classification Results for Dmoz-Sports-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
Naïve Bayes	75.859	24.141	1.614	0.759	0.759	0.769	0.759	0.759	0.764
Multinomial NB	83.763	16.237	1.106	0.838	0.838	0.843	0.838	0.838	0.840
J48	83.889	16.111	0.971	0.839	0.839	0.842	0.839	0.839	0.840
SMO-Linear-0	3.704	96.296	0.341	0.037	0.037	0.001	0.037	0.037	0.003
SMO-Linear- 10^{-5}	2.985	97.015	0.707	0.030	0.030	0.046	0.030	0.030	0.036
SMO-Linear- 10^{-4}	2.985	97.015	0.707	0.030	0.030	0.046	0.030	0.030	0.036
SMO-Linear- 10^{-3}	2.778	97.222	0.191	0.028	0.028	0.009	0.028	0.028	0.014
SMO-Linear- 10^{-2}	57.156	42.845	2.212	0.572	0.572	0.686	0.572	0.572	0.623
SMO-Linear- 10^{-1}	82.659	17.341	1.190	0.827	0.827	0.850	0.827	0.827	0.838
SMO-Linear- 10^0	85.489	14.511	1.072	0.855	0.855	0.860	0.855	0.855	0.858
SMO-Linear- 10^1	83.244	16.756	1.252	0.832	0.832	0.836	0.832	0.832	0.834
SMO-Linear- 10^2	82.978	17.022	1.262	0.830	0.830	0.833	0.830	0.830	0.831
SMO-Linear- 10^3	82.978	17.022	1.262	0.830	0.830	0.833	0.830	0.830	0.831
SMO-Linear- 10^4	82.978	17.022	1.262	0.830	0.830	0.833	0.830	0.830	0.831
SMO-Linear- 10^5	82.978	17.022	1.262	0.830	0.830	0.833	0.830	0.830	0.831
SMO-Poly-0	3.704	96.296	0.341	0.037	0.037	0.001	0.037	0.037	0.003
SMO-Poly- 10^{-5}	2.793	97.208	0.210	0.028	0.028	0.046	0.028	0.028	0.035
SMO-Poly- 10^{-4}	4.622	95.378	2.095	0.046	0.046	0.117	0.046	0.046	0.066
SMO-Poly- 10^{-3}	4.007	95.993	0.784	0.040	0.040	0.126	0.040	0.040	0.061
SMO-Poly- 10^{-2}	46.407	53.593	2.557	0.464	0.464	0.709	0.464	0.464	0.561
SMO-Poly- 10^{-1}	73.533	26.467	1.206	0.735	0.735	0.791	0.735	0.735	0.762
SMO-Poly- 10^0	74.133	25.867	1.381	0.741	0.741	0.774	0.741	0.741	0.757
SMO-Poly- 10^1	72.822	27.178	1.202	0.728	0.728	0.755	0.728	0.728	0.742
SMO-Poly- 10^2	72.563	27.437	1.152	0.726	0.726	0.752	0.726	0.726	0.738
SMO-Poly- 10^3	72.563	27.437	1.152	0.726	0.726	0.752	0.726	0.726	0.738
SMO-Poly- 10^4	72.563	27.437	1.152	0.726	0.726	0.752	0.726	0.726	0.738
SMO-Poly- 10^5	72.563	27.437	1.152	0.726	0.726	0.752	0.726	0.726	0.738
SMO-RBF-0	3.704	96.296	0.341	0.037	0.037	0.001	0.037	0.037	0.003
SMO-RBF- 10^{-5}	2.978	97.022	0.686	0.030	0.030	0.046	0.030	0.030	0.036
SMO-RBF- 10^{-4}	2.978	97.022	0.686	0.030	0.030	0.046	0.030	0.030	0.036
SMO-RBF- 10^{-3}	2.978	97.022	0.686	0.030	0.030	0.046	0.030	0.030	0.036
SMO-RBF- 10^{-2}	2.978	97.022	0.686	0.030	0.030	0.046	0.030	0.030	0.036
SMO-RBF- 10^{-1}	3.496	96.504	0.800	0.035	0.035	0.156	0.035	0.035	0.057
SMO-RBF- 10^0	68.459	31.541	1.300	0.685	0.685	0.758	0.685	0.685	0.719
SMO-RBF- 10^1	84.756	15.245	1.378	0.848	0.848	0.861	0.848	0.848	0.854
SMO-RBF- 10^2	84.770	15.230	1.143	0.848	0.848	0.852	0.848	0.848	0.850
SMO-RBF- 10^3	83.356	16.645	1.194	0.834	0.834	0.836	0.834	0.834	0.835
SMO-RBF- 10^4	83.274	16.726	1.183	0.833	0.833	0.835	0.833	0.833	0.834
SMO-RBF- 10^5	83.274	16.726	1.183	0.833	0.833	0.835	0.833	0.833	0.834

Table 3.20: Classification Results for Dmoz-Sports-500 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	74.815	25.185	1.368	0.748	0.748	0.751	0.748	0.748	0.750
IB3	76.133	23.867	1.370	0.761	0.761	0.783	0.761	0.761	0.772
IB5	77.904	22.096	1.081	0.779	0.779	0.794	0.779	0.779	0.787
IB7	78.800	21.200	0.930	0.788	0.788	0.800	0.788	0.788	0.794
IB9	78.941	21.059	1.466	0.789	0.789	0.801	0.789	0.789	0.795
IB11	78.874	21.126	1.456	0.789	0.789	0.801	0.789	0.789	0.795
IB13	78.874	21.126	1.324	0.789	0.789	0.801	0.789	0.789	0.795
IB15	78.474	21.526	1.514	0.785	0.785	0.797	0.785	0.785	0.791
IB17	78.467	21.533	1.677	0.785	0.785	0.797	0.785	0.785	0.791
IB19	78.230	21.770	1.646	0.782	0.782	0.795	0.782	0.782	0.788
IB21	78.222	21.778	1.489	0.782	0.782	0.795	0.782	0.782	0.789
IB25	77.889	22.111	1.479	0.779	0.779	0.794	0.779	0.779	0.786
IB29	77.474	22.526	1.577	0.775	0.775	0.790	0.775	0.775	0.782
IB35	77.104	22.896	1.692	0.771	0.771	0.788	0.771	0.771	0.779
IB41	76.667	23.333	1.614	0.767	0.767	0.785	0.767	0.767	0.776
IB49	76.252	23.748	1.661	0.763	0.763	0.781	0.763	0.763	0.772
IB57	75.563	24.437	1.425	0.756	0.756	0.775	0.756	0.756	0.765
IB73	74.659	25.341	1.542	0.747	0.747	0.768	0.747	0.747	0.757
IB89	73.859	26.141	1.390	0.739	0.739	0.764	0.739	0.739	0.751
IB1-WV	74.815	25.185	1.368	0.748	0.748	0.751	0.748	0.748	0.750
IB3-WV	77.585	22.415	1.371	0.776	0.776	0.779	0.776	0.776	0.778
IB5-WV	79.207	20.793	1.040	0.792	0.792	0.797	0.792	0.792	0.794
IB7-WV	79.881	20.119	0.991	0.799	0.799	0.805	0.799	0.799	0.802
IB9-WV	80.096	19.904	1.202	0.801	0.801	0.808	0.801	0.801	0.804
IB11-WV	79.970	20.030	1.285	0.800	0.800	0.807	0.800	0.800	0.804
IB13-WV	79.904	20.096	1.321	0.799	0.799	0.807	0.799	0.799	0.803
IB15-WV	79.578	20.422	1.461	0.796	0.796	0.805	0.796	0.796	0.800
IB17-WV	79.504	20.496	1.416	0.795	0.795	0.804	0.795	0.795	0.800
IB19-WV	79.474	20.526	1.504	0.795	0.795	0.804	0.795	0.795	0.799
IB21-WV	79.415	20.585	1.506	0.794	0.794	0.804	0.794	0.794	0.799
IB25-WV	79.037	20.963	1.480	0.790	0.790	0.802	0.790	0.790	0.796
IB29-WV	78.719	21.282	1.629	0.787	0.787	0.800	0.787	0.787	0.794
IB35-WV	78.193	21.808	1.680	0.782	0.782	0.797	0.782	0.782	0.789
IB41-WV	77.674	22.326	1.607	0.777	0.777	0.793	0.777	0.777	0.785
IB49-WV	77.481	22.519	1.505	0.775	0.775	0.792	0.775	0.775	0.783
IB57-WV	76.659	23.341	1.373	0.767	0.767	0.785	0.767	0.767	0.776
IB73-WV	75.800	24.200	1.366	0.758	0.758	0.779	0.758	0.758	0.768
IB89-WV	75.081	24.919	1.343	0.751	0.751	0.775	0.751	0.751	0.763

3.10 Enron-Top-20 Collection

Table 3.21 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.22 presents the classification results obtained by IBk for Enron-Top-20 collection.

Table 3.21: Classification Results for Enron-Top-20 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	51.542	48.458	1.846	0.515	0.515	0.503	0.513	0.515	0.508
Multinomial NB	72.596	27.404	1.017	0.726	0.726	0.706	0.703	0.726	0.704
J48	64.467	35.533	1.882	0.645	0.645	0.625	0.611	0.645	0.618
SMO-Linear-0	2.871	97.129	0.362	0.029	0.029	0.001	0.050	0.029	0.003
SMO-Linear- 10^{-5}	9.031	90.969	0.671	0.090	0.090	0.005	0.050	0.090	0.008
SMO-Linear- 10^{-4}	9.031	90.969	0.671	0.090	0.090	0.005	0.050	0.090	0.008
SMO-Linear- 10^{-3}	9.122	90.878	0.698	0.091	0.091	0.055	0.051	0.091	0.053
SMO-Linear- 10^{-2}	24.812	75.188	1.121	0.248	0.248	0.689	0.177	0.248	0.281
SMO-Linear- 10^{-1}	56.838	43.162	1.264	0.568	0.568	0.568	0.492	0.568	0.527
SMO-Linear- 10^0	65.982	34.018	1.280	0.660	0.660	0.669	0.619	0.660	0.643
SMO-Linear- 10^1	65.550	34.450	1.014	0.656	0.656	0.645	0.629	0.656	0.637
SMO-Linear- 10^2	63.103	36.897	0.794	0.631	0.631	0.617	0.609	0.631	0.613
SMO-Linear- 10^3	60.467	39.533	1.259	0.605	0.605	0.592	0.586	0.605	0.589
SMO-Linear- 10^4	59.656	40.344	1.024	0.597	0.597	0.586	0.581	0.597	0.583
SMO-Linear- 10^5	59.535	40.465	1.037	0.595	0.595	0.586	0.580	0.595	0.583
SMO-Poly-0	2.871	97.129	0.362	0.029	0.029	0.001	0.050	0.029	0.003
SMO-Poly- 10^{-5}	9.084	90.916	0.657	0.0908	0.0908	0.1045	0.0504	0.091	0.068
SMO-Poly- 10^{-4}	9.205	90.795	0.630	0.0920	0.0920	0.1753	0.0516	0.092	0.080
SMO-Poly- 10^{-3}	10.205	89.795	0.704	0.1020	0.1020	0.4757	0.0645	0.102	0.114
SMO-Poly- 10^{-2}	22.835	77.165	2.422	0.228	0.228	0.673	0.185	0.228	0.290
SMO-Poly- 10^{-1}	40.313	59.687	1.487	0.403	0.403	0.669	0.350	0.403	0.460
SMO-Poly- 10^0	47.791	52.209	1.491	0.478	0.478	0.629	0.443	0.478	0.520
SMO-Poly- 10^1	53.565	46.435	1.099	0.536	0.536	0.604	0.509	0.536	0.553
SMO-Poly- 10^2	55.065	44.935	0.993	0.551	0.551	0.570	0.527	0.551	0.548
SMO-Poly- 10^3	54.549	45.451	1.224	0.545	0.545	0.545	0.527	0.545	0.536
SMO-Poly- 10^4	52.807	47.193	1.239	0.528	0.528	0.517	0.509	0.528	0.513
SMO-Poly- 10^5	51.458	48.542	1.489	0.515	0.515	0.498	0.497	0.515	0.497
SMO-RBF-0	2.871	97.129	0.362	0.029	0.029	0.001	0.050	0.029	0.003
SMO-RBF- 10^{-5}	9.031	90.969	0.671	0.090	0.090	0.005	0.050	0.090	0.008
SMO-RBF- 10^{-4}	9.031	90.969	0.671	0.090	0.090	0.005	0.050	0.090	0.008
SMO-RBF- 10^{-3}	9.031	90.969	0.671	0.090	0.090	0.005	0.050	0.090	0.008
SMO-RBF- 10^{-2}	9.031	90.969	0.671	0.090	0.090	0.005	0.050	0.090	0.008
SMO-RBF- 10^{-1}	9.849	90.151	0.622	0.098	0.098	0.021	0.055	0.098	0.030
SMO-RBF- 10^0	33.192	66.808	1.136	0.332	0.332	0.636	0.257	0.332	0.366
SMO-RBF- 10^1	61.353	38.647	0.943	0.614	0.614	0.677	0.550	0.614	0.607
SMO-RBF- 10^2	65.270	34.730	1.158	0.653	0.653	0.661	0.616	0.653	0.637
SMO-RBF- 10^3	64.422	35.578	0.704	0.644	0.644	0.642	0.619	0.644	0.630
SMO-RBF- 10^4	61.739	38.261	1.283	0.617	0.617	0.620	0.593	0.617	0.606
SMO-RBF- 10^5	59.739	40.261	1.196	0.597	0.597	0.596	0.577	0.597	0.586

Table 3.22: Classification Results for Enron-Top-20 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	66.717	33.283	0.907	0.667	0.667	0.655	0.649	0.667	0.652
IB3	64.937	35.063	0.920	0.649	0.649	0.634	0.626	0.649	0.630
IB5	65.414	34.586	0.754	0.654	0.654	0.633	0.628	0.654	0.631
IB7	65.452	34.548	0.984	0.655	0.655	0.635	0.626	0.655	0.631
IB9	65.194	34.806	0.971	0.652	0.652	0.632	0.623	0.652	0.627
IB11	64.808	35.192	1.057	0.648	0.648	0.629	0.618	0.648	0.623
IB13	64.512	35.488	0.946	0.645	0.645	0.627	0.615	0.645	0.621
IB15	64.103	35.897	1.259	0.641	0.641	0.622	0.610	0.641	0.616
IB17	63.762	36.238	1.406	0.638	0.638	0.616	0.605	0.638	0.611
IB19	63.482	36.518	1.098	0.635	0.635	0.615	0.601	0.635	0.608
IB21	63.482	36.518	0.896	0.635	0.635	0.616	0.600	0.635	0.608
IB25	62.944	37.056	0.782	0.629	0.629	0.609	0.594	0.629	0.601
IB29	62.793	37.208	0.766	0.628	0.628	0.604	0.592	0.628	0.598
IB35	62.308	37.692	0.852	0.623	0.623	0.599	0.587	0.623	0.593
IB41	61.739	38.261	1.139	0.617	0.617	0.597	0.579	0.617	0.588
IB49	61.452	38.548	1.183	0.615	0.615	0.597	0.576	0.615	0.586
IB57	60.936	39.064	0.956	0.609	0.609	0.592	0.571	0.609	0.581
IB73	60.285	39.715	1.001	0.603	0.603	0.585	0.562	0.603	0.573
IB89	59.565	40.435	1.147	0.596	0.596	0.580	0.554	0.596	0.567
IB1-WV	66.717	33.283	0.907	0.667	0.667	0.655	0.649	0.667	0.652
IB3-WV	66.717	33.283	0.809	0.667	0.667	0.658	0.647	0.667	0.652
IB5-WV	66.634	33.366	0.805	0.666	0.666	0.660	0.643	0.666	0.651
IB7-WV	66.331	33.669	0.833	0.663	0.663	0.657	0.638	0.663	0.647
IB9-WV	66.164	33.836	0.807	0.662	0.662	0.656	0.636	0.662	0.646
IB11-WV	65.656	34.344	0.569	0.657	0.657	0.653	0.631	0.657	0.642
IB13-WV	65.187	34.813	0.895	0.652	0.652	0.648	0.625	0.652	0.636
IB15-WV	64.785	35.215	0.917	0.648	0.648	0.643	0.620	0.648	0.631
IB17-WV	64.528	35.472	1.058	0.645	0.645	0.640	0.616	0.645	0.628
IB19-WV	63.868	36.132	1.092	0.639	0.639	0.633	0.608	0.639	0.620
IB21-WV	63.717	36.283	0.985	0.637	0.637	0.632	0.606	0.637	0.618
IB25-WV	63.164	36.836	1.046	0.632	0.632	0.625	0.598	0.632	0.611
IB29-WV	62.868	37.132	0.944	0.629	0.629	0.621	0.595	0.629	0.608
IB35-WV	62.118	37.882	1.196	0.621	0.621	0.611	0.587	0.621	0.599
IB41-WV	61.717	38.283	1.392	0.617	0.617	0.608	0.582	0.617	0.595
IB49-WV	61.179	38.821	1.496	0.612	0.612	0.605	0.576	0.612	0.590
IB57-WV	60.467	39.533	1.327	0.605	0.605	0.595	0.567	0.605	0.581
IB73-WV	59.929	40.071	1.355	0.599	0.599	0.588	0.561	0.599	0.574
IB89-WV	59.413	40.587	1.317	0.594	0.594	0.586	0.555	0.594	0.570

3.11 FBIS Collection

Table 3.23 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.24 presents the classification results obtained by IBk for FBIS collection.

Table 3.23: Classification Results for FBIS collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	61.796	38.204	2.522	0.618	0.618	0.533	0.552	0.618	0.543
Multinomial NB	77.183	22.817	2.596	0.772	0.772	0.730	0.750	0.772	0.739
J48	71.498	28.502	2.234	0.715	0.715	0.642	0.635	0.715	0.638
SMO-Linear-0	3.736	96.264	0.869	0.037	0.037	0.002	0.059	0.037	0.004
SMO-Linear- 10^{-5}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-Linear- 10^{-4}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-Linear- 10^{-3}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-Linear- 10^{-2}	41.496	58.504	2.903	0.415	0.415	0.350	0.160	0.415	0.220
SMO-Linear- 10^{-1}	64.190	35.810	1.614	0.642	0.642	0.649	0.402	0.642	0.496
SMO-Linear- 10^0	78.241	21.759	2.565	0.782	0.782	0.761	0.660	0.782	0.707
SMO-Linear- 10^1	78.930	21.070	2.593	0.789	0.789	0.756	0.716	0.789	0.735
SMO-Linear- 10^2	78.605	21.395	2.419	0.786	0.786	0.753	0.714	0.786	0.733
SMO-Linear- 10^3	78.605	21.395	2.419	0.786	0.786	0.753	0.714	0.786	0.733
SMO-Linear- 10^4	78.605	21.395	2.419	0.786	0.786	0.753	0.714	0.786	0.733
SMO-Linear- 10^5	78.605	21.395	2.419	0.786	0.786	0.753	0.714	0.786	0.733
SMO-Poly-0	3.736	96.264	0.869	0.037	0.037	0.002	0.059	0.037	0.004
SMO-Poly- 10^{-5}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-Poly- 10^{-4}	20.665	79.336	1.996	0.207	0.207	0.130	0.060	0.207	0.082
SMO-Poly- 10^{-3}	24.157	75.843	2.851	0.242	0.242	0.293	0.078	0.242	0.123
SMO-Poly- 10^{-2}	49.412	50.588	2.610	0.494	0.494	0.514	0.249	0.494	0.335
SMO-Poly- 10^{-1}	68.981	31.019	1.948	0.690	0.690	0.709	0.495	0.690	0.583
SMO-Poly- 10^0	76.087	23.913	1.982	0.761	0.761	0.724	0.651	0.761	0.685
SMO-Poly- 10^1	76.046	23.954	2.009	0.760	0.760	0.714	0.655	0.760	0.683
SMO-Poly- 10^2	76.046	23.954	2.009	0.760	0.760	0.714	0.655	0.760	0.683
SMO-Poly- 10^3	76.046	23.954	2.009	0.760	0.760	0.714	0.655	0.760	0.683
SMO-Poly- 10^4	76.046	23.954	2.009	0.760	0.760	0.714	0.655	0.760	0.683
SMO-Poly- 10^5	76.046	23.954	2.009	0.760	0.760	0.714	0.655	0.760	0.683
SMO-RBF-0	3.736	96.264	0.869	0.037	0.037	0.002	0.059	0.037	0.004
SMO-RBF- 10^{-5}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-RBF- 10^{-4}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-RBF- 10^{-3}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-RBF- 10^{-2}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-RBF- 10^{-1}	20.543	79.457	2.069	0.205	0.205	0.012	0.059	0.205	0.020
SMO-RBF- 10^0	46.408	53.592	2.765	0.464	0.464	0.370	0.201	0.464	0.261
SMO-RBF- 10^1	69.509	30.491	2.375	0.695	0.695	0.683	0.481	0.695	0.564
SMO-RBF- 10^2	78.890	21.110	2.779	0.789	0.789	0.761	0.687	0.789	0.722
SMO-RBF- 10^3	78.361	21.639	2.578	0.784	0.784	0.756	0.705	0.784	0.729
SMO-RBF- 10^4	78.442	21.558	2.504	0.784	0.784	0.756	0.705	0.784	0.729
SMO-RBF- 10^5	78.442	21.558	2.504	0.784	0.784	0.756	0.705	0.784	0.729

Table 3.24: Classification Results for FBIS collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	76.331	23.669	2.805	0.763	0.763	0.729	0.696	0.763	0.712
IB3	78.075	21.925	2.513	0.781	0.781	0.778	0.674	0.781	0.722
IB5	79.293	20.707	2.853	0.793	0.793	0.794	0.672	0.793	0.728
IB7	79.659	20.342	2.810	0.797	0.797	0.808	0.668	0.797	0.731
IB9	78.521	21.479	3.228	0.785	0.785	0.815	0.646	0.785	0.720
IB11	77.993	22.007	2.953	0.780	0.780	0.784	0.630	0.780	0.699
IB13	78.197	21.803	3.274	0.782	0.782	0.799	0.633	0.782	0.706
IB15	77.710	22.290	3.699	0.777	0.777	0.801	0.618	0.777	0.698
IB17	77.751	22.249	3.466	0.778	0.778	0.807	0.624	0.778	0.704
IB19	77.670	22.331	3.146	0.777	0.777	0.818	0.621	0.777	0.706
IB21	77.426	22.575	3.454	0.774	0.774	0.806	0.615	0.774	0.697
IB25	76.978	23.022	3.372	0.770	0.770	0.786	0.602	0.770	0.682
IB29	76.532	23.468	3.029	0.765	0.765	0.784	0.592	0.765	0.675
IB35	76.532	23.469	2.709	0.765	0.765	0.736	0.589	0.765	0.654
IB41	76.167	23.833	2.568	0.762	0.762	0.747	0.580	0.762	0.653
IB49	75.882	24.118	2.680	0.759	0.759	0.750	0.570	0.759	0.648
IB57	75.476	24.524	2.579	0.755	0.755	0.743	0.557	0.755	0.637
IB73	74.217	25.783	2.903	0.742	0.742	0.746	0.537	0.742	0.624
IB89	73.446	26.554	2.909	0.734	0.734	0.746	0.523	0.734	0.615
IB1-WV	76.331	23.669	2.805	0.763	0.763	0.729	0.696	0.763	0.712
IB3-WV	79.618	20.382	2.641	0.796	0.796	0.772	0.708	0.796	0.738
IB5-WV	80.430	19.570	2.551	0.804	0.804	0.796	0.705	0.804	0.748
IB7-WV	80.999	19.001	2.549	0.810	0.810	0.819	0.701	0.810	0.756
IB9-WV	80.713	19.287	2.923	0.807	0.807	0.837	0.691	0.807	0.757
IB11-WV	80.023	19.977	2.610	0.800	0.800	0.829	0.679	0.800	0.746
IB13-WV	79.943	20.057	2.515	0.799	0.799	0.818	0.672	0.799	0.738
IB15-WV	79.740	20.260	2.597	0.797	0.797	0.827	0.666	0.797	0.738
IB17-WV	79.456	20.544	2.870	0.795	0.795	0.822	0.658	0.795	0.731
IB19-WV	79.253	20.747	2.827	0.793	0.793	0.835	0.654	0.793	0.733
IB21-WV	78.888	21.112	2.971	0.789	0.789	0.833	0.646	0.789	0.728
IB25-WV	78.643	21.357	3.011	0.786	0.786	0.815	0.634	0.786	0.713
IB29-WV	78.156	21.844	2.983	0.782	0.782	0.792	0.623	0.782	0.698
IB35-WV	78.196	21.804	2.359	0.782	0.782	0.795	0.616	0.782	0.694
IB41-WV	77.506	22.494	2.961	0.775	0.775	0.812	0.602	0.775	0.692
IB49-WV	77.302	22.698	2.836	0.773	0.773	0.756	0.595	0.773	0.666
IB57-WV	76.816	23.184	2.402	0.768	0.768	0.754	0.583	0.768	0.658
IB73-WV	75.516	24.484	2.669	0.755	0.755	0.753	0.557	0.755	0.641
IB89-WV	74.663	25.337	2.735	0.747	0.747	0.751	0.540	0.747	0.628

3.12 Hitech Collection

Table 3.25 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.26 presents the classification results obtained by IBk for Hitech collection.

Table 3.25: Classification Results for Hitech collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	62.930	37.070	2.924	0.629	0.629	0.583	0.577	0.629	0.580
Multinomial NB	72.924	27.076	2.548	0.729	0.729	0.711	0.667	0.729	0.689
J48	56.760	43.240	3.534	0.568	0.568	0.533	0.519	0.568	0.526
SMO-Linear-0	18.643	81.358	2.642	0.186	0.186	0.031	0.167	0.186	0.052
SMO-Linear- 10^{-5}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-Linear- 10^{-4}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-Linear- 10^{-3}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-Linear- 10^{-2}	44.287	55.713	3.526	0.443	0.443	0.695	0.314	0.443	0.432
SMO-Linear- 10^{-1}	66.449	33.552	2.792	0.664	0.664	0.649	0.571	0.664	0.607
SMO-Linear- 10^0	66.144	33.856	2.810	0.661	0.661	0.642	0.593	0.661	0.617
SMO-Linear- 10^1	64.883	35.118	3.480	0.649	0.649	0.619	0.583	0.649	0.600
SMO-Linear- 10^2	64.883	35.118	3.480	0.649	0.649	0.619	0.583	0.649	0.600
SMO-Linear- 10^3	64.883	35.118	3.480	0.649	0.649	0.619	0.583	0.649	0.600
SMO-Linear- 10^4	64.883	35.118	3.480	0.649	0.649	0.619	0.583	0.649	0.600
SMO-Linear- 10^5	64.883	35.118	3.480	0.649	0.649	0.619	0.583	0.649	0.600
SMO-Poly-0	18.643	81.358	2.642	0.186	0.186	0.031	0.167	0.186	0.052
SMO-Poly- 10^{-5}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-Poly- 10^{-4}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-Poly- 10^{-3}	32.291	67.709	2.164	0.323	0.323	0.666	0.224	0.323	0.335
SMO-Poly- 10^{-2}	51.804	48.196	1.940	0.518	0.518	0.647	0.406	0.518	0.499
SMO-Poly- 10^{-1}	53.108	46.892	1.268	0.531	0.531	0.620	0.438	0.531	0.513
SMO-Poly- 10^0	45.111	54.889	2.352	0.451	0.451	0.545	0.418	0.451	0.473
SMO-Poly- 10^1	45.197	54.803	3.301	0.452	0.452	0.574	0.394	0.452	0.467
SMO-Poly- 10^2	43.284	56.716	3.513	0.433	0.433	0.573	0.377	0.433	0.455
SMO-Poly- 10^3	43.284	56.716	3.513	0.433	0.433	0.573	0.377	0.433	0.455
SMO-Poly- 10^4	43.284	56.716	3.513	0.433	0.433	0.573	0.377	0.433	0.455
SMO-Poly- 10^5	43.284	56.716	3.513	0.433	0.433	0.573	0.377	0.433	0.455
SMO-RBF-0	18.643	81.358	2.642	0.186	0.186	0.031	0.167	0.186	0.052
SMO-RBF- 10^{-5}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-RBF- 10^{-4}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-RBF- 10^{-3}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-RBF- 10^{-2}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-RBF- 10^{-1}	26.207	73.793	1.953	0.262	0.262	0.044	0.167	0.262	0.069
SMO-RBF- 10^0	49.198	50.802	2.995	0.492	0.492	0.530	0.358	0.492	0.427
SMO-RBF- 10^1	64.536	35.464	2.683	0.645	0.645	0.649	0.558	0.645	0.600
SMO-RBF- 10^2	65.144	34.856	2.864	0.651	0.651	0.660	0.569	0.651	0.611
SMO-RBF- 10^3	63.970	36.030	3.881	0.640	0.640	0.633	0.560	0.640	0.594
SMO-RBF- 10^4	63.970	36.030	3.881	0.640	0.640	0.633	0.560	0.640	0.594
SMO-RBF- 10^5	63.970	36.030	3.881	0.640	0.640	0.633	0.560	0.640	0.594

Table 3.26: Classification Results for Hitech collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	68.274	31.726	2.261	0.683	0.683	0.660	0.653	0.683	0.657
IB3	68.839	31.161	2.618	0.688	0.688	0.675	0.642	0.688	0.658
IB5	68.622	31.378	2.649	0.686	0.686	0.687	0.632	0.686	0.658
IB7	69.666	30.334	3.053	0.697	0.697	0.698	0.634	0.697	0.664
IB9	70.058	29.942	3.098	0.701	0.701	0.709	0.629	0.701	0.666
IB11	69.580	30.420	3.349	0.696	0.696	0.703	0.613	0.696	0.655
IB13	69.839	30.161	2.727	0.698	0.698	0.716	0.616	0.698	0.662
IB15	69.753	30.247	3.348	0.698	0.698	0.722	0.609	0.698	0.660
IB17	69.797	30.203	3.491	0.698	0.698	0.723	0.608	0.698	0.661
IB19	69.449	30.551	2.988	0.694	0.694	0.724	0.600	0.694	0.656
IB21	69.580	30.420	2.945	0.696	0.696	0.725	0.595	0.696	0.654
IB25	69.361	30.639	3.206	0.694	0.694	0.739	0.589	0.694	0.655
IB29	69.188	30.812	3.110	0.692	0.692	0.745	0.581	0.692	0.653
IB35	68.188	31.812	3.189	0.682	0.682	0.752	0.563	0.682	0.644
IB41	68.275	31.725	2.897	0.683	0.683	0.741	0.559	0.683	0.637
IB49	67.449	32.551	2.655	0.674	0.674	0.751	0.548	0.674	0.633
IB57	66.754	33.246	2.468	0.668	0.668	0.751	0.539	0.668	0.627
IB73	65.538	34.463	2.834	0.655	0.655	0.736	0.520	0.655	0.609
IB89	65.755	34.245	2.949	0.658	0.658	0.744	0.515	0.658	0.608
IB1-WV	68.274	31.726	2.261	0.683	0.683	0.660	0.653	0.683	0.657
IB3-WV	70.882	29.118	2.860	0.709	0.709	0.692	0.667	0.709	0.679
IB5-WV	71.013	28.987	2.557	0.710	0.710	0.707	0.663	0.710	0.684
IB7-WV	71.795	28.205	2.378	0.718	0.718	0.722	0.665	0.718	0.692
IB9-WV	71.578	28.422	3.333	0.716	0.716	0.725	0.657	0.716	0.689
IB11-WV	71.622	28.378	3.137	0.716	0.716	0.724	0.652	0.716	0.686
IB13-WV	71.534	28.466	2.707	0.715	0.715	0.730	0.642	0.715	0.683
IB15-WV	71.274	28.726	3.250	0.713	0.713	0.728	0.636	0.713	0.679
IB17-WV	71.491	28.509	2.974	0.715	0.715	0.737	0.636	0.715	0.683
IB19-WV	71.405	28.595	3.164	0.714	0.714	0.741	0.630	0.714	0.681
IB21-WV	70.927	29.073	2.803	0.709	0.709	0.740	0.621	0.709	0.675
IB25-WV	70.709	29.291	3.094	0.707	0.707	0.742	0.611	0.707	0.670
IB29-WV	70.404	29.596	3.020	0.704	0.704	0.746	0.601	0.704	0.666
IB35-WV	68.927	31.073	3.569	0.689	0.689	0.731	0.581	0.689	0.647
IB41-WV	69.404	30.596	2.908	0.694	0.694	0.741	0.579	0.694	0.650
IB49-WV	68.579	31.422	2.612	0.686	0.686	0.737	0.564	0.686	0.639
IB57-WV	67.927	32.074	2.463	0.679	0.679	0.745	0.557	0.679	0.637
IB73-WV	66.884	33.116	2.826	0.669	0.669	0.760	0.538	0.669	0.630
IB89-WV	66.668	33.332	3.076	0.667	0.667	0.764	0.531	0.667	0.627

3.13 Industry-Sector Collection

Table 3.27 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.28 presents the classification results obtained by IBk for Industry-Sector collection.

Table 3.27: Classification Results for Industry-Sector collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	41.545	58.455	1.655	0.415	0.415	0.478	0.463	0.415	0.471
Multinomial NB	76.228	23.772	1.225	0.762	0.762	0.760	0.741	0.762	0.750
J48	57.559	42.441	1.930	0.576	0.576	0.564	0.531	0.576	0.547
SMO-Linear-0	4.015	95.985	0.392	0.040	0.040	0.003	0.083	0.040	0.006
SMO-Linear- 10^{-5}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-Linear- 10^{-4}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-Linear- 10^{-3}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-Linear- 10^{-2}	31.677	68.323	1.509	0.317	0.317	0.778	0.112	0.317	0.195
SMO-Linear- 10^{-1}	50.777	49.223	1.839	0.508	0.508	0.852	0.359	0.508	0.505
SMO-Linear- 10^0	66.349	33.651	1.226	0.663	0.663	0.789	0.587	0.663	0.673
SMO-Linear- 10^1	70.341	29.659	1.010	0.703	0.703	0.748	0.670	0.703	0.707
SMO-Linear- 10^2	67.607	32.393	1.572	0.676	0.676	0.662	0.662	0.676	0.662
SMO-Linear- 10^3	62.356	37.644	1.721	0.624	0.624	0.596	0.623	0.624	0.609
SMO-Linear- 10^4	58.376	41.624	1.296	0.584	0.584	0.565	0.590	0.584	0.577
SMO-Linear- 10^5	57.412	42.588	1.620	0.574	0.574	0.558	0.583	0.574	0.570
SMO-Poly-0	4.015	95.985	0.392	0.040	0.040	0.003	0.083	0.040	0.006
SMO-Poly- 10^{-5}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-Poly- 10^{-4}	29.477	70.523	1.357	0.295	0.295	0.066	0.084	0.295	0.074
SMO-Poly- 10^{-3}	23.305	76.695	1.881	0.233	0.233	0.586	0.233	0.233	0.333
SMO-Poly- 10^{-2}	28.172	71.828	2.470	0.282	0.282	0.626	0.626	0.282	0.626
SMO-Poly- 10^{-1}	27.199	72.801	3.481	0.272	0.272	0.605	0.279	0.272	0.382
SMO-Poly- 10^0	18.182	81.818	4.631	0.182	0.182	0.535	0.187	0.182	0.277
SMO-Poly- 10^1	56.300	43.700	0.939	0.563	0.563	0.668	0.486	0.563	0.563
SMO-Poly- 10^2	57.128	42.872	1.639	0.571	0.571	0.618	0.534	0.571	0.573
SMO-Poly- 10^3	56.947	43.053	1.380	0.569	0.569	0.570	0.557	0.569	0.564
SMO-Poly- 10^4	54.418	45.582	1.781	0.544	0.544	0.527	0.552	0.544	0.539
SMO-Poly- 10^5	50.969	49.031	2.320	0.510	0.510	0.496	0.532	0.510	0.513
SMO-RBF-0	4.015	95.985	0.392	0.040	0.040	0.003	0.083	0.040	0.006
SMO-RBF- 10^{-5}	2.004	97.996	0.512	0.020	0.020	0.003	0.017	0.020	0.005
SMO-RBF- 10^{-4}	2.004	97.996	0.512	0.020	0.020	0.003	0.017	0.020	0.005
SMO-RBF- 10^{-3}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-RBF- 10^{-2}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-RBF- 10^{-1}	29.432	70.568	1.378	0.294	0.294	0.025	0.083	0.294	0.038
SMO-RBF- 10^0	31.995	68.005	1.386	0.320	0.320	0.530	0.111	0.320	0.184
SMO-RBF- 10^1	55.223	44.777	2.040	0.552	0.552	0.841	0.417	0.552	0.558
SMO-RBF- 10^2	66.780	33.220	1.220	0.668	0.668	0.815	0.589	0.668	0.684
SMO-RBF- 10^3	69.332	30.668	1.213	0.693	0.693	0.776	0.638	0.693	0.700
SMO-RBF- 10^4	64.920	35.080	1.910	0.649	0.649	0.685	0.602	0.649	0.641
SMO-RBF- 10^5	58.648	41.352	1.847	0.586	0.586	0.609	0.548	0.586	0.577

Table 3.28: Classification Results for Industry-Sector collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	77.577	22.423	1.097	0.776	0.776	0.770	0.766	0.776	0.768
IB3	75.751	24.249	0.923	0.758	0.758	0.789	0.727	0.758	0.757
IB5	75.071	24.929	0.971	0.751	0.751	0.787	0.717	0.751	0.750
IB7	74.209	25.791	0.971	0.742	0.742	0.790	0.971	0.742	0.871
IB9	73.041	26.960	1.031	0.730	0.730	0.770	0.690	0.730	0.728
IB11	72.258	27.742	0.772	0.723	0.723	0.767	0.675	0.723	0.718
IB13	71.498	28.502	0.968	0.715	0.715	0.761	0.662	0.715	0.708
IB15	70.137	29.863	0.921	0.701	0.701	0.769	0.645	0.701	0.702
IB17	68.844	31.156	0.719	0.688	0.688	0.763	0.628	0.688	0.689
IB19	67.506	32.494	0.638	0.675	0.675	0.756	0.612	0.675	0.676
IB21	66.451	33.549	0.873	0.665	0.665	0.754	0.596	0.665	0.666
IB25	64.489	35.511	0.965	0.645	0.645	0.743	0.566	0.645	0.643
IB29	63.003	36.997	1.127	0.630	0.630	0.731	0.538	0.630	0.620
IB35	60.894	39.106	0.922	0.609	0.609	0.717	0.506	0.609	0.594
IB41	59.340	40.660	1.282	0.593	0.593	0.701	0.478	0.593	0.568
IB49	57.275	42.725	1.493	0.573	0.573	0.684	0.447	0.573	0.541
IB57	55.744	44.256	1.087	0.557	0.557	0.646	0.418	0.557	0.507
IB73	53.181	46.819	1.041	0.532	0.532	0.625	0.379	0.532	0.471
IB89	51.219	48.781	1.085	0.512	0.512	0.632	0.353	0.512	0.453
IB1-WV	77.577	77.577	1.097	0.776	0.776	0.770	0.766	0.776	0.768
IB3-WV	77.078	77.078	0.979	0.771	0.771	0.773	0.758	0.771	0.765
IB5-WV	76.942	76.942	0.882	0.769	0.769	0.793	0.751	0.769	0.771
IB7-WV	76.421	76.421	0.882	0.764	0.764	0.737	0.882	0.764	0.803
IB9-WV	75.536	75.536	1.097	0.755	0.755	0.797	0.726	0.755	0.760
IB11-WV	74.799	74.799	0.896	0.748	0.748	0.794	0.711	0.748	0.750
IB13-WV	74.163	74.163	1.102	0.742	0.742	0.785	0.698	0.742	0.739
IB15-WV	73.437	73.437	1.061	0.734	0.734	0.783	0.690	0.734	0.733
IB17-WV	71.612	28.388	0.530	0.716	0.716	0.781	0.664	0.716	0.718
IB19-WV	70.534	29.466	0.595	0.705	0.705	0.774	0.651	0.705	0.707
IB21-WV	69.298	30.702	0.816	0.693	0.693	0.772	0.632	0.693	0.695
IB25-WV	68.141	31.859	0.958	0.681	0.681	0.767	0.617	0.681	0.684
IB29-WV	66.383	33.617	0.724	0.664	0.664	0.758	0.584	0.664	0.660
IB35-WV	64.863	35.137	1.136	0.649	0.649	0.754	0.557	0.649	0.641
IB41-WV	64.863	35.137	1.136	0.649	0.649	0.754	0.557	0.649	0.641
IB49-WV	62.697	37.303	1.326	0.627	0.627	0.732	0.521	0.627	0.609
IB57-WV	61.222	38.778	1.535	0.612	0.612	0.727	0.489	0.612	0.585
IB73-WV	58.274	41.726	1.329	0.583	0.583	0.709	0.440	0.583	0.543
IB89-WV	55.733	44.267	1.084	0.557	0.557	0.688	0.406	0.557	0.511

3.14 Irish-Sentiment Collection

Table 3.29 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.30 presents the classification results obtained by IBk for Irish-Sentiment collection.

Table 3.29: Classification Results for Irish-Sentiment collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	59.759	40.241	3.672	0.598	0.598	0.589	0.592	0.598	0.590
Multinomial NB	67.651	32.349	2.640	0.677	0.677	0.669	0.669	0.677	0.669
J48	51.084	48.916	3.682	0.511	0.511	0.499	0.496	0.511	0.497
SMO-Linear-0	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Linear- 10^{-5}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Linear- 10^{-4}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Linear- 10^{-3}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Linear- 10^{-2}	54.217	45.783	3.714	0.542	0.542	0.619	0.480	0.542	0.541
SMO-Linear- 10^{-1}	65.542	34.458	2.581	0.655	0.655	0.644	0.636	0.655	0.640
SMO-Linear- 10^0	61.807	38.193	2.516	0.618	0.618	0.608	0.603	0.618	0.606
SMO-Linear- 10^1	61.145	38.856	2.487	0.611	0.611	0.601	0.596	0.611	0.599
SMO-Linear- 10^2	61.145	38.856	2.487	0.611	0.611	0.601	0.596	0.611	0.599
SMO-Linear- 10^3	61.145	38.856	2.487	0.611	0.611	0.601	0.596	0.611	0.599
SMO-Linear- 10^4	61.145	38.856	2.487	0.611	0.611	0.601	0.596	0.611	0.599
SMO-Linear- 10^5	61.145	38.856	2.487	0.611	0.611	0.601	0.596	0.611	0.599
SMO-Poly-0	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Poly- 10^{-5}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Poly- 10^{-4}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-Poly- 10^{-3}	43.072	56.928	3.607	0.431	0.431	0.672	0.370	0.431	0.477
SMO-Poly- 10^{-2}	54.096	45.904	2.330	0.541	0.541	0.622	0.495	0.541	0.551
SMO-Poly- 10^{-1}	54.940	45.060	2.048	0.549	0.549	0.598	0.516	0.549	0.554
SMO-Poly- 10^0	53.313	46.687	2.143	0.533	0.533	0.584	0.502	0.533	0.540
SMO-Poly- 10^1	53.072	46.928	2.196	0.531	0.531	0.582	0.499	0.531	0.538
SMO-Poly- 10^2	53.072	46.928	2.196	0.531	0.531	0.582	0.499	0.531	0.538
SMO-Poly- 10^3	53.072	46.928	2.196	0.531	0.531	0.582	0.499	0.531	0.538
SMO-Poly- 10^4	53.072	46.928	2.196	0.531	0.531	0.582	0.499	0.531	0.538
SMO-Poly- 10^5	53.072	46.928	2.196	0.531	0.531	0.582	0.499	0.531	0.538
SMO-RBF-0	39.458	60.542	2.196	0.395	0.395	0.132	0.333	0.395	0.189
SMO-RBF- 10^{-5}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-RBF- 10^{-4}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-RBF- 10^{-3}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-RBF- 10^{-2}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-RBF- 10^{-1}	39.458	60.542	3.075	0.395	0.395	0.132	0.333	0.395	0.189
SMO-RBF- 10^0	56.205	43.795	3.075	0.562	0.562	0.624	0.511	0.562	0.562
SMO-RBF- 10^1	64.518	35.482	2.800	0.645	0.645	0.636	0.628	0.645	0.632
SMO-RBF- 10^2	62.410	37.590	1.894	0.624	0.624	0.613	0.609	0.624	0.610
SMO-RBF- 10^3	62.169	37.831	1.687	0.622	0.622	0.609	0.606	0.622	0.607
SMO-RBF- 10^4	62.169	37.831	1.901	0.622	0.622	0.609	0.606	0.622	0.607
SMO-RBF- 10^5	62.169	37.831	1.901	0.622	0.622	0.609	0.606	0.622	0.607

Table 3.30: Classification Results for Irish-Sentiment collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	59.036	40.964	1.901	0.590	0.590	0.596	0.580	0.590	0.588
IB3	57.831	42.169	2.707	0.578	0.578	0.611	0.559	0.578	0.583
IB5	58.373	41.627	3.752	0.584	0.584	0.613	0.560	0.584	0.585
IB7	57.831	42.169	4.238	0.578	0.578	0.608	0.559	0.578	0.582
IB9	56.867	43.133	2.889	0.569	0.569	0.604	0.547	0.569	0.574
IB11	57.470	42.530	2.763	0.575	0.575	0.614	0.556	0.575	0.584
IB13	58.193	41.807	3.504	0.582	0.582	0.624	0.566	0.582	0.594
IB15	57.711	42.289	3.706	0.577	0.577	0.620	0.561	0.577	0.589
IB17	57.410	42.590	3.572	0.574	0.574	0.616	0.557	0.574	0.585
IB19	57.169	42.831	2.939	0.572	0.572	0.616	0.554	0.572	0.583
IB21	56.506	43.494	3.326	0.565	0.565	0.611	0.547	0.565	0.577
IB25	56.446	43.554	2.924	0.564	0.564	0.618	0.545	0.564	0.579
IB29	56.205	43.795	3.595	0.562	0.562	0.626	0.543	0.562	0.581
IB35	55.602	44.398	4.024	0.556	0.556	0.626	0.535	0.556	0.577
IB41	55.542	44.458	3.267	0.555	0.555	0.631	0.531	0.555	0.577
IB49	54.277	45.723	3.219	0.543	0.543	0.626	0.518	0.543	0.567
IB57	53.614	46.386	3.589	0.536	0.536	0.626	0.509	0.536	0.561
IB73	52.289	47.711	3.471	0.523	0.523	0.616	0.492	0.523	0.547
IB89	51.687	48.313	3.490	0.517	0.517	0.619	0.483	0.517	0.543
IB1-WV	59.036	40.964	2.707	0.590	0.590	0.596	0.580	0.590	0.588
IB3-WV	60.482	39.518	4.052	0.605	0.605	0.615	0.590	0.605	0.602
IB5-WV	60.602	39.398	3.686	0.606	0.606	0.625	0.591	0.606	0.607
IB7-WV	60.904	39.096	3.370	0.609	0.609	0.630	0.594	0.609	0.611
IB9-WV	58.855	41.145	3.211	0.589	0.589	0.612	0.572	0.589	0.592
IB11-WV	59.759	40.241	3.939	0.598	0.598	0.624	0.584	0.598	0.603
IB13-WV	60.422	39.578	4.122	0.604	0.604	0.636	0.592	0.604	0.613
IB15-WV	59.699	40.301	3.619	0.597	0.597	0.631	0.584	0.597	0.607
IB17-WV	59.398	40.603	3.606	0.594	0.594	0.630	0.580	0.594	0.604
IB19-WV	59.096	40.904	3.215	0.591	0.591	0.631	0.576	0.591	0.602
IB21-WV	58.795	41.205	3.389	0.588	0.588	0.626	0.573	0.588	0.598
IB25-WV	58.072	41.928	3.324	0.581	0.581	0.623	0.565	0.581	0.593
IB29-WV	57.410	42.590	3.096	0.574	0.574	0.629	0.558	0.574	0.591
IB35-WV	57.289	42.711	3.619	0.573	0.573	0.635	0.555	0.573	0.592
IB41-WV	56.446	43.554	3.492	0.564	0.564	0.630	0.544	0.564	0.584
IB49-WV	55.602	44.398	3.772	0.556	0.556	0.633	0.534	0.556	0.579
IB57-WV	54.819	45.181	3.429	0.548	0.548	0.628	0.524	0.548	0.572
IB73-WV	53.434	46.566	3.131	0.534	0.534	0.624	0.507	0.534	0.559
IB89-WV	52.711	47.289	3.421	0.527	0.527	0.628	0.496	0.527	0.554

3.15 La1s Collection

Table 3.31 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.32 presents the classification results obtained by IBk for La1s collection.

Table 3.31: Classification Results for La1s collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	75.218	24.783	2.361	0.752	0.752	0.712	0.735	0.752	0.723
Multinomial NB	88.171	11.829	1.672	0.882	0.882	0.857	0.856	0.882	0.857
J48	76.656	23.344	2.432	0.767	0.767	0.741	0.733	0.767	0.737
SMO-Linear-0	11.050	88.950	1.907	0.110	0.110	0.018	0.167	0.110	0.033
SMO-Linear- 10^{-5}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Linear- 10^{-4}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Linear- 10^{-3}	29.524	70.476	2.837	0.295	0.295	0.216	0.167	0.295	0.188
SMO-Linear- 10^{-2}	63.201	36.799	4.432	0.632	0.632	0.850	0.514	0.632	0.640
SMO-Linear- 10^{-1}	81.649	18.351	2.841	0.816	0.816	0.815	0.767	0.816	0.790
SMO-Linear- 10^0	84.301	15.699	2.250	0.843	0.843	0.823	0.820	0.843	0.821
SMO-Linear- 10^1	84.021	15.979	2.175	0.840	0.840	0.819	0.818	0.840	0.818
SMO-Linear- 10^2	84.021	15.979	2.175	0.840	0.840	0.819	0.818	0.840	0.818
SMO-Linear- 10^3	84.021	15.979	2.175	0.840	0.840	0.819	0.818	0.840	0.818
SMO-Linear- 10^4	84.021	15.979	2.175	0.840	0.840	0.819	0.818	0.840	0.818
SMO-Linear- 10^5	84.021	15.979	2.175	0.840	0.840	0.819	0.818	0.840	0.818
SMO-Poly-0	11.050	88.950	1.907	0.110	0.110	0.018	0.167	0.110	0.033
SMO-Poly- 10^{-5}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Poly- 10^{-4}	29.712	70.288	2.735	0.297	0.297	0.549	0.170	0.297	0.259
SMO-Poly- 10^{-3}	41.415	58.585	3.695	0.414	0.414	0.772	0.279	0.414	0.410
SMO-Poly- 10^{-2}	66.854	33.146	3.191	0.669	0.669	0.747	0.578	0.669	0.652
SMO-Poly- 10^{-1}	72.972	27.028	2.118	0.730	0.730	0.723	0.682	0.730	0.702
SMO-Poly- 10^0	72.816	27.184	2.855	0.728	0.728	0.714	0.693	0.728	0.703
SMO-Poly- 10^1	72.971	27.029	2.657	0.730	0.730	0.715	0.691	0.730	0.703
SMO-Poly- 10^2	72.971	27.029	2.549	0.730	0.730	0.715	0.691	0.730	0.703
SMO-Poly- 10^3	72.971	27.029	2.549	0.730	0.730	0.715	0.691	0.730	0.703
SMO-Poly- 10^4	72.971	27.029	2.549	0.730	0.730	0.715	0.691	0.730	0.703
SMO-Poly- 10^5	72.971	27.029	2.549	0.730	0.730	0.715	0.691	0.730	0.703
SMO-RBF-0	11.050	88.950	1.907	0.110	0.110	0.018	0.167	0.110	0.033
SMO-RBF- 10^{-5}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-4}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-3}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-2}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-1}	29.431	70.569	2.826	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^0	66.229	33.771	4.131	0.662	0.662	0.727	0.574	0.662	0.642
SMO-RBF- 10^1	80.244	19.756	2.083	0.802	0.802	0.790	0.772	0.802	0.781
SMO-RBF- 10^2	81.210	18.790	2.327	0.812	0.812	0.794	0.786	0.812	0.790
SMO-RBF- 10^3	81.586	18.415	2.289	0.816	0.816	0.798	0.791	0.816	0.794
SMO-RBF- 10^4	81.586	18.415	2.289	0.816	0.816	0.798	0.791	0.816	0.794
SMO-RBF- 10^5	81.586	18.415	2.289	0.816	0.816	0.798	0.791	0.816	0.794

Table 3.32: Classification Results for Lals collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	80.399	19.601	2.334	0.804	0.804	0.780	0.785	0.804	0.783
IB3	79.401	20.600	2.208	0.794	0.794	0.771	0.768	0.794	0.770
IB5	79.245	20.756	1.763	0.792	0.792	0.775	0.760	0.792	0.768
IB7	78.590	21.410	2.070	0.786	0.786	0.774	0.748	0.786	0.760
IB9	77.966	22.034	1.967	0.780	0.780	0.773	0.740	0.780	0.756
IB11	77.216	22.784	2.162	0.772	0.772	0.769	0.733	0.772	0.751
IB13	76.560	23.440	2.051	0.766	0.766	0.759	0.724	0.766	0.741
IB15	75.749	24.251	1.988	0.757	0.757	0.755	0.717	0.757	0.735
IB17	75.094	24.906	1.967	0.751	0.751	0.758	0.711	0.751	0.734
IB19	74.438	25.562	1.985	0.744	0.744	0.751	0.704	0.744	0.726
IB21	73.596	26.404	1.743	0.736	0.736	0.748	0.696	0.736	0.721
IB25	72.815	27.185	1.802	0.728	0.728	0.746	0.688	0.728	0.716
IB29	72.502	27.498	2.113	0.725	0.725	0.745	0.684	0.725	0.714
IB35	71.785	28.215	2.098	0.718	0.718	0.740	0.676	0.718	0.706
IB41	71.410	28.590	2.143	0.714	0.714	0.745	0.672	0.714	0.707
IB49	71.410	28.590	2.462	0.714	0.714	0.742	0.670	0.714	0.704
IB57	71.566	28.434	2.671	0.716	0.716	0.746	0.670	0.716	0.706
IB73	72.065	27.935	2.260	0.721	0.721	0.749	0.674	0.721	0.710
IB89	72.503	27.497	2.277	0.725	0.725	0.756	0.675	0.725	0.713
IB1-WV	80.399	19.601	2.334	0.804	0.804	0.780	0.785	0.804	0.783
IB3-WV	80.555	19.445	1.726	0.806	0.806	0.784	0.780	0.806	0.782
IB5-WV	80.524	19.476	1.906	0.805	0.805	0.785	0.775	0.805	0.780
IB7-WV	79.651	20.349	1.937	0.797	0.797	0.782	0.762	0.797	0.772
IB9-WV	78.964	21.036	2.059	0.790	0.790	0.779	0.753	0.790	0.766
IB11-WV	78.434	21.566	1.892	0.784	0.784	0.775	0.745	0.784	0.760
IB13-WV	77.872	22.128	2.108	0.779	0.779	0.768	0.739	0.779	0.753
IB15-WV	76.873	23.128	1.984	0.769	0.769	0.763	0.730	0.769	0.746
IB17-WV	76.217	23.783	1.890	0.762	0.762	0.763	0.722	0.762	0.742
IB19-WV	75.655	24.345	1.874	0.757	0.757	0.759	0.716	0.757	0.737
IB21-WV	74.875	25.125	2.218	0.749	0.749	0.759	0.709	0.749	0.733
IB25-WV	73.908	26.093	1.968	0.739	0.739	0.753	0.698	0.739	0.724
IB29-WV	73.563	26.437	1.896	0.736	0.736	0.753	0.693	0.736	0.722
IB35-WV	72.815	27.185	1.880	0.728	0.728	0.743	0.685	0.728	0.713
IB41-WV	72.721	27.279	1.933	0.727	0.727	0.749	0.683	0.727	0.714
IB49-WV	72.471	27.529	2.338	0.725	0.725	0.745	0.680	0.725	0.711
IB57-WV	72.471	27.529	2.368	0.725	0.725	0.750	0.678	0.725	0.712
IB73-WV	72.689	27.311	2.344	0.727	0.727	0.752	0.679	0.727	0.714
IB89-WV	72.877	27.123	2.213	0.729	0.729	0.754	0.678	0.729	0.714

3.16 La2s Collection

Table 3.33 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.34 presents the classification results obtained by IBk for La2s collection.

Table 3.33: Classification Results for La2s collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	75.251	24.749	1.973	0.753	0.753	0.736	0.733	0.753	0.734
Multinomial NB	89.917	10.083	1.765	0.899	0.899	0.879	0.879	0.899	0.879
J48	76.845	23.155	1.778	0.768	0.768	0.742	0.739	0.768	0.740
SMO-Linear-0	12.196	87.804	1.318	0.122	0.122	0.020	0.167	0.122	0.036
SMO-Linear- 10^{-5}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Linear- 10^{-4}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Linear- 10^{-3}	29.722	70.278	3.965	0.297	0.297	0.216	0.169	0.297	0.189
SMO-Linear- 10^{-2}	69.983	30.017	2.098	0.700	0.700	0.833	0.597	0.700	0.695
SMO-Linear- 10^{-1}	83.122	16.878	1.771	0.831	0.831	0.832	0.776	0.831	0.803
SMO-Linear- 10^0	86.764	13.236	2.160	0.868	0.868	0.847	0.838	0.868	0.843
SMO-Linear- 10^1	86.406	13.594	2.364	0.864	0.864	0.841	0.835	0.864	0.838
SMO-Linear- 10^2	86.406	13.594	2.364	0.864	0.864	0.841	0.835	0.864	0.838
SMO-Linear- 10^3	86.406	13.594	2.364	0.864	0.864	0.841	0.835	0.864	0.838
SMO-Linear- 10^4	86.406	13.594	2.364	0.864	0.864	0.841	0.835	0.864	0.838
SMO-Linear- 10^5	86.406	13.594	2.364	0.864	0.864	0.841	0.835	0.864	0.838
SMO-Poly-0	12.196	87.804	1.318	0.122	0.122	0.020	0.167	0.122	0.036
SMO-Poly- 10^{-5}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.861	0.076
SMO-Poly- 10^{-4}	29.592	70.408	3.980	0.296	0.296	0.501	0.168	0.296	0.252
SMO-Poly- 10^{-3}	48.195	51.805	4.160	0.482	0.482	0.820	0.356	0.482	0.496
SMO-Poly- 10^{-2}	70.049	29.951	2.660	0.700	0.700	0.767	0.622	0.700	0.687
SMO-Poly- 10^{-1}	77.463	22.537	2.061	0.775	0.775	0.761	0.725	0.775	0.742
SMO-Poly- 10^0	77.950	22.050	2.227	0.780	0.780	0.751	0.732	0.780	0.742
SMO-Poly- 10^1	78.178	21.822	2.107	0.782	0.782	0.751	0.736	0.782	0.743
SMO-Poly- 10^2	78.178	21.822	2.107	0.782	0.782	0.751	0.736	0.782	0.743
SMO-Poly- 10^3	78.178	21.822	2.107	0.782	0.782	0.751	0.736	0.782	0.743
SMO-Poly- 10^4	78.178	21.822	2.107	0.782	0.782	0.751	0.736	0.782	0.743
SMO-Poly- 10^5	78.178	21.822	2.107	0.782	0.782	0.751	0.736	0.782	0.743
SMO-RBF-0	12.196	87.804	1.318	0.122	0.122	0.020	0.167	0.122	0.036
SMO-RBF- 10^{-5}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-4}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-3}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-2}	29.429	70.571	4.027	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-1}	29.462	70.538	4.038	0.295	0.295	0.216	0.167	0.295	0.188
SMO-RBF- 10^0	68.812	31.188	3.012	0.688	0.688	0.771	0.609	0.688	0.681
SMO-RBF- 10^1	80.488	19.513	2.041	0.805	0.805	0.805	0.762	0.805	0.782
SMO-RBF- 10^2	83.088	16.912	1.941	0.831	0.831	0.821	0.795	0.831	0.808
SMO-RBF- 10^3	83.023	16.977	2.041	0.830	0.830	0.821	0.794	0.830	0.807
SMO-RBF- 10^4	83.023	16.977	2.041	0.830	0.830	0.821	0.794	0.830	0.807
SMO-RBF- 10^5	83.023	16.977	2.041	0.830	0.830	0.821	0.794	0.830	0.807

Table 3.34: Classification Results for La2s collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	82.794	17.206	3.358	0.828	0.828	0.805	0.806	0.828	0.806
IB3	80.812	19.188	2.344	0.808	0.808	0.790	0.784	0.808	0.787
IB5	80.942	19.058	1.907	0.809	0.809	0.797	0.785	0.809	0.791
IB7	80.486	19.514	2.335	0.805	0.805	0.801	0.781	0.805	0.791
IB9	79.446	20.554	1.823	0.794	0.794	0.802	0.771	0.794	0.786
IB11	79.479	20.521	1.821	0.795	0.795	0.806	0.767	0.795	0.786
IB13	79.706	20.294	2.101	0.797	0.797	0.808	0.769	0.797	0.788
IB15	78.991	21.009	1.580	0.790	0.790	0.801	0.759	0.790	0.779
IB17	78.861	21.139	2.138	0.789	0.789	0.803	0.759	0.789	0.780
IB19	77.983	22.017	2.232	0.780	0.780	0.796	0.745	0.780	0.770
IB21	77.593	22.408	2.185	0.776	0.776	0.795	0.740	0.776	0.767
IB25	76.943	23.057	2.181	0.769	0.769	0.801	0.736	0.769	0.767
IB29	76.358	23.642	1.922	0.764	0.764	0.791	0.725	0.764	0.757
IB35	76.325	23.675	2.576	0.763	0.763	0.790	0.720	0.763	0.753
IB41	76.650	23.350	2.515	0.767	0.767	0.789	0.719	0.767	0.752
IB49	76.780	23.220	2.589	0.768	0.768	0.787	0.714	0.768	0.749
IB57	76.584	23.416	2.575	0.766	0.766	0.785	0.710	0.766	0.745
IB73	77.105	22.896	2.205	0.771	0.771	0.791	0.715	0.771	0.751
IB89	77.234	22.766	2.551	0.772	0.772	0.792	0.715	0.772	0.752
IB1-WV	82.794	17.206	3.358	0.828	0.828	0.805	0.806	0.828	0.806
IB3-WV	82.048	17.953	2.536	0.820	0.820	0.799	0.795	0.820	0.797
IB5-WV	82.242	17.758	2.436	0.822	0.822	0.808	0.801	0.822	0.804
IB7-WV	81.397	18.603	2.252	0.814	0.814	0.805	0.790	0.814	0.797
IB9-WV	80.584	19.416	2.027	0.806	0.806	0.804	0.783	0.806	0.793
IB11-WV	80.324	19.676	1.829	0.803	0.803	0.806	0.777	0.803	0.791
IB13-WV	80.844	19.156	1.856	0.808	0.808	0.817	0.783	0.808	0.800
IB15-WV	80.162	19.838	1.803	0.802	0.802	0.810	0.774	0.802	0.792
IB17-WV	79.999	20.001	2.223	0.800	0.800	0.808	0.771	0.800	0.789
IB19-WV	79.316	20.684	2.041	0.793	0.793	0.805	0.763	0.793	0.783
IB21-WV	78.796	21.204	2.162	0.788	0.788	0.804	0.756	0.788	0.780
IB25-WV	77.853	22.147	2.051	0.779	0.779	0.802	0.745	0.779	0.772
IB29-WV	77.756	22.244	2.024	0.778	0.778	0.799	0.740	0.778	0.769
IB35-WV	77.529	22.472	2.629	0.775	0.775	0.796	0.731	0.775	0.762
IB41-WV	77.268	22.732	2.320	0.773	0.773	0.792	0.726	0.773	0.758
IB49-WV	77.528	22.472	2.681	0.775	0.775	0.795	0.723	0.775	0.757
IB57-WV	77.397	22.603	2.631	0.774	0.774	0.790	0.720	0.774	0.754
IB73-WV	77.723	22.278	2.121	0.777	0.777	0.793	0.723	0.777	0.756
IB89-WV	77.819	22.181	2.261	0.778	0.778	0.798	0.723	0.778	0.758

3.17 LATimes Collection

Table 3.35 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.36 presents the classification results obtained by IBk for LATimes collection.

Table 3.35: Classification Results for LATimes collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	70.760	29.240	1.698	0.708	0.708	0.680	0.688	0.708	0.684
Multinomial NB	84.998	15.002	1.521	0.850	0.850	0.821	0.835	0.850	0.828
J48	75.681	24.319	1.859	0.757	0.757	0.725	0.726	0.757	0.726
SMO-Linear-0	11.610	88.390	1.702	0.116	0.116	0.019	0.167	0.116	0.035
SMO-Linear- 10^{-5}	29.431	70.569	1.697	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Linear- 10^{-4}	29.431	70.569	1.697	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Linear- 10^{-3}	29.447	70.553	1.701	0.294	0.294	0.216	0.167	0.294	0.188
SMO-Linear- 10^{-2}	51.600	48.400	2.265	0.516	0.516	0.785	0.396	0.516	0.527
SMO-Linear- 10^{-1}	70.887	29.113	1.507	0.709	0.709	0.777	0.651	0.709	0.709
SMO-Linear- 10^0	81.112	18.888	1.553	0.811	0.811	0.785	0.777	0.811	0.781
SMO-Linear- 10^1	77.719	22.281	1.478	0.777	0.777	0.743	0.759	0.777	0.751
SMO-Linear- 10^2	73.371	26.629	1.622	0.734	0.734	0.699	0.720	0.734	0.709
SMO-Linear- 10^3	71.285	28.715	1.699	0.713	0.713	0.680	0.702	0.713	0.691
SMO-Linear- 10^4	70.489	29.511	2.051	0.705	0.705	0.676	0.694	0.705	0.684
SMO-Linear- 10^5	70.505	29.495	2.056	0.705	0.705	0.676	0.694	0.705	0.685
SMO-Poly-0	11.610	88.390	1.702	0.116	0.116	0.019	0.167	0.116	0.035
SMO-Poly- 10^{-5}	29.431	70.569	1.679	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Poly- 10^{-4}	29.431	70.569	1.679	0.294	0.294	0.049	0.167	0.294	0.076
SMO-Poly- 10^{-3}	32.075	67.925	1.605	0.321	0.321	0.833	0.190	0.321	0.309
SMO-Poly- 10^{-2}	56.490	43.510	1.839	0.565	0.565	0.755	0.479	0.565	0.586
SMO-Poly- 10^{-1}	67.543	32.457	0.814	0.675	0.675	0.748	0.621	0.675	0.678
SMO-Poly- 10^0	76.891	23.109	2.217	0.769	0.769	0.736	0.740	0.769	0.738
SMO-Poly- 10^1	72.782	27.218	1.513	0.728	0.728	0.692	0.710	0.728	0.701
SMO-Poly- 10^2	68.801	31.200	1.705	0.688	0.688	0.653	0.677	0.688	0.665
SMO-Poly- 10^3	66.874	33.127	2.095	0.669	0.669	0.638	0.662	0.669	0.650
SMO-Poly- 10^4	65.536	34.464	2.024	0.655	0.655	0.628	0.651	0.655	0.640
SMO-Poly- 10^5	65.536	34.464	2.019	0.655	0.655	0.628	0.651	0.655	0.639
SMO-RBF-0	11.610	88.390	1.702	0.116	0.116	0.019	0.167	0.116	0.035
SMO-RBF- 10^{-5}	29.431	70.569	1.697	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-4}	29.431	70.569	1.697	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-3}	29.431	70.569	1.605	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-2}	29.431	70.569	1.839	0.294	0.294	0.049	0.167	0.294	0.076
SMO-RBF- 10^{-1}	29.558	70.442	0.814	0.296	0.296	0.216	0.168	0.296	0.189
SMO-RBF- 10^0	56.824	43.176	2.217	0.568	0.568	0.713	0.480	0.568	0.574
SMO-RBF- 10^1	72.655	27.345	1.513	0.727	0.727	0.755	0.670	0.727	0.710
SMO-RBF- 10^2	78.293	21.707	1.705	0.783	0.783	0.763	0.751	0.783	0.757
SMO-RBF- 10^3	75.044	24.957	2.095	0.750	0.750	0.723	0.723	0.750	0.723
SMO-RBF- 10^4	72.161	27.839	2.024	0.722	0.722	0.692	0.697	0.722	0.695
SMO-RBF- 10^5	70.569	29.431	2.019	0.706	0.706	0.680	0.689	0.706	0.684

Table 3.36: Classification Results for LATimes collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	76.828	23.172	2.097	0.768	0.768	0.748	0.751	0.768	0.749
IB3	73.866	26.134	2.248	0.739	0.739	0.732	0.724	0.739	0.728
IB5	72.082	27.918	2.267	0.721	0.721	0.724	0.707	0.721	0.715
IB7	71.366	28.634	2.301	0.714	0.714	0.722	0.694	0.714	0.708
IB9	71.015	28.985	2.272	0.710	0.710	0.730	0.687	0.710	0.708
IB11	69.773	30.227	2.079	0.698	0.698	0.724	0.724	0.698	0.724
IB13	68.626	31.374	2.107	0.686	0.686	0.726	0.661	0.686	0.692
IB15	67.878	32.122	2.151	0.679	0.679	0.721	0.653	0.679	0.685
IB17	67.257	32.743	2.393	0.673	0.673	0.724	0.648	0.673	0.684
IB19	66.556	33.444	2.179	0.666	0.666	0.722	0.639	0.666	0.678
IB21	66.110	33.890	2.142	0.661	0.661	0.723	0.633	0.661	0.675
IB25	64.215	35.786	2.304	0.642	0.642	0.717	0.609	0.642	0.659
IB29	63.355	36.646	2.009	0.634	0.634	0.718	0.597	0.634	0.652
IB35	61.858	38.143	2.181	0.619	0.619	0.715	0.581	0.619	0.641
IB41	60.552	39.448	2.068	0.606	0.606	0.712	0.564	0.606	0.630
IB49	59.405	40.595	1.943	0.594	0.594	0.721	0.551	0.594	0.625
IB57	59.023	40.977	2.111	0.590	0.590	0.722	0.545	0.590	0.621
IB73	58.879	41.121	2.187	0.589	0.589	0.729	0.540	0.589	0.620
IB89	59.596	40.404	2.401	0.596	0.596	0.740	0.543	0.596	0.626
IB1-WV	76.828	23.172	2.097	0.768	0.768	0.748	0.751	0.768	0.749
IB3-WV	74.710	25.290	2.334	0.747	0.747	0.742	0.730	0.747	0.736
IB5-WV	73.452	26.548	2.011	0.735	0.735	0.739	0.716	0.735	0.727
IB7-WV	72.433	27.567	2.026	0.724	0.724	0.734	0.703	0.724	0.718
IB9-WV	71.859	28.141	2.148	0.719	0.719	0.737	0.696	0.719	0.716
IB11-WV	70.601	29.399	2.246	0.706	0.706	0.735	0.683	0.706	0.708
IB13-WV	69.407	30.593	2.206	0.694	0.694	0.731	0.668	0.694	0.698
IB15-WV	68.594	31.406	2.131	0.686	0.686	0.726	0.659	0.686	0.691
IB17-WV	67.910	32.091	2.202	0.679	0.679	0.727	0.653	0.679	0.688
IB19-WV	67.272	32.728	2.152	0.673	0.673	0.726	0.645	0.673	0.683
IB21-WV	66.763	33.237	2.082	0.668	0.668	0.724	0.639	0.668	0.679
IB25-WV	65.027	34.973	2.127	0.650	0.650	0.725	0.617	0.650	0.667
IB29-WV	64.087	35.913	2.155	0.641	0.641	0.726	0.605	0.641	0.660
IB35-WV	62.622	37.378	2.122	0.626	0.626	0.724	0.588	0.626	0.649
IB41-WV	61.428	38.572	2.133	0.614	0.614	0.725	0.573	0.614	0.640
IB49-WV	60.424	39.576	2.065	0.604	0.604	0.729	0.560	0.604	0.634
IB57-WV	59.978	40.022	2.082	0.600	0.600	0.732	0.555	0.600	0.631
IB73-WV	60.106	39.895	2.222	0.601	0.601	0.737	0.552	0.601	0.631
IB89-WV	60.711	39.289	2.371	0.607	0.607	0.744	0.555	0.607	0.636

3.18 Multi-Domain-Sentiment Collection

Table 3.37 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.38 presents the classification results obtained by IBk for Multi-Domain-Sentiment collection.

Table 3.37: Classification Results for Multi-Domain-Sentiment collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	72.850	27.150	0.000	0.729	0.729	0.729	0.729	0.729	0.729
Multinomial NB	78.650	21.350	0.000	0.787	0.787	0.787	0.787	0.787	0.787
J48	74.950	25.050	0.750	0.750	0.750	0.750	0.750	0.750	0.750
SMO-Linear-0	50.000	50.000	1.365	0.500	0.500	0.250	0.500	0.500	0.333
SMO-Linear- 10^{-5}	49.175	50.825	1.609	0.492	0.492	0.491	0.492	0.492	0.492
SMO-Linear- 10^{-4}	51.625	48.375	0.889	0.516	0.516	0.516	0.516	0.516	0.516
SMO-Linear- 10^{-3}	48.800	51.200	0.650	0.488	0.488	0.488	0.488	0.488	0.488
SMO-Linear- 10^{-2}	79.875	20.125	1.719	0.799	0.799	0.801	0.799	0.799	0.800
SMO-Linear- 10^{-1}	81.738	18.263	0.959	0.817	0.817	0.818	0.818	0.817	0.818
SMO-Linear- 10^0	80.175	19.825	1.331	0.802	0.802	0.802	0.802	0.802	0.802
SMO-Linear- 10^1	77.138	22.863	1.600	0.771	0.771	0.772	0.771	0.771	0.772
SMO-Linear- 10^2	76.225	23.775	1.744	0.762	0.762	0.763	0.762	0.762	0.762
SMO-Linear- 10^3	76.188	23.813	1.726	0.762	0.762	0.762	0.762	0.762	0.762
SMO-Linear- 10^4	76.188	23.813	1.726	0.762	0.762	0.762	0.762	0.762	0.762
SMO-Linear- 10^5	76.188	23.813	1.726	0.762	0.762	0.762	0.762	0.762	0.762
SMO-Poly-0	50.000	50.000	1.365	0.500	0.500	0.250	0.500	0.500	0.333
SMO-Poly- 10^{-5}	48.800	51.200	0.650	0.488	0.488	0.244	0.500	0.488	0.328
SMO-Poly- 10^{-4}	48.825	51.175	0.692	0.488	0.488	0.394	0.500	0.488	0.441
SMO-Poly- 10^{-3}	53.425	46.575	5.365	0.534	0.534	0.686	0.540	0.534	0.604
SMO-Poly- 10^{-2}	67.038	32.963	5.120	0.670	0.670	0.731	0.672	0.670	0.700
SMO-Poly- 10^{-1}	71.638	28.363	1.985	0.716	0.716	0.734	0.717	0.716	0.725
SMO-Poly- 10^0	73.400	26.600	1.279	0.734	0.734	0.740	0.734	0.734	0.737
SMO-Poly- 10^1	72.788	27.213	1.286	0.728	0.728	0.731	0.728	0.728	0.729
SMO-Poly- 10^2	73.688	26.313	1.434	0.737	0.737	0.739	0.737	0.737	0.738
SMO-Poly- 10^3	73.538	26.463	2.005	0.735	0.735	0.737	0.735	0.735	0.736
SMO-Poly- 10^4	73.538	26.463	2.005	0.735	0.735	0.737	0.735	0.735	0.736
SMO-Poly- 10^5	73.538	26.463	2.005	0.735	0.735	0.737	0.735	0.735	0.736
SMO-RBF-0	50.000	50.000	1.365	0.500	0.500	0.250	0.500	0.500	0.333
SMO-RBF- 10^{-5}	49.600	50.400	2.846	0.496	0.496	0.286	0.508	0.496	0.366
SMO-RBF- 10^{-4}	49.600	50.400	2.846	0.496	0.496	0.286	0.508	0.496	0.366
SMO-RBF- 10^{-3}	49.450	50.550	2.405	0.495	0.495	0.286	0.507	0.495	0.365
SMO-RBF- 10^{-2}	49.450	50.550	2.405	0.495	0.495	0.286	0.507	0.495	0.365
SMO-RBF- 10^{-1}	51.300	48.700	7.192	0.513	0.513	0.407	0.524	0.513	0.458
SMO-RBF- 10^0	81.025	18.975	0.882	0.810	0.810	0.812	0.811	0.810	0.811
SMO-RBF- 10^1	81.900	18.100	1.098	0.819	0.819	0.820	0.819	0.819	0.819
SMO-RBF- 10^2	79.450	20.550	1.429	0.795	0.795	0.795	0.795	0.795	0.795
SMO-RBF- 10^3	77.188	22.813	1.343	0.772	0.772	0.772	0.772	0.772	0.772
SMO-RBF- 10^4	76.625	23.375	1.908	0.766	0.766	0.767	0.766	0.766	0.766
SMO-RBF- 10^5	76.625	23.375	1.908	0.766	0.766	0.767	0.766	0.766	0.766

Table 3.38: Classification Results for Multi-Domain-Sentiment collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	61.863	38.138	2.046	0.619	0.619	0.619	0.619	0.619	0.619
IB3	63.875	36.125	1.398	0.639	0.639	0.639	0.639	0.639	0.639
IB5	64.625	35.375	1.187	0.646	0.646	0.647	0.646	0.646	0.647
IB7	66.075	33.925	1.234	0.661	0.661	0.661	0.661	0.661	0.661
IB9	66.600	33.400	1.383	0.666	0.666	0.666	0.666	0.666	0.666
IB11	66.788	33.213	1.543	0.668	0.668	0.668	0.668	0.668	0.668
IB13	67.188	32.813	1.478	0.672	0.672	0.672	0.672	0.672	0.672
IB15	67.025	32.975	1.519	0.670	0.670	0.671	0.670	0.670	0.670
IB17	67.663	32.338	1.224	0.677	0.677	0.677	0.677	0.677	0.677
IB19	67.913	32.088	1.003	0.679	0.679	0.679	0.679	0.679	0.679
IB21	68.363	31.638	0.839	0.684	0.684	0.684	0.684	0.684	0.684
IB25	68.175	31.825	0.915	0.682	0.682	0.682	0.682	0.682	0.682
IB29	69.050	30.950	1.309	0.691	0.691	0.691	0.691	0.691	0.691
IB35	69.288	30.713	0.855	0.693	0.693	0.693	0.693	0.693	0.693
IB41	69.388	30.613	1.253	0.694	0.694	0.694	0.694	0.694	0.694
IB49	70.025	29.975	1.375	0.700	0.700	0.700	0.700	0.700	0.700
IB57	70.500	29.500	1.137	0.705	0.705	0.705	0.705	0.705	0.705
IB73	71.000	29.000	1.288	0.710	0.710	0.711	0.710	0.710	0.710
IB89	70.800	29.200	1.465	0.708	0.708	0.709	0.708	0.708	0.708
IB1-WV	61.863	38.138	2.046	0.619	0.619	0.619	0.619	0.619	0.619
IB3-WV	64.650	35.350	1.218	0.647	0.647	0.647	0.647	0.647	0.647
IB5-WV	65.275	34.725	1.226	0.653	0.653	0.654	0.653	0.653	0.653
IB7-WV	66.713	33.288	1.204	0.667	0.667	0.668	0.667	0.667	0.667
IB9-WV	67.263	32.738	1.377	0.673	0.673	0.673	0.673	0.673	0.673
IB11-WV	67.463	32.538	1.499	0.675	0.675	0.675	0.675	0.675	0.675
IB13-WV	67.988	32.013	1.459	0.680	0.680	0.680	0.680	0.680	0.680
IB15-WV	67.838	32.163	1.342	0.678	0.678	0.679	0.678	0.678	0.678
IB17-WV	68.438	31.563	0.952	0.684	0.684	0.685	0.684	0.684	0.684
IB19-WV	68.613	31.388	0.951	0.686	0.686	0.686	0.686	0.686	0.686
IB21-WV	69.000	31.000	0.871	0.690	0.690	0.690	0.690	0.690	0.690
IB25-WV	68.825	31.175	0.941	0.688	0.688	0.688	0.688	0.688	0.688
IB29-WV	69.525	30.475	1.317	0.695	0.695	0.695	0.695	0.695	0.695
IB35-WV	69.788	30.213	0.983	0.698	0.698	0.698	0.698	0.698	0.698
IB41-WV	69.863	30.138	1.322	0.699	0.699	0.699	0.699	0.699	0.699
IB49-WV	70.313	29.688	1.322	0.703	0.703	0.703	0.703	0.703	0.703
IB57-WV	70.813	29.188	1.094	0.708	0.708	0.708	0.708	0.708	0.708
IB73-WV	71.400	28.600	1.400	0.714	0.714	0.715	0.714	0.714	0.714
IB89-WV	71.063	28.938	1.368	0.711	0.711	0.712	0.710	0.711	0.711

3.19 New3 Collection

Table 3.39 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.40 presents the classification results obtained by IBk for New3 collection.

Table 3.39: Classification Results for New3 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
Naïve Bayes	56.727	43.272	0.938	0.567	0.567	0.589	0.576	0.5672	0.582
Multinomial NB	79.169	20.830	1.225	0.791	0.791	0.824	0.778	0.7916	0.801
J48	70.851	29.148	1.427	0.708	0.708	0.708	0.699	0.7085	0.704
SMO-Linear-0	1.318	98.681	0.315	0.013	0.013	0.000	0.022	0.0131	0.000
SMO-Linear- 10^{-5}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-Linear- 10^{-4}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-Linear- 10^{-3}	7.323	92.676	0.884	0.073	0.073	0.030	0.023	0.0732	0.026
SMO-Linear- 10^{-2}	26.020	73.979	1.558	0.260	0.260	0.535	0.152	0.2602	0.237
SMO-Linear- 10^{-1}	60.378	39.621	1.280	0.603	0.603	0.776	0.528	0.6037	0.629
SMO-Linear- 10^0	71.939	28.060	1.310	0.719	0.719	0.758	0.687	0.7193	0.721
SMO-Linear- 10^1	71.918	28.081	1.810	0.719	0.719	0.741	0.696	0.7191	0.718
SMO-Linear- 10^2	71.102	28.897	1.554	0.711	0.711	0.735	0.690	0.7110	0.712
SMO-Linear- 10^3	71.113	28.886	1.583	0.711	0.711	0.735	0.690	0.7111	0.712
SMO-Linear- 10^4	71.113	28.886	1.583	0.711	0.711	0.735	0.690	0.7111	0.712
SMO-Linear- 10^5	71.113	28.886	1.583	0.711	0.711	0.735	0.690	0.7111	0.712
SMO-Poly-0	1.318	98.681	0.315	0.013	0.013	0.000	0.022	0.0131	0.000
SMO-Poly- 10^{-5}	7.334	92.665	0.881	0.073	0.073	0.069	0.023	0.0733	0.034
SMO-Poly- 10^{-4}	7.417	92.582	0.915	0.074	0.074	0.076	0.024	0.0741	0.037
SMO-Poly- 10^{-3}	8.809	91.190	0.959	0.088	0.088	0.455	0.035	0.0880	0.065
SMO-Poly- 10^{-2}	32.349	67.650	1.688	0.323	0.323	0.705	0.232	0.3234	0.349
SMO-Poly- 10^{-1}	48.545	51.454	1.508	0.485	0.485	0.705	0.423	0.4854	0.529
SMO-Poly- 10^0	52.667	47.332	1.590	0.526	0.526	0.655	0.493	0.5266	0.562
SMO-Poly- 10^1	53.797	46.202	1.320	0.537	0.537	0.645	0.509	0.5379	0.569
SMO-Poly- 10^2	54.498	45.501	1.860	0.544	0.544	0.632	0.512	0.5449	0.565
SMO-Poly- 10^3	53.891	46.108	2.007	0.538	0.538	0.625	0.506	0.5389	0.560
SMO-Poly- 10^4	53.661	46.338	1.919	0.536	0.536	0.631	0.504	0.5366	0.561
SMO-Poly- 10^5	53.661	46.338	1.919	0.536	0.536	0.631	0.504	0.5366	0.561
SMO-RBF-0	1.318	98.681	0.315	0.013	0.013	0.000	0.022	0.0131	0.000
SMO-RBF- 10^{-5}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-RBF- 10^{-4}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-RBF- 10^{-3}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-RBF- 10^{-2}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-RBF- 10^{-1}	7.281	92.718	0.899	0.072	0.072	0.001	0.022	0.0728	0.003
SMO-RBF- 10^0	29.849	70.150	1.463	0.298	0.298	0.597	0.192	0.2984	0.291
SMO-RBF- 10^1	63.601	36.398	1.673	0.636	0.636	0.786	0.579	0.6360	0.667
SMO-RBF- 10^2	69.251	30.749	1.583	0.692	0.692	0.768	0.659	0.6925	0.709
SMO-RBF- 10^3	68.937	31.062	1.830	0.689	0.689	0.753	0.662	0.6893	0.705
SMO-RBF- 10^4	68.466	31.533	1.727	0.684	0.684	0.751	0.658	0.6846	0.701
SMO-RBF- 10^5	68.466	31.533	1.727	0.684	0.684	0.751	0.658	0.6846	0.701

Table 3.40: Classification Results for New3 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	79.263	20.736	0.765	0.792	0.792	0.798	0.789	0.7926	0.794
IB3	78.489	21.510	1.006	0.784	0.784	0.805	0.776	0.7848	0.790
IB5	79.378	20.621	1.109	0.793	0.793	0.810	0.782	0.7937	0.796
IB7	79.127	20.872	1.130	0.791	0.791	0.813	0.779	0.7912	0.796
IB9	78.970	21.029	0.899	0.789	0.789	0.814	0.776	0.7897	0.795
IB11	78.813	21.186	0.874	0.788	0.788	0.817	0.773	0.7881	0.794
IB13	78.562	21.437	1.007	0.785	0.785	0.816	0.769	0.7856	0.792
IB15	78.154	21.845	1.208	0.781	0.781	0.812	0.765	0.7815	0.788
IB17	77.976	22.023	1.323	0.779	0.779	0.816	0.764	0.7797	0.789
IB19	77.505	22.494	1.250	0.775	0.775	0.809	0.758	0.7750	0.782
IB21	77.421	22.578	1.331	0.774	0.774	0.813	0.757	0.7742	0.784
IB25	77.244	22.756	1.347	0.772	0.772	0.809	0.755	0.7724	0.781
IB29	76.773	23.226	1.507	0.767	0.767	0.809	0.749	0.7677	0.778
IB35	76.333	23.666	1.375	0.763	0.763	0.810	0.740	0.7633	0.773
IB41	75.925	24.074	1.217	0.759	0.759	0.809	0.735	0.7592	0.770
IB49	75.590	24.409	1.432	0.755	0.755	0.810	0.732	0.7559	0.769
IB57	75.423	24.576	1.621	0.754	0.754	0.811	0.727	0.7542	0.766
IB73	74.492	25.507	1.452	0.744	0.744	0.805	0.713	0.7449	0.756
IB89	73.969	26.030	1.430	0.739	0.739	0.799	0.710	0.7396	0.752
IB1-WV	79.263	20.736	0.765	0.792	0.792	0.798	0.789	0.7926	0.794
IB3-WV	80.393	19.606	1.005	0.803	0.803	0.810	0.796	0.8039	0.803
IB5-WV	81.031	18.968	0.887	0.810	0.810	0.821	0.801	0.8103	0.811
IB7-WV	81.073	18.926	0.919	0.810	0.810	0.828	0.800	0.8107	0.813
IB9-WV	81.073	18.926	0.919	0.810	0.810	0.824	0.799	0.8107	0.812
IB11-WV	80.550	19.449	0.922	0.805	0.805	0.826	0.791	0.8055	0.808
IB13-WV	80.246	19.753	0.791	0.802	0.802	0.827	0.787	0.8024	0.807
IB15-WV	79.817	20.182	1.042	0.798	0.798	0.824	0.782	0.7981	0.803
IB17-WV	79.556	20.443	1.024	0.795	0.795	0.826	0.781	0.7955	0.803
IB19-WV	79.116	20.883	1.017	0.791	0.791	0.820	0.775	0.7911	0.797
IB21-WV	78.792	21.207	1.147	0.787	0.787	0.819	0.770	0.7879	0.793
IB25-WV	78.468	21.531	1.219	0.784	0.784	0.819	0.766	0.7846	0.792
IB29-WV	78.112	21.887	1.505	0.781	0.781	0.820	0.761	0.7811	0.789
IB35-WV	77.442	22.557	1.407	0.774	0.774	0.817	0.751	0.7744	0.782
IB41-WV	77.223	22.776	1.347	0.772	0.772	0.818	0.748	0.7722	0.781
IB49-WV	76.616	23.383	1.340	0.766	0.766	0.819	0.741	0.7661	0.778
IB57-WV	76.375	23.624	1.339	0.763	0.763	0.817	0.736	0.7637	0.775
IB73-WV	75.528	24.472	1.502	0.755	0.755	0.814	0.726	0.7552	0.767
IB89-WV	75.057	24.942	1.435	0.750	0.750	0.812	0.721	0.7505	0.764

3.20 NSF Collection

Table 3.41 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.42 presents the classification results obtained by IBk for NSF collection.

Table 3.41: Classification Results for NSF collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	70.866	29.134	1.157	0.709	0.709	0.704	0.726	0.709	0.715
Multinomial NB	83.846	16.154	1.065	0.838	0.838	0.856	0.815	0.838	0.835
J48	70.743	29.257	1.585	0.707	0.707	0.715	0.680	0.707	0.697
SMO-Linear-0	12.724	87.277	1.100	0.127	0.127	0.008	0.063	0.127	0.014
SMO-Linear- 10^{-5}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-Linear- 10^{-4}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-Linear- 10^{-3}	13.398	86.602	0.809	0.134	0.134	0.071	0.063	0.134	0.066
SMO-Linear- 10^{-2}	37.638	62.362	0.991	0.376	0.376	0.727	0.215	0.376	0.331
SMO-Linear- 10^{-1}	78.858	21.142	1.160	0.789	0.789	0.848	0.741	0.789	0.791
SMO-Linear- 10^0	81.880	18.121	1.106	0.819	0.819	0.831	0.803	0.819	0.817
SMO-Linear- 10^1	79.627	20.373	1.434	0.796	0.796	0.804	0.784	0.796	0.794
SMO-Linear- 10^2	78.592	21.408	1.638	0.786	0.786	0.783	0.775	0.786	0.779
SMO-Linear- 10^3	78.592	21.408	1.638	0.786	0.786	0.783	0.775	0.786	0.779
SMO-Linear- 10^4	78.592	21.408	1.638	0.786	0.786	0.783	0.775	0.786	0.779
SMO-Linear- 10^5	78.592	21.408	1.638	0.786	0.786	0.783	0.775	0.786	0.779
SMO-Poly-0	12.724	86.612	1.100	0.127	0.127	0.008	0.063	0.134	0.015
SMO-Poly- 10^{-5}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-Poly- 10^{-4}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-Poly- 10^{-3}	13.464	86.536	0.829	0.135	0.135	0.071	0.063	0.135	0.067
SMO-Poly- 10^{-2}	35.975	64.025	1.429	0.360	0.360	0.889	0.224	0.360	0.358
SMO-Poly- 10^{-1}	72.102	27.898	1.700	0.721	0.721	0.851	0.664	0.721	0.746
SMO-Poly- 10^0	76.207	23.793	0.983	0.762	0.762	0.826	0.723	0.762	0.771
SMO-Poly- 10^1	75.827	24.174	0.839	0.758	0.758	0.809	0.723	0.758	0.763
SMO-Poly- 10^2	74.905	25.095	0.961	0.749	0.749	0.770	0.718	0.749	0.743
SMO-Poly- 10^3	74.905	25.095	0.961	0.749	0.749	0.770	0.718	0.749	0.743
SMO-Poly- 10^4	74.905	25.095	0.961	0.749	0.749	0.770	0.718	0.749	0.743
SMO-Poly- 10^5	74.905	25.095	0.961	0.749	0.749	0.770	0.718	0.749	0.743
SMO-RBF-0	12.724	86.612	1.100	0.127	0.127	0.008	0.063	0.134	0.015
SMO-RBF- 10^{-5}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-RBF- 10^{-4}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-RBF- 10^{-3}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-RBF- 10^{-2}	13.388	86.612	0.800	0.134	0.134	0.008	0.063	0.134	0.015
SMO-RBF- 10^{-1}	16.619	83.381	0.891	0.166	0.166	0.068	0.079	0.166	0.073
SMO-RBF- 10^0	56.595	43.406	1.906	0.566	0.566	0.781	0.427	0.566	0.552
SMO-RBF- 10^1	81.823	18.178	1.286	0.818	0.818	0.854	0.791	0.818	0.821
SMO-RBF- 10^2	81.642	18.358	1.164	0.816	0.816	0.826	0.803	0.816	0.814
SMO-RBF- 10^3	79.590	20.410	1.693	0.796	0.796	0.797	0.782	0.796	0.790
SMO-RBF- 10^4	79.162	20.838	1.426	0.792	0.792	0.789	0.779	0.792	0.784
SMO-RBF- 10^5	79.162	20.838	1.426	0.792	0.792	0.789	0.779	0.792	0.784

Table 3.42: Classification Results for NSF collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	75.456	24.544	0.783	0.755	0.755	0.748	0.740	0.755	0.744
IB3	75.608	24.392	1.220	0.756	0.756	0.775	0.735	0.756	0.755
IB5	76.007	23.993	1.599	0.760	0.760	0.780	0.737	0.760	0.758
IB7	76.567	23.433	1.512	0.766	0.766	0.789	0.742	0.766	0.765
IB9	76.701	23.299	1.281	0.767	0.767	0.791	0.742	0.767	0.765
IB11	76.169	23.831	1.276	0.762	0.762	0.788	0.733	0.762	0.759
IB13	75.941	24.060	0.948	0.759	0.759	0.786	0.726	0.759	0.754
IB15	75.342	24.658	1.043	0.753	0.753	0.782	0.718	0.753	0.749
IB17	74.962	25.038	1.127	0.750	0.750	0.782	0.714	0.750	0.747
IB19	74.572	25.428	1.230	0.746	0.746	0.780	0.708	0.746	0.742
IB21	74.487	25.513	1.124	0.745	0.745	0.780	0.706	0.745	0.741
IB25	73.822	26.179	1.287	0.738	0.738	0.776	0.698	0.738	0.735
IB29	73.327	26.673	1.397	0.733	0.733	0.774	0.688	0.733	0.729
IB35	72.681	27.319	1.243	0.727	0.727	0.770	0.679	0.727	0.721
IB41	72.187	27.813	1.385	0.722	0.722	0.770	0.671	0.722	0.717
IB49	71.019	28.982	1.399	0.710	0.710	0.762	0.657	0.710	0.706
IB57	70.135	29.865	1.176	0.701	0.701	0.759	0.645	0.701	0.697
IB73	69.071	30.930	1.284	0.691	0.691	0.761	0.630	0.691	0.690
IB89	68.082	31.918	1.282	0.681	0.681	0.760	0.615	0.681	0.680
IB1-WV	75.456	24.544	0.783	0.755	0.755	0.748	0.740	0.755	0.744
IB3-WV	77.527	22.473	0.791	0.775	0.775	0.777	0.758	0.775	0.767
IB5-WV	78.164	21.836	1.258	0.782	0.782	0.791	0.762	0.782	0.776
IB7-WV	78.829	21.171	1.259	0.788	0.788	0.801	0.768	0.788	0.784
IB9-WV	78.886	21.114	1.142	0.789	0.789	0.802	0.766	0.789	0.783
IB11-WV	78.354	21.646	0.951	0.784	0.784	0.798	0.757	0.784	0.777
IB13-WV	78.183	21.817	0.886	0.782	0.782	0.799	0.752	0.782	0.775
IB15-WV	77.898	22.102	0.919	0.779	0.779	0.800	0.747	0.779	0.772
IB17-WV	77.575	22.425	1.008	0.776	0.776	0.800	0.744	0.776	0.771
IB19-WV	77.129	22.872	0.928	0.771	0.771	0.798	0.738	0.771	0.767
IB21-WV	76.805	23.195	0.935	0.768	0.768	0.796	0.732	0.768	0.763
IB25-WV	76.207	23.793	1.218	0.762	0.762	0.794	0.724	0.762	0.757
IB29-WV	75.665	24.335	1.358	0.757	0.757	0.792	0.715	0.757	0.752
IB35-WV	74.715	25.285	1.317	0.747	0.747	0.786	0.704	0.747	0.742
IB41-WV	74.097	25.903	1.261	0.741	0.741	0.784	0.693	0.741	0.736
IB49-WV	73.100	26.901	1.276	0.731	0.731	0.779	0.682	0.731	0.727
IB57-WV	72.149	27.851	1.096	0.721	0.721	0.774	0.669	0.721	0.718
IB73-WV	70.819	29.181	1.237	0.708	0.708	0.774	0.650	0.708	0.707
IB89-WV	69.774	30.226	1.328	0.698	0.698	0.773	0.634	0.698	0.697

3.21 Oh0 Collection

Table 3.43 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.44 presents the classification results obtained by IBk for Oh0 collection.

Table 3.43: Classification Results for Oh0 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	79.666	20.334	3.956	0.797	0.797	0.753	0.767	0.797	0.760
Multinomial NB	89.835	10.165	1.963	0.898	0.898	0.876	0.874	0.898	0.875
J48	80.953	19.047	2.321	0.810	0.810	0.824	0.793	0.810	0.808
SMO-Linear-0	5.684	94.316	1.272	0.057	0.057	0.006	0.100	0.057	0.011
SMO-Linear- 10^{-5}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-Linear- 10^{-4}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-Linear- 10^{-3}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-Linear- 10^{-2}	33.405	66.595	4.015	0.334	0.334	0.484	0.183	0.334	0.265
SMO-Linear- 10^{-1}	81.259	18.741	3.066	0.813	0.813	0.846	0.741	0.813	0.790
SMO-Linear- 10^0	81.555	18.445	2.864	0.816	0.816	0.820	0.760	0.816	0.789
SMO-Linear- 10^1	81.555	18.445	2.864	0.816	0.816	0.820	0.760	0.816	0.789
SMO-Linear- 10^2	81.555	18.445	2.864	0.816	0.816	0.820	0.760	0.816	0.789
SMO-Linear- 10^3	81.555	18.445	2.864	0.816	0.816	0.820	0.760	0.816	0.789
SMO-Linear- 10^4	81.555	18.445	2.864	0.816	0.816	0.820	0.760	0.816	0.789
SMO-Linear- 10^5	81.555	18.445	2.864	0.816	0.816	0.820	0.760	0.816	0.789
SMO-Poly-0	5.684	94.316	1.272	0.057	0.057	0.006	0.100	0.057	0.011
SMO-Poly- 10^{-5}	20.042	79.959	5.074	0.200	0.200	0.041	0.104	0.200	0.059
SMO-Poly- 10^{-4}	20.934	79.066	4.990	0.209	0.209	0.104	0.109	0.209	0.106
SMO-Poly- 10^{-3}	19.640	80.360	3.308	0.196	0.196	0.079	0.102	0.196	0.089
SMO-Poly- 10^{-2}	46.964	53.036	5.702	0.470	0.470	0.806	0.328	0.470	0.466
SMO-Poly- 10^{-1}	52.045	47.956	4.094	0.520	0.520	0.723	0.398	0.520	0.513
SMO-Poly- 10^0	52.639	47.361	4.202	0.526	0.526	0.709	0.403	0.526	0.514
SMO-Poly- 10^1	52.639	47.361	4.202	0.526	0.526	0.709	0.403	0.526	0.514
SMO-Poly- 10^2	52.639	47.361	4.202	0.526	0.526	0.709	0.403	0.526	0.514
SMO-Poly- 10^3	52.639	47.361	4.202	0.526	0.526	0.709	0.403	0.526	0.514
SMO-Poly- 10^4	52.639	47.361	4.202	0.526	0.526	0.709	0.403	0.526	0.514
SMO-Poly- 10^5	52.639	47.361	4.202	0.526	0.526	0.709	0.403	0.526	0.514
SMO-RBF-0	5.684	94.316	1.272	0.057	0.057	0.006	0.100	0.057	0.011
SMO-RBF- 10^{-5}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-RBF- 10^{-4}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-RBF- 10^{-3}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-RBF- 10^{-2}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-RBF- 10^{-1}	19.342	80.659	3.408	0.193	0.193	0.019	0.100	0.193	0.032
SMO-RBF- 10^0	45.572	54.428	3.395	0.456	0.456	0.634	0.301	0.456	0.408
SMO-RBF- 10^1	80.061	19.939	2.884	0.801	0.801	0.807	0.744	0.801	0.774
SMO-RBF- 10^2	79.760	20.240	3.318	0.798	0.798	0.806	0.744	0.798	0.774
SMO-RBF- 10^3	79.760	20.240	3.318	0.798	0.798	0.806	0.744	0.798	0.774
SMO-RBF- 10^4	79.760	20.240	3.318	0.798	0.798	0.806	0.744	0.798	0.774
SMO-RBF- 10^5	79.760	20.240	3.318	0.798	0.798	0.806	0.744	0.798	0.774

Table 3.44: Classification Results for Oh0 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	77.762	22.238	2.402	0.778	0.778	0.764	0.742	0.778	0.752
IB3	77.864	22.136	3.350	0.779	0.779	0.778	0.729	0.779	0.753
IB5	80.058	19.942	2.099	0.801	0.801	0.813	0.754	0.801	0.783
IB7	79.855	20.145	2.547	0.799	0.799	0.813	0.744	0.799	0.777
IB9	79.953	20.047	3.134	0.800	0.800	0.826	0.748	0.800	0.785
IB11	80.651	19.349	2.862	0.807	0.807	0.830	0.757	0.807	0.792
IB13	80.652	19.348	2.891	0.807	0.807	0.828	0.755	0.807	0.790
IB15	79.951	20.049	3.622	0.800	0.800	0.824	0.745	0.800	0.782
IB17	80.848	19.153	3.197	0.809	0.809	0.835	0.751	0.809	0.791
IB19	81.250	18.751	3.181	0.813	0.813	0.848	0.758	0.813	0.800
IB21	80.851	19.149	3.049	0.809	0.809	0.846	0.751	0.809	0.796
IB25	79.854	20.146	3.511	0.799	0.799	0.844	0.738	0.799	0.788
IB29	78.657	21.343	4.403	0.787	0.787	0.839	0.726	0.787	0.779
IB35	77.267	22.733	3.429	0.773	0.773	0.838	0.710	0.773	0.769
IB41	76.174	23.826	3.078	0.762	0.762	0.838	0.697	0.762	0.761
IB49	75.078	24.922	3.594	0.751	0.751	0.837	0.682	0.751	0.752
IB57	74.082	25.918	2.698	0.741	0.741	0.841	0.669	0.741	0.745
IB73	72.188	27.812	3.212	0.722	0.722	0.835	0.641	0.722	0.725
IB89	69.894	30.106	2.835	0.699	0.699	0.832	0.612	0.699	0.705
IB1-WV	77.762	22.238	2.402	0.778	0.778	0.764	0.742	0.778	0.752
IB3-WV	80.361	19.639	2.793	0.804	0.804	0.790	0.767	0.804	0.778
IB5-WV	81.358	18.642	2.161	0.814	0.814	0.811	0.775	0.814	0.793
IB7-WV	81.553	18.447	2.012	0.816	0.816	0.819	0.770	0.816	0.794
IB9-WV	80.954	19.046	3.043	0.810	0.810	0.810	0.767	0.810	0.788
IB11-WV	80.753	19.247	2.869	0.808	0.808	0.825	0.766	0.808	0.794
IB13-WV	81.354	18.646	2.318	0.814	0.814	0.836	0.766	0.814	0.800
IB15-WV	80.651	19.349	2.724	0.807	0.807	0.825	0.761	0.807	0.792
IB17-WV	81.249	18.752	3.153	0.813	0.813	0.837	0.764	0.813	0.798
IB19-WV	81.450	18.550	2.715	0.815	0.815	0.844	0.765	0.815	0.802
IB21-WV	81.851	18.149	2.865	0.819	0.819	0.854	0.768	0.819	0.809
IB25-WV	80.554	19.446	3.368	0.806	0.806	0.851	0.751	0.806	0.798
IB29-WV	79.655	20.345	4.019	0.797	0.797	0.851	0.741	0.797	0.792
IB35-WV	78.562	21.438	3.643	0.786	0.786	0.848	0.728	0.786	0.783
IB41-WV	77.272	22.728	3.117	0.773	0.773	0.848	0.713	0.773	0.774
IB49-WV	76.274	23.726	3.423	0.763	0.763	0.846	0.698	0.763	0.765
IB57-WV	75.775	24.225	3.099	0.758	0.758	0.851	0.692	0.758	0.763
IB73-WV	73.283	26.717	3.086	0.733	0.733	0.843	0.657	0.733	0.738
IB89-WV	71.391	28.609	3.197	0.714	0.714	0.840	0.633	0.714	0.722

3.22 Oh5 Collection

Table 3.45 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.46 presents the classification results obtained by IBk for Oh5 collection.

Table 3.45: Classification Results for Oh5 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	78.766	21.234	4.540	0.788	0.788	0.793	0.777	0.788	0.785
Multinomial NB	86.274	13.726	3.313	0.863	0.863	0.861	0.858	0.863	0.860
J48	80.395	19.605	4.879	0.804	0.804	0.805	0.796	0.804	0.800
SMO-Linear-0	8.058	91.942	2.176	0.081	0.081	0.008	0.100	0.081	0.015
SMO-Linear- 10^{-5}	16.338	83.662	6.086	0.163	0.163	0.037	0.101	0.163	0.054
SMO-Linear- 10^{-4}	16.338	83.662	6.086	0.163	0.163	0.037	0.101	0.163	0.054
SMO-Linear- 10^{-3}	14.925	85.075	2.439	0.149	0.149	0.027	0.092	0.149	0.042
SMO-Linear- 10^{-2}	25.489	74.512	4.480	0.255	0.255	0.261	0.160	0.255	0.198
SMO-Linear- 10^{-1}	77.242	22.758	4.621	0.772	0.772	0.814	0.734	0.772	0.772
SMO-Linear- 10^0	76.806	23.194	5.716	0.768	0.768	0.782	0.739	0.768	0.760
SMO-Linear- 10^1	76.806	23.194	5.716	0.768	0.768	0.782	0.739	0.768	0.760
SMO-Linear- 10^2	76.806	23.194	5.716	0.768	0.768	0.782	0.739	0.768	0.760
SMO-Linear- 10^3	76.806	23.194	5.716	0.768	0.768	0.782	0.739	0.768	0.760
SMO-Linear- 10^4	76.806	23.194	5.716	0.768	0.768	0.782	0.739	0.768	0.760
SMO-Linear- 10^5	76.806	23.194	5.716	0.768	0.768	0.782	0.739	0.768	0.760
SMO-Poly-0	8.058	91.942	2.176	0.081	0.081	0.008	0.100	0.081	0.015
SMO-Poly- 10^{-5}	19.070	80.931	6.726	0.191	0.191	0.049	0.119	0.191	0.069
SMO-Poly- 10^{-4}	18.729	81.271	6.872	0.187	0.187	0.038	0.117	0.187	0.057
SMO-Poly- 10^{-3}	15.903	84.097	3.639	0.159	0.159	0.033	0.099	0.159	0.049
SMO-Poly- 10^{-2}	49.234	50.766	6.985	0.492	0.492	0.749	0.387	0.492	0.510
SMO-Poly- 10^{-1}	51.409	48.591	7.546	0.514	0.514	0.751	0.443	0.514	0.557
SMO-Poly- 10^0	48.249	51.751	7.191	0.483	0.483	0.721	0.418	0.483	0.529
SMO-Poly- 10^1	48.249	51.751	7.191	0.483	0.483	0.721	0.418	0.483	0.529
SMO-Poly- 10^2	48.249	51.751	7.191	0.483	0.483	0.721	0.418	0.483	0.529
SMO-Poly- 10^3	48.249	51.751	7.191	0.483	0.483	0.721	0.418	0.483	0.529
SMO-Poly- 10^4	48.249	51.751	7.191	0.483	0.483	0.721	0.418	0.483	0.529
SMO-Poly- 10^5	48.249	51.751	7.191	0.483	0.483	0.721	0.418	0.483	0.529
SMO-RBF-0	8.058	91.942	2.176	0.081	0.081	0.008	0.100	0.081	0.015
SMO-RBF- 10^{-5}	16.229	83.771	5.777	0.162	0.162	0.036	0.101	0.162	0.053
SMO-RBF- 10^{-4}	16.229	83.771	5.777	0.162	0.162	0.036	0.101	0.162	0.053
SMO-RBF- 10^{-3}	16.229	83.771	5.777	0.162	0.162	0.036	0.101	0.162	0.053
SMO-RBF- 10^{-2}	18.186	81.814	6.998	0.182	0.182	0.040	0.113	0.182	0.059
SMO-RBF- 10^{-1}	17.425	82.575	6.813	0.174	0.174	0.036	0.108	0.174	0.054
SMO-RBF- 10^0	35.300	64.700	5.811	0.353	0.353	0.529	0.238	0.353	0.328
SMO-RBF- 10^1	75.825	24.175	6.061	0.758	0.758	0.782	0.728	0.758	0.754
SMO-RBF- 10^2	75.938	24.062	5.685	0.759	0.759	0.775	0.730	0.759	0.752
SMO-RBF- 10^3	75.938	24.062	5.685	0.759	0.759	0.775	0.730	0.759	0.752
SMO-RBF- 10^4	75.938	24.062	5.685	0.759	0.759	0.775	0.730	0.759	0.752
SMO-RBF- 10^5	75.938	24.062	5.685	0.759	0.759	0.775	0.730	0.759	0.752

Table 3.46: Classification Results for Oh5 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	76.914	23.087	3.058	0.769	0.769	0.769	0.759	0.769	0.764
IB3	75.817	24.183	2.740	0.758	0.758	0.775	0.739	0.758	0.757
IB5	77.999	22.001	3.925	0.780	0.780	0.806	0.764	0.780	0.784
IB7	77.778	22.222	4.098	0.778	0.778	0.805	0.763	0.778	0.784
IB9	76.908	23.093	3.095	0.769	0.769	0.801	0.752	0.769	0.776
IB11	77.453	22.547	3.211	0.775	0.775	0.804	0.756	0.775	0.779
IB13	78.215	21.785	4.205	0.782	0.782	0.813	0.761	0.782	0.786
IB15	78.212	21.788	4.304	0.782	0.782	0.816	0.761	0.782	0.788
IB17	77.891	22.110	3.991	0.779	0.779	0.816	0.759	0.779	0.786
IB19	78.433	21.567	3.844	0.784	0.784	0.822	0.765	0.784	0.792
IB21	78.763	21.238	3.489	0.788	0.788	0.824	0.768	0.788	0.795
IB25	77.455	22.545	3.748	0.775	0.775	0.813	0.753	0.775	0.782
IB29	77.235	22.765	3.475	0.772	0.772	0.812	0.749	0.772	0.779
IB35	78.002	21.998	3.760	0.780	0.780	0.817	0.755	0.780	0.785
IB41	77.238	22.762	4.573	0.772	0.772	0.819	0.744	0.772	0.780
IB49	77.347	22.653	3.468	0.773	0.773	0.817	0.746	0.773	0.780
IB57	77.022	22.978	3.957	0.770	0.770	0.816	0.740	0.770	0.776
IB73	78.005	21.995	4.438	0.780	0.780	0.833	0.750	0.780	0.789
IB89	76.476	23.524	4.130	0.765	0.765	0.824	0.733	0.765	0.776
IB1-WV	76.914	23.087	3.058	0.769	0.769	0.769	0.759	0.769	0.764
IB3-WV	78.433	21.567	2.255	0.784	0.784	0.789	0.769	0.784	0.779
IB5-WV	78.981	21.019	2.760	0.790	0.790	0.804	0.775	0.790	0.789
IB7-WV	78.432	21.568	4.004	0.784	0.784	0.807	0.772	0.784	0.789
IB9-WV	77.889	22.111	3.245	0.779	0.779	0.803	0.765	0.779	0.783
IB11-WV	78.545	21.455	3.614	0.785	0.785	0.811	0.770	0.785	0.790
IB13-WV	78.546	21.454	4.325	0.785	0.785	0.813	0.765	0.785	0.789
IB15-WV	78.542	21.459	3.698	0.785	0.785	0.814	0.765	0.785	0.789
IB17-WV	79.197	20.803	3.949	0.792	0.792	0.824	0.772	0.792	0.797
IB19-WV	79.415	20.585	3.702	0.794	0.794	0.831	0.775	0.794	0.802
IB21-WV	79.418	20.582	3.872	0.794	0.794	0.828	0.774	0.794	0.800
IB25-WV	79.089	20.911	3.409	0.791	0.791	0.825	0.770	0.791	0.796
IB29-WV	78.434	21.566	3.582	0.784	0.784	0.821	0.762	0.784	0.790
IB35-WV	78.546	21.454	3.739	0.785	0.785	0.820	0.760	0.785	0.789
IB41-WV	78.002	21.998	4.205	0.780	0.780	0.821	0.752	0.780	0.785
IB49-WV	77.892	22.108	3.409	0.779	0.779	0.819	0.750	0.779	0.783
IB57-WV	78.005	21.995	4.163	0.780	0.780	0.823	0.750	0.780	0.784
IB73-WV	78.984	21.017	4.291	0.790	0.790	0.835	0.760	0.790	0.796
IB89-WV	77.785	22.215	4.687	0.778	0.778	0.833	0.745	0.778	0.786

3.23 Oh10 Collection

Table 3.47 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.48 presents the classification results obtained by IBk for Oh10 collection.

Table 3.47: Classification Results for Oh10 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	72.381	27.619	4.018	0.724	0.724	0.688	0.686	0.724	0.687
Multinomial NB	80.667	19.333	3.513	0.807	0.807	0.775	0.779	0.807	0.777
J48	72.095	27.905	5.455	0.721	0.721	0.720	0.697	0.721	0.708
SMO-Linear-0	12.000	88.000	3.107	0.120	0.120	0.012	0.100	0.120	0.021
SMO-Linear- 10^{-5}	14.095	85.905	2.364	0.141	0.141	0.028	0.090	0.141	0.043
SMO-Linear- 10^{-4}	14.952	85.048	2.089	0.150	0.150	0.030	0.095	0.150	0.045
SMO-Linear- 10^{-3}	14.095	85.905	2.364	0.141	0.141	0.028	0.090	0.141	0.043
SMO-Linear- 10^{-2}	30.571	69.429	3.200	0.306	0.306	0.252	0.204	0.306	0.226
SMO-Linear- 10^{-1}	76.000	24.000	2.687	0.760	0.760	0.777	0.699	0.760	0.736
SMO-Linear- 10^0	75.048	24.952	2.513	0.750	0.750	0.734	0.708	0.750	0.721
SMO-Linear- 10^1	75.048	24.952	2.513	0.750	0.750	0.734	0.708	0.750	0.721
SMO-Linear- 10^2	75.048	24.952	2.513	0.750	0.750	0.734	0.708	0.750	0.721
SMO-Linear- 10^3	75.048	24.952	2.513	0.750	0.750	0.734	0.708	0.750	0.721
SMO-Linear- 10^4	75.048	24.952	2.513	0.750	0.750	0.734	0.708	0.750	0.721
SMO-Linear- 10^5	75.048	24.952	2.513	0.750	0.750	0.734	0.708	0.750	0.721
SMO-Poly-0	12.000	88.000	3.107	0.120	0.120	0.012	0.100	0.120	0.021
SMO-Poly- 10^{-5}	19.714	80.286	3.640	0.197	0.197	0.039	0.125	0.197	0.060
SMO-Poly- 10^{-4}	18.095	81.905	5.902	0.181	0.181	0.036	0.115	0.181	0.055
SMO-Poly- 10^{-3}	16.190	83.810	4.302	0.162	0.162	0.132	0.103	0.162	0.116
SMO-Poly- 10^{-2}	52.286	47.714	4.161	0.523	0.523	0.738	0.396	0.523	0.516
SMO-Poly- 10^{-1}	57.714	42.286	3.862	0.577	0.577	0.692	0.492	0.577	0.575
SMO-Poly- 10^0	60.190	39.810	4.796	0.602	0.602	0.638	0.537	0.602	0.583
SMO-Poly- 10^1	60.286	39.714	4.588	0.603	0.603	0.644	0.539	0.603	0.587
SMO-Poly- 10^2	60.286	39.714	4.588	0.603	0.603	0.644	0.539	0.603	0.587
SMO-Poly- 10^3	60.286	39.714	4.588	0.603	0.603	0.644	0.539	0.603	0.587
SMO-Poly- 10^4	60.286	39.714	4.588	0.603	0.603	0.644	0.539	0.603	0.587
SMO-Poly- 10^5	60.286	39.714	4.588	0.603	0.603	0.644	0.539	0.603	0.587
SMO-RBF-0	12.000	88.000	3.107	0.120	0.120	0.012	0.100	0.120	0.021
SMO-RBF- 10^{-5}	14.095	85.905	2.364	0.141	0.141	0.028	0.090	0.141	0.043
SMO-RBF- 10^{-4}	14.095	85.905	2.364	0.141	0.141	0.028	0.090	0.141	0.043
SMO-RBF- 10^{-3}	14.095	85.905	2.364	0.141	0.141	0.028	0.090	0.141	0.043
SMO-RBF- 10^{-2}	18.095	81.905	7.389	0.181	0.181	0.036	0.115	0.181	0.055
SMO-RBF- 10^{-1}	18.000	82.000	6.015	0.180	0.180	0.036	0.115	0.180	0.055
SMO-RBF- 10^0	45.810	54.191	5.189	0.458	0.458	0.431	0.325	0.458	0.370
SMO-RBF- 10^1	74.190	25.810	3.228	0.742	0.742	0.740	0.691	0.742	0.715
SMO-RBF- 10^2	72.952	27.048	3.386	0.730	0.730	0.731	0.674	0.730	0.701
SMO-RBF- 10^3	72.952	27.048	3.386	0.730	0.730	0.731	0.674	0.730	0.701
SMO-RBF- 10^4	72.952	27.048	3.386	0.730	0.730	0.731	0.674	0.730	0.701
SMO-RBF- 10^5	72.952	27.048	3.386	0.730	0.730	0.731	0.674	0.730	0.701

Table 3.48: Classification Results for Oh10 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	66.286	33.714	2.957	0.663	0.663	0.628	0.614	0.663	0.621
IB3	69.714	30.286	3.095	0.697	0.697	0.724	0.637	0.697	0.677
IB5	70.952	29.048	3.049	0.710	0.710	0.712	0.654	0.710	0.682
IB7	71.048	28.952	4.410	0.710	0.710	0.707	0.648	0.710	0.676
IB9	71.333	28.667	4.073	0.713	0.713	0.712	0.648	0.713	0.678
IB11	72.095	27.905	4.386	0.721	0.721	0.722	0.654	0.721	0.686
IB13	72.381	27.619	4.280	0.724	0.724	0.735	0.661	0.724	0.696
IB15	71.619	28.381	3.609	0.716	0.716	0.728	0.647	0.716	0.685
IB17	72.190	27.810	4.463	0.722	0.722	0.733	0.651	0.722	0.690
IB19	72.000	28.000	4.242	0.720	0.720	0.738	0.649	0.720	0.691
IB21	72.571	27.429	3.659	0.726	0.726	0.754	0.656	0.726	0.701
IB25	72.381	27.619	4.406	0.724	0.724	0.746	0.653	0.724	0.696
IB29	72.190	27.810	4.815	0.722	0.722	0.751	0.647	0.722	0.695
IB35	71.714	28.286	5.635	0.717	0.717	0.718	0.637	0.717	0.675
IB41	72.476	27.524	5.048	0.725	0.725	0.759	0.647	0.725	0.698
IB49	71.810	28.191	5.621	0.718	0.718	0.788	0.635	0.718	0.703
IB57	71.238	28.762	5.550	0.712	0.712	0.793	0.631	0.712	0.702
IB73	69.714	30.286	5.484	0.697	0.697	0.756	0.609	0.697	0.674
IB89	70.190	29.810	5.094	0.702	0.702	0.798	0.615	0.702	0.694
IB1-WV	66.286	33.714	2.957	0.663	0.663	0.628	0.614	0.663	0.621
IB3-WV	70.571	29.429	2.934	0.706	0.706	0.681	0.655	0.706	0.668
IB5-WV	72.190	27.810	2.619	0.722	0.722	0.714	0.671	0.722	0.692
IB7-WV	72.381	27.619	3.354	0.724	0.724	0.716	0.673	0.724	0.694
IB9-WV	72.571	27.429	4.255	0.726	0.726	0.713	0.668	0.726	0.690
IB11-WV	72.667	27.333	4.508	0.727	0.727	0.717	0.663	0.727	0.689
IB13-WV	73.619	26.381	3.763	0.736	0.736	0.738	0.676	0.736	0.706
IB15-WV	72.190	27.810	4.036	0.722	0.722	0.724	0.657	0.722	0.689
IB17-WV	72.952	27.048	4.410	0.730	0.730	0.739	0.664	0.730	0.699
IB19-WV	73.048	26.952	3.951	0.730	0.730	0.744	0.664	0.730	0.702
IB21-WV	73.143	26.857	3.659	0.731	0.731	0.744	0.666	0.731	0.703
IB25-WV	73.238	26.762	4.161	0.732	0.732	0.744	0.664	0.732	0.702
IB29-WV	72.762	27.238	4.952	0.728	0.728	0.746	0.656	0.728	0.698
IB35-WV	72.286	27.714	5.378	0.723	0.723	0.740	0.648	0.723	0.691
IB41-WV	72.667	27.333	5.076	0.727	0.727	0.753	0.650	0.727	0.698
IB49-WV	72.286	27.714	5.327	0.723	0.723	0.777	0.645	0.723	0.705
IB57-WV	72.571	27.429	5.451	0.726	0.726	0.779	0.650	0.726	0.709
IB73-WV	71.048	28.952	5.605	0.710	0.710	0.781	0.628	0.710	0.696
IB89-WV	71.143	28.857	5.217	0.711	0.711	0.802	0.628	0.711	0.705

3.24 Oh15 Collection

Table 3.49 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.50 presents the classification results obtained by IBk for Oh15 collection.

Table 3.49: Classification Results for Oh15 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	75.038	24.962	3.549	0.750	0.750	0.734	0.725	0.750	0.729
Multinomial NB	83.683	16.318	2.736	0.837	0.837	0.833	0.812	0.837	0.822
J48	75.784	24.217	4.885	0.758	0.758	0.772	0.760	0.758	0.766
SMO-Linear-0	6.137	93.863	2.607	0.061	0.061	0.006	0.100	0.061	0.011
SMO-Linear- 10^{-5}	14.560	85.440	2.552	0.146	0.146	0.028	0.085	0.146	0.042
SMO-Linear- 10^{-4}	17.169	82.831	7.198	0.172	0.172	0.035	0.101	0.172	0.052
SMO-Linear- 10^{-3}	14.560	85.440	2.552	0.146	0.146	0.028	0.085	0.146	0.042
SMO-Linear- 10^{-2}	33.632	66.368	3.249	0.336	0.336	0.270	0.199	0.336	0.229
SMO-Linear- 10^{-1}	75.035	24.965	4.218	0.750	0.750	0.787	0.698	0.750	0.740
SMO-Linear- 10^0	74.707	25.293	5.076	0.747	0.747	0.763	0.704	0.747	0.732
SMO-Linear- 10^1	74.925	25.075	4.942	0.749	0.749	0.764	0.706	0.749	0.734
SMO-Linear- 10^2	74.925	25.075	4.942	0.749	0.749	0.764	0.706	0.749	0.734
SMO-Linear- 10^3	74.925	25.075	4.942	0.749	0.749	0.764	0.706	0.749	0.734
SMO-Linear- 10^4	74.925	25.075	4.942	0.749	0.749	0.764	0.706	0.749	0.734
SMO-Linear- 10^5	74.925	25.075	4.942	0.749	0.749	0.764	0.706	0.749	0.734
SMO-Poly-0	6.137	93.863	2.607	0.061	0.061	0.006	0.100	0.061	0.011
SMO-Poly- 10^{-5}	20.684	79.316	7.284	0.207	0.207	0.049	0.121	0.207	0.070
SMO-Poly- 10^{-4}	19.375	80.625	6.028	0.194	0.194	0.043	0.113	0.194	0.062
SMO-Poly- 10^{-3}	16.309	83.691	3.629	0.163	0.163	0.032	0.095	0.163	0.048
SMO-Poly- 10^{-2}	40.096	59.905	3.870	0.401	0.401	0.728	0.290	0.401	0.414
SMO-Poly- 10^{-1}	44.477	55.523	4.195	0.445	0.445	0.747	0.350	0.445	0.476
SMO-Poly- 10^0	47.217	52.783	3.128	0.472	0.472	0.695	0.375	0.472	0.488
SMO-Poly- 10^1	47.327	52.673	3.108	0.473	0.473	0.695	0.376	0.473	0.488
SMO-Poly- 10^2	47.327	52.673	3.108	0.473	0.473	0.695	0.376	0.473	0.488
SMO-Poly- 10^3	47.327	52.673	3.108	0.473	0.473	0.695	0.376	0.473	0.488
SMO-Poly- 10^4	47.327	52.673	3.108	0.473	0.473	0.695	0.376	0.473	0.488
SMO-Poly- 10^5	47.327	52.673	3.108	0.473	0.473	0.695	0.376	0.473	0.488
SMO-RBF-0	6.137	93.863	2.607	0.061	0.061	0.006	0.100	0.061	0.011
SMO-RBF- 10^{-5}	14.560	85.440	2.552	0.146	0.146	0.028	0.085	0.146	0.042
SMO-RBF- 10^{-4}	14.560	85.440	2.552	0.146	0.146	0.028	0.085	0.146	0.042
SMO-RBF- 10^{-3}	14.560	85.440	2.552	0.146	0.146	0.028	0.085	0.146	0.042
SMO-RBF- 10^{-2}	14.560	85.440	2.552	0.146	0.146	0.028	0.085	0.146	0.042
SMO-RBF- 10^{-1}	15.973	84.027	5.699	0.160	0.160	0.032	0.094	0.160	0.047
SMO-RBF- 10^0	39.001	60.999	5.014	0.390	0.390	0.395	0.254	0.390	0.309
SMO-RBF- 10^1	72.406	27.594	4.529	0.724	0.724	0.749	0.676	0.724	0.711
SMO-RBF- 10^2	71.745	28.255	4.838	0.717	0.717	0.735	0.672	0.717	0.702
SMO-RBF- 10^3	71.745	28.255	4.838	0.717	0.717	0.735	0.672	0.717	0.702
SMO-RBF- 10^4	71.745	28.255	4.838	0.717	0.717	0.735	0.672	0.717	0.702
SMO-RBF- 10^5	71.745	28.255	4.838	0.717	0.717	0.735	0.672	0.717	0.702

Table 3.50: Classification Results for Oh15 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	69.116	30.884	4.669	0.691	0.691	0.673	0.661	0.691	0.667
IB3	71.844	28.156	3.376	0.719	0.719	0.712	0.689	0.719	0.700
IB5	73.600	26.400	3.569	0.736	0.736	0.744	0.698	0.736	0.720
IB7	74.371	25.630	3.399	0.744	0.744	0.764	0.706	0.744	0.734
IB9	74.479	25.521	4.170	0.745	0.745	0.768	0.704	0.745	0.735
IB11	74.043	25.957	4.728	0.740	0.740	0.773	0.696	0.740	0.732
IB13	72.839	27.161	4.608	0.728	0.728	0.759	0.683	0.728	0.719
IB15	73.165	26.835	4.725	0.732	0.732	0.770	0.686	0.732	0.725
IB17	72.946	27.055	4.309	0.729	0.729	0.773	0.680	0.729	0.723
IB19	72.944	27.056	4.817	0.729	0.729	0.770	0.681	0.729	0.722
IB21	72.832	27.168	4.329	0.728	0.728	0.772	0.675	0.728	0.720
IB25	72.505	27.495	4.279	0.725	0.725	0.784	0.669	0.725	0.722
IB29	71.521	28.480	4.288	0.715	0.715	0.774	0.655	0.715	0.709
IB35	70.973	29.027	4.174	0.710	0.710	0.796	0.645	0.710	0.713
IB41	70.534	29.466	4.301	0.705	0.705	0.789	0.641	0.705	0.707
IB49	67.906	32.094	4.144	0.679	0.679	0.789	0.610	0.679	0.688
IB57	67.035	32.965	3.713	0.670	0.670	0.790	0.597	0.670	0.680
IB73	63.311	36.689	3.644	0.633	0.633	0.803	0.548	0.633	0.651
IB89	62.218	37.782	4.148	0.622	0.622	0.812	0.533	0.622	0.644
IB1-WV	69.116	30.884	4.669	0.691	0.691	0.673	0.661	0.691	0.667
IB3-WV	73.383	26.617	3.153	0.734	0.734	0.727	0.700	0.734	0.713
IB5-WV	75.024	24.976	3.370	0.750	0.750	0.760	0.715	0.750	0.736
IB7-WV	75.571	24.429	4.122	0.756	0.756	0.779	0.718	0.756	0.747
IB9-WV	75.026	24.974	4.254	0.750	0.750	0.776	0.709	0.750	0.741
IB11-WV	74.700	25.300	4.637	0.747	0.747	0.782	0.704	0.747	0.741
IB13-WV	74.369	25.631	4.893	0.744	0.744	0.789	0.703	0.744	0.743
IB15-WV	74.038	25.962	4.646	0.740	0.740	0.782	0.697	0.740	0.737
IB17-WV	73.599	26.401	4.318	0.736	0.736	0.783	0.689	0.736	0.733
IB19-WV	74.036	25.964	5.032	0.740	0.740	0.790	0.692	0.740	0.738
IB21-WV	73.708	26.293	4.553	0.737	0.737	0.793	0.686	0.737	0.736
IB25-WV	73.275	26.725	3.927	0.733	0.733	0.789	0.679	0.733	0.730
IB29-WV	72.068	27.932	4.212	0.721	0.721	0.784	0.662	0.721	0.718
IB35-WV	70.972	29.028	3.889	0.710	0.710	0.798	0.647	0.710	0.714
IB41-WV	71.082	28.918	4.172	0.711	0.711	0.800	0.648	0.711	0.716
IB49-WV	69.001	30.999	3.841	0.690	0.690	0.798	0.623	0.690	0.700
IB57-WV	67.473	32.528	3.640	0.675	0.675	0.786	0.600	0.675	0.680
IB73-WV	64.736	35.264	3.996	0.647	0.647	0.802	0.568	0.647	0.665
IB89-WV	62.985	37.015	4.982	0.630	0.630	0.814	0.543	0.630	0.651

3.25 Ohscal Collection

Table 3.51 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.52 presents the classification results obtained by IBk for Ohscal collection.

Table 3.51: Classification Results for Ohscal collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	62.784	37.216	1.095	0.628	0.628	0.615	0.621	0.628	0.618
Multinomial NB	74.736	25.264	0.781	0.747	0.747	0.734	0.739	0.747	0.736
J48	71.304	28.696	1.127	0.713	0.713	0.699	0.700	0.713	0.700
SMO-Linear-0	10.383	89.617	0.985	0.104	0.104	0.010	0.100	0.104	0.019
SMO-Linear- 10^{-5}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-Linear- 10^{-4}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-Linear- 10^{-3}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-Linear- 10^{-2}	52.643	47.357	1.331	0.526	0.526	0.695	0.458	0.526	0.552
SMO-Linear- 10^{-1}	76.698	23.302	1.066	0.767	0.767	0.762	0.751	0.767	0.756
SMO-Linear- 10^0	74.359	25.641	1.024	0.744	0.744	0.734	0.733	0.744	0.733
SMO-Linear- 10^1	73.866	26.134	1.166	0.739	0.739	0.729	0.728	0.739	0.728
SMO-Linear- 10^2	73.866	26.134	1.166	0.739	0.739	0.729	0.728	0.739	0.728
SMO-Linear- 10^3	73.866	26.134	1.166	0.739	0.739	0.729	0.728	0.739	0.728
SMO-Linear- 10^4	73.866	26.134	1.166	0.739	0.739	0.729	0.728	0.739	0.728
SMO-Linear- 10^5	73.866	26.134	1.166	0.739	0.739	0.729	0.728	0.739	0.728
SMO-Poly-0	10.383	89.617	0.985	0.104	0.104	0.010	0.100	0.104	0.019
SMO-Poly- 10^{-5}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-Poly- 10^{-4}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-Poly- 10^{-3}	14.540	85.460	1.051	0.145	0.145	0.215	0.100	0.145	0.137
SMO-Poly- 10^{-2}	36.678	63.322	1.301	0.367	0.367	0.698	0.312	0.367	0.431
SMO-Poly- 10^{-1}	60.858	39.142	1.513	0.609	0.609	0.684	0.575	0.609	0.625
SMO-Poly- 10^0	63.429	36.571	1.376	0.634	0.634	0.671	0.607	0.634	0.637
SMO-Poly- 10^1	63.842	36.159	1.525	0.638	0.638	0.662	0.612	0.638	0.636
SMO-Poly- 10^2	63.850	36.150	1.514	0.639	0.639	0.661	0.612	0.639	0.636
SMO-Poly- 10^3	63.850	36.150	1.514	0.639	0.639	0.661	0.612	0.639	0.636
SMO-Poly- 10^4	63.850	36.150	1.514	0.639	0.639	0.661	0.612	0.639	0.636
SMO-Poly- 10^5	63.850	36.150	1.514	0.639	0.639	0.661	0.612	0.639	0.636
SMO-RBF-0	10.383	89.617	0.985	0.104	0.104	0.104	0.100	0.104	0.102
SMO-RBF- 10^{-5}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-RBF- 10^{-4}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-RBF- 10^{-3}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-RBF- 10^{-2}	14.523	85.478	1.039	0.145	0.145	0.015	0.100	0.145	0.025
SMO-RBF- 10^{-1}	14.523	85.478	1.038	0.145	0.145	0.015	0.100	0.145	0.025
SMO-RBF- 10^0	65.598	34.402	1.101	0.656	0.656	0.709	0.610	0.656	0.656
SMO-RBF- 10^1	76.160	23.840	0.948	0.762	0.762	0.756	0.749	0.762	0.752
SMO-RBF- 10^2	74.099	25.901	1.198	0.741	0.741	0.734	0.729	0.741	0.731
SMO-RBF- 10^3	74.028	25.972	1.116	0.740	0.740	0.733	0.728	0.740	0.731
SMO-RBF- 10^4	74.028	25.972	1.116	0.740	0.740	0.733	0.728	0.740	0.731
SMO-RBF- 10^5	74.028	25.972	1.116	0.740	0.740	0.733	0.728	0.740	0.731

Table 3.52: Classification Results for Ohscal collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	60.939	39.061	0.880	0.609	0.609	0.598	0.597	0.609	0.598
IB3	63.053	36.947	0.777	0.631	0.631	0.629	0.622	0.631	0.625
IB5	65.732	34.268	1.322	0.657	0.657	0.652	0.649	0.657	0.650
IB7	66.691	33.309	1.273	0.667	0.667	0.660	0.658	0.667	0.659
IB9	67.237	32.763	1.285	0.672	0.672	0.666	0.661	0.672	0.664
IB11	67.613	32.387	1.315	0.676	0.676	0.671	0.664	0.676	0.667
IB13	68.142	31.858	1.453	0.681	0.681	0.678	0.670	0.681	0.674
IB15	67.927	32.073	1.411	0.679	0.679	0.677	0.668	0.679	0.672
IB17	67.945	32.055	1.427	0.679	0.679	0.679	0.667	0.679	0.673
IB19	68.133	31.867	1.300	0.681	0.681	0.682	0.669	0.681	0.675
IB21	68.312	31.688	1.297	0.683	0.683	0.685	0.670	0.683	0.677
IB25	67.954	32.046	1.357	0.680	0.680	0.684	0.665	0.680	0.674
IB29	67.945	32.055	1.292	0.679	0.679	0.686	0.664	0.679	0.675
IB35	67.649	32.351	1.155	0.676	0.676	0.686	0.661	0.676	0.673
IB41	67.524	32.476	1.117	0.675	0.675	0.687	0.660	0.675	0.673
IB49	67.801	32.199	0.941	0.678	0.678	0.692	0.662	0.678	0.677
IB57	67.425	32.575	1.150	0.674	0.674	0.693	0.658	0.674	0.675
IB73	67.067	32.933	1.285	0.671	0.671	0.693	0.653	0.671	0.672
IB89	66.807	33.193	1.338	0.668	0.668	0.695	0.649	0.668	0.671
IB1-WV	60.939	39.061	0.880	0.609	0.609	0.598	0.597	0.609	0.598
IB3-WV	64.308	35.693	1.089	0.643	0.643	0.633	0.631	0.643	0.632
IB5-WV	66.386	33.614	1.374	0.664	0.664	0.657	0.652	0.664	0.654
IB7-WV	67.345	32.655	1.125	0.673	0.673	0.668	0.662	0.673	0.665
IB9-WV	67.649	32.351	1.351	0.676	0.676	0.671	0.663	0.676	0.667
IB11-WV	68.053	31.948	1.416	0.681	0.681	0.677	0.667	0.681	0.672
IB13-WV	68.411	31.589	1.370	0.684	0.684	0.682	0.670	0.684	0.676
IB15-WV	68.411	31.589	1.086	0.684	0.684	0.684	0.671	0.684	0.677
IB17-WV	68.357	31.643	1.132	0.684	0.684	0.685	0.670	0.684	0.677
IB19-WV	68.464	31.536	1.370	0.685	0.685	0.686	0.671	0.685	0.678
IB21-WV	68.653	31.348	1.453	0.687	0.687	0.690	0.672	0.687	0.681
IB25-WV	68.536	31.464	1.312	0.685	0.685	0.692	0.670	0.685	0.681
IB29-WV	68.196	31.804	1.246	0.682	0.682	0.691	0.666	0.682	0.678
IB35-WV	67.819	32.181	0.989	0.678	0.678	0.691	0.662	0.678	0.676
IB41-WV	67.640	32.360	1.026	0.676	0.676	0.690	0.660	0.676	0.675
IB49-WV	67.855	32.145	1.030	0.679	0.679	0.695	0.662	0.679	0.678
IB57-WV	67.604	32.396	1.126	0.676	0.676	0.695	0.659	0.676	0.676
IB73-WV	67.309	32.691	1.266	0.673	0.673	0.695	0.654	0.673	0.674
IB89-WV	67.174	32.826	1.263	0.672	0.672	0.699	0.653	0.672	0.675

3.26 Ohsumed-400 Collection

Table 3.53 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.54 presents the classification results obtained by IBk for Ohsumed-400 collection.

Table 3.53: Classification Results for Ohsumed-400 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	34.750	65.250	1.269	0.348	0.348	0.346	0.348	0.348	0.347
Multinomial NB	40.511	59.489	1.449	0.405	0.405	0.404	0.405	0.405	0.405
J48	30.913	69.087	1.676	0.309	0.309	0.307	0.310	0.309	0.308
SMO-Linear-0	4.348	95.652	0.543	0.043	0.043	0.002	0.043	0.043	0.003
SMO-Linear- 10^{-5}	3.424	96.576	0.824	0.034	0.034	0.018	0.034	0.034	0.024
SMO-Linear- 10^{-4}	3.109	96.891	0.263	0.031	0.031	0.011	0.031	0.031	0.016
SMO-Linear- 10^{-3}	3.239	96.761	0.539	0.032	0.032	0.011	0.032	0.032	0.017
SMO-Linear- 10^{-2}	7.783	92.217	2.243	0.078	0.078	0.169	0.078	0.078	0.106
SMO-Linear- 10^{-1}	35.707	64.293	1.686	0.357	0.357	0.368	0.358	0.357	0.363
SMO-Linear- 10^0	32.500	67.500	1.580	0.325	0.325	0.333	0.326	0.325	0.329
SMO-Linear- 10^1	32.250	67.750	1.795	0.323	0.323	0.330	0.323	0.323	0.327
SMO-Linear- 10^2	32.272	67.728	1.786	0.323	0.323	0.330	0.323	0.323	0.327
SMO-Linear- 10^3	32.250	67.750	1.795	0.323	0.323	0.330	0.323	0.323	0.327
SMO-Linear- 10^4	32.261	67.739	1.791	0.323	0.323	0.330	0.323	0.323	0.327
SMO-Linear- 10^5	32.250	67.750	1.795	0.323	0.323	0.330	0.323	0.323	0.327
SMO-Poly-0	4.348	95.652	0.543	0.043	0.043	0.002	0.043	0.043	0.004
SMO-Poly- 10^{-5}	3.913	96.087	0.726	0.039	0.039	0.016	0.053	0.039	0.024
SMO-Poly- 10^{-4}	5.196	94.804	1.340	0.052	0.052	0.073	0.067	0.052	0.069
SMO-Poly- 10^{-3}	3.587	96.413	0.748	0.036	0.036	0.007	0.050	0.036	0.012
SMO-Poly- 10^{-2}	18.293	81.707	2.023	0.183	0.183	0.245	0.191	0.183	0.215
SMO-Poly- 10^{-1}	24.326	75.674	1.321	0.243	0.243	0.265	0.246	0.243	0.255
SMO-Poly- 10^0	23.391	76.609	0.880	0.234	0.234	0.245	0.235	0.234	0.240
SMO-Poly- 10^1	22.467	77.533	0.863	0.225	0.225	0.241	0.226	0.225	0.233
SMO-Poly- 10^2	22.467	77.533	0.850	0.225	0.225	0.241	0.226	0.225	0.233
SMO-Poly- 10^3	22.457	77.543	0.858	0.225	0.225	0.241	0.226	0.225	0.233
SMO-Poly- 10^4	22.467	77.533	0.850	0.225	0.225	0.241	0.226	0.225	0.233
SMO-Poly- 10^5	22.457	77.543	0.858	0.225	0.225	0.240	0.226	0.225	0.233
SMO-RBF-0	4.348	95.652	0.543	0.043	0.043	0.002	0.043	0.043	0.004
SMO-RBF- 10^{-5}	3.380	96.620	0.760	0.034	0.034	0.003	0.047	0.034	0.006
SMO-RBF- 10^{-4}	3.380	96.620	0.760	0.034	0.034	0.003	0.047	0.034	0.006
SMO-RBF- 10^{-3}	3.380	96.620	0.760	0.034	0.034	0.003	0.047	0.034	0.006
SMO-RBF- 10^{-2}	4.217	95.783	1.247	0.042	0.042	0.007	0.058	0.042	0.013
SMO-RBF- 10^{-1}	5.609	94.391	2.342	0.056	0.056	0.030	0.074	0.056	0.042
SMO-RBF- 10^0	21.935	78.065	1.129	0.219	0.219	0.361	0.236	0.219	0.285
SMO-RBF- 10^1	32.000	68.000	1.401	0.320	0.320	0.331	0.321	0.320	0.326
SMO-RBF- 10^2	30.913	69.087	1.717	0.309	0.309	0.325	0.310	0.309	0.317
SMO-RBF- 10^3	30.924	69.076	1.737	0.309	0.309	0.324	0.310	0.309	0.317
SMO-RBF- 10^4	30.913	69.087	1.732	0.309	0.309	0.324	0.310	0.309	0.317
SMO-RBF- 10^5	30.913	69.087	1.732	0.309	0.309	0.324	0.310	0.309	0.317

Table 3.54: Classification Results for Ohsumed-400 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	23.848	76.152	1.042	0.238	0.238	0.252	0.238	0.238	0.245
IB3	28.207	71.793	1.152	0.282	0.282	0.328	0.282	0.282	0.304
IB5	32.522	67.478	1.324	0.325	0.325	0.350	0.326	0.325	0.337
IB7	34.207	65.793	1.739	0.342	0.342	0.357	0.343	0.342	0.350
IB9	35.348	64.652	1.823	0.353	0.353	0.366	0.354	0.353	0.360
IB11	35.902	64.098	1.498	0.359	0.359	0.371	0.360	0.359	0.366
IB13	36.370	63.630	1.597	0.364	0.364	0.375	0.365	0.364	0.370
IB15	36.967	63.033	1.387	0.370	0.370	0.382	0.372	0.370	0.377
IB17	37.424	62.576	1.278	0.374	0.374	0.385	0.376	0.374	0.381
IB19	37.891	62.109	1.379	0.379	0.379	0.388	0.381	0.379	0.384
IB21	37.761	62.239	1.598	0.378	0.378	0.387	0.380	0.378	0.383
IB25	37.761	62.239	1.490	0.378	0.378	0.391	0.380	0.378	0.385
IB29	37.935	62.065	1.079	0.379	0.379	0.389	0.381	0.379	0.385
IB35	37.815	62.185	1.156	0.378	0.378	0.388	0.380	0.378	0.384
IB41	38.163	61.837	1.273	0.382	0.382	0.394	0.384	0.382	0.389
IB49	38.098	61.902	0.987	0.381	0.381	0.395	0.383	0.381	0.389
IB57	38.120	61.880	1.054	0.381	0.381	0.395	0.384	0.381	0.389
IB73	37.967	62.033	1.027	0.380	0.380	0.400	0.382	0.380	0.391
IB89	37.935	62.065	1.184	0.379	0.379	0.404	0.381	0.379	0.392
IB1-WV	23.848	76.152	1.042	0.238	0.238	0.252	0.238	0.238	0.245
IB3-WV	25.500	74.500	1.067	0.255	0.255	0.265	0.255	0.255	0.260
IB5-WV	27.467	72.533	0.983	0.275	0.275	0.283	0.274	0.275	0.279
IB7-WV	28.196	71.804	1.133	0.282	0.282	0.289	0.282	0.282	0.286
IB9-WV	28.891	71.109	1.262	0.289	0.289	0.297	0.289	0.289	0.293
IB11-WV	29.087	70.913	1.445	0.291	0.291	0.299	0.291	0.291	0.295
IB13-WV	29.630	70.370	1.600	0.296	0.296	0.305	0.298	0.296	0.301
IB15-WV	30.054	69.946	1.772	0.301	0.301	0.308	0.302	0.301	0.305
IB17-WV	30.576	69.424	1.586	0.306	0.306	0.314	0.307	0.306	0.310
IB19-WV	30.957	69.043	1.553	0.310	0.310	0.317	0.310	0.310	0.314
IB21-WV	31.239	68.761	1.713	0.312	0.312	0.319	0.313	0.312	0.316
IB25-WV	31.522	68.478	1.308	0.315	0.315	0.322	0.317	0.315	0.319
IB29-WV	32.207	67.793	1.184	0.322	0.322	0.327	0.323	0.322	0.325
IB35-WV	32.793	67.207	0.989	0.328	0.328	0.335	0.329	0.328	0.332
IB41-WV	33.239	66.761	0.843	0.332	0.332	0.339	0.334	0.332	0.336
IB49-WV	33.587	66.413	0.750	0.336	0.336	0.344	0.337	0.336	0.340
IB57-WV	33.739	66.261	0.974	0.337	0.337	0.347	0.339	0.337	0.343
IB73-WV	34.587	65.413	0.804	0.346	0.346	0.357	0.347	0.346	0.352
IB89-WV	34.663	65.337	1.109	0.347	0.347	0.359	0.348	0.347	0.353

3.27 Opinosis Collection

Table 3.55 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.56 presents the classification results obtained by IBk for Opinosis collection.

Table 3.55: Classification Results for Opinosis collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	60.740	39.260	0.939	0.607	0.607	0.600	0.618	0.607	0.609
Multinomial NB	59.563	40.437	1.859	0.596	0.596	0.644	0.532	0.596	0.582
J48	60.833	39.167	1.615	0.608	0.608	0.620	0.608	0.608	0.614
SMO-Linear-0	1.595	98.405	0.495	0.016	0.016	0.000	0.020	0.016	0.001
SMO-Linear- 10^{-5}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-Linear- 10^{-4}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-Linear- 10^{-3}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-Linear- 10^{-2}	9.292	90.708	0.877	0.093	0.093	0.150	0.025	0.093	0.042
SMO-Linear- 10^{-1}	56.079	43.922	1.451	0.561	0.561	0.605	0.477	0.561	0.533
SMO-Linear- 10^0	61.034	38.966	1.925	0.610	0.610	0.617	0.611	0.610	0.614
SMO-Linear- 10^1	59.671	40.329	1.720	0.597	0.597	0.604	0.612	0.597	0.608
SMO-Linear- 10^2	57.998	42.002	1.637	0.580	0.580	0.589	0.605	0.580	0.597
SMO-Linear- 10^3	57.472	42.528	1.285	0.575	0.575	0.586	0.603	0.575	0.594
SMO-Linear- 10^4	57.286	42.714	1.306	0.573	0.573	0.585	0.602	0.573	0.593
SMO-Linear- 10^5	57.286	42.714	1.324	0.573	0.573	0.585	0.602	0.573	0.594
SMO-Poly-0	1.595	98.405	0.495	0.016	0.016	0.000	0.020	0.016	0.001
SMO-Poly- 10^{-5}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-Poly- 10^{-4}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-Poly- 10^{-3}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-Poly- 10^{-2}	8.952	91.048	0.771	0.090	0.090	0.136	0.025	0.090	0.042
SMO-Poly- 10^{-1}	28.202	71.798	1.846	0.282	0.282	0.425	0.184	0.282	0.257
SMO-Poly- 10^0	40.995	59.006	1.851	0.410	0.410	0.453	0.357	0.410	0.399
SMO-Poly- 10^1	45.949	54.051	2.025	0.460	0.460	0.485	0.436	0.460	0.459
SMO-Poly- 10^2	47.544	52.456	1.963	0.475	0.475	0.499	0.467	0.475	0.482
SMO-Poly- 10^3	46.956	53.044	1.956	0.470	0.470	0.498	0.466	0.470	0.481
SMO-Poly- 10^4	46.956	53.044	1.956	0.470	0.470	0.498	0.466	0.470	0.481
SMO-Poly- 10^5	46.956	53.044	1.944	0.470	0.470	0.498	0.466	0.470	0.481
SMO-RBF-0	1.595	98.405	0.495	0.016	0.016	0.000	0.020	0.016	0.001
SMO-RBF- 10^{-5}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-RBF- 10^{-4}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-RBF- 10^{-3}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-RBF- 10^{-2}	8.177	91.823	0.599	0.082	0.082	0.002	0.020	0.082	0.003
SMO-RBF- 10^{-1}	28.202	71.798	0.599	0.282	0.282	0.425	0.184	0.282	0.257
SMO-RBF- 10^0	40.995	59.006	1.350	0.410	0.410	0.453	0.357	0.410	0.399
SMO-RBF- 10^1	45.949	54.051	1.815	0.460	0.460	0.485	0.436	0.460	0.459
SMO-RBF- 10^2	47.544	52.456	1.968	0.475	0.475	0.499	0.467	0.475	0.482
SMO-RBF- 10^3	46.956	53.044	1.985	0.470	0.470	0.498	0.466	0.470	0.481
SMO-RBF- 10^4	46.956	53.044	1.272	0.470	0.470	0.498	0.466	0.470	0.481
SMO-RBF- 10^5	46.971	53.029	1.543	0.470	0.470	0.498	0.466	0.470	0.481

Table 3.56: Classification Results for Opinions collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	51.478	48.522	2.055	0.515	0.515	0.521	0.514	0.515	0.518
IB3	56.512	43.488	2.167	0.565	0.565	0.582	0.560	0.565	0.570
IB5	60.074	39.926	1.758	0.601	0.601	0.627	0.590	0.601	0.608
IB7	62.273	37.727	1.767	0.623	0.623	0.660	0.610	0.623	0.634
IB9	62.304	37.696	1.858	0.623	0.623	0.666	0.607	0.623	0.635
IB11	62.583	37.417	1.866	0.626	0.626	0.671	0.605	0.626	0.636
IB13	62.521	37.479	2.072	0.625	0.625	0.669	0.603	0.625	0.634
IB15	62.769	37.231	1.835	0.628	0.628	0.673	0.604	0.628	0.637
IB17	62.661	37.339	1.981	0.627	0.627	0.682	0.601	0.627	0.639
IB19	62.521	37.479	1.897	0.625	0.625	0.678	0.598	0.625	0.635
IB21	62.428	37.572	1.893	0.624	0.624	0.678	0.595	0.624	0.634
IB25	62.877	37.123	2.052	0.629	0.629	0.682	0.594	0.629	0.635
IB29	62.862	37.138	1.972	0.629	0.629	0.671	0.591	0.629	0.628
IB35	62.784	37.216	2.031	0.628	0.628	0.686	0.586	0.628	0.632
IB41	62.319	37.681	1.863	0.623	0.623	0.677	0.578	0.623	0.624
IB49	62.149	37.851	1.759	0.621	0.621	0.675	0.574	0.621	0.620
IB57	61.654	38.346	1.595	0.617	0.617	0.693	0.567	0.617	0.623
IB73	61.809	38.191	1.373	0.618	0.618	0.658	0.564	0.618	0.607
IB89	61.391	38.610	1.387	0.614	0.614	0.674	0.556	0.614	0.609
IB1-WV	51.478	48.522	2.055	0.515	0.515	0.521	0.514	0.515	0.518
IB3-WV	54.901	45.099	2.122	0.549	0.549	0.557	0.546	0.549	0.551
IB5-WV	56.868	43.132	1.921	0.569	0.569	0.578	0.560	0.569	0.569
IB7-WV	58.169	41.831	1.974	0.582	0.582	0.596	0.573	0.582	0.584
IB9-WV	58.865	41.135	2.543	0.589	0.589	0.605	0.576	0.589	0.590
IB11-WV	58.865	41.135	2.311	0.589	0.589	0.604	0.573	0.589	0.588
IB13-WV	59.067	40.933	2.358	0.591	0.591	0.608	0.574	0.591	0.590
IB15-WV	59.377	40.623	2.235	0.594	0.594	0.611	0.574	0.594	0.591
IB17-WV	59.408	40.592	2.410	0.594	0.594	0.613	0.571	0.594	0.591
IB19-WV	59.206	40.794	2.534	0.592	0.592	0.610	0.567	0.592	0.588
IB21-WV	59.361	40.639	2.425	0.594	0.594	0.615	0.567	0.594	0.590
IB25-WV	59.779	40.221	2.224	0.598	0.598	0.614	0.568	0.598	0.590
IB29-WV	60.074	39.926	2.041	0.601	0.601	0.613	0.566	0.601	0.589
IB35-WV	60.167	39.833	2.116	0.602	0.602	0.615	0.563	0.602	0.587
IB41-WV	59.903	40.097	2.073	0.599	0.599	0.608	0.558	0.599	0.581
IB49-WV	60.167	39.834	2.006	0.602	0.602	0.613	0.556	0.602	0.583
IB57-WV	60.105	39.895	1.593	0.601	0.601	0.621	0.554	0.601	0.585
IB73-WV	60.213	39.787	1.586	0.602	0.602	0.614	0.548	0.602	0.579
IB89-WV	60.244	39.756	1.596	0.602	0.602	0.630	0.545	0.602	0.584

3.28 Pubmed-Cancer Collection

Table 3.57 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.58 presents the classification results obtained by IBk for Pubmed-Cancer collection⁶.

Table 3.57: Classification Results for Pubmed-Cancer collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	69.902	30.098	0.444	0.699	0.699	0.518	0.691	0.699	0.592
Multinomial NB	86.609	13.391	0.349	0.866	0.866	0.763	0.761	0.866	0.762
J48	96.798	3.202	0.180	0.968	0.968	0.943	0.920	0.968	0.931
SMO-Linear-0	3.647	96.353	0.358	0.036	0.036	0.003	0.083	0.036	0.006
SMO-Linear- 10^{-5}	26.746	73.254	0.298	0.267	0.267	0.022	0.083	0.267	0.035
SMO-Linear- 10^{-4}	48.998	51.002	0.688	0.490	0.490	0.213	0.163	0.490	0.184
SMO-Linear- 10^{-3}	48.998	51.002	0.688	0.490	0.490	0.213	0.163	0.490	0.184
SMO-Linear- 10^{-2}	82.202	17.798	0.459	0.822	0.822	0.771	0.468	0.822	0.582
SMO-Linear- 10^{-1}	95.048	4.952	0.286	0.950	0.950	0.940	0.825	0.950	0.878
SMO-Linear- 10^0	95.980	4.020	0.275	0.960	0.960	0.929	0.898	0.960	0.913
SMO-Linear- 10^1	95.474	4.526	0.304	0.955	0.955	0.920	0.893	0.955	0.906
SMO-Linear- 10^2	95.395	4.605	0.311	0.954	0.954	0.918	0.893	0.954	0.905
SMO-Linear- 10^3	95.369	4.631	0.316	0.954	0.954	0.917	0.893	0.954	0.905
SMO-Linear- 10^4	95.369	4.631	0.316	0.954	0.954	0.917	0.893	0.954	0.905
SMO-Linear- 10^5	95.368	4.632	0.316	0.954	0.954	0.917	0.893	0.954	0.905
SMO-Poly-0	3.647	96.353	0.358	0.036	0.036	0.003	0.083	0.036	0.006
SMO-Poly- 10^{-5}	26.746	73.254	0.298	0.267	0.267	0.022	0.083	0.267	0.035
SMO-Poly- 10^{-4}	26.746	73.254	0.298	0.267	0.267	0.022	0.083	0.267	0.035
SMO-Poly- 10^{-3}	26.961	73.039	0.265	0.270	0.270	0.188	0.084	0.270	0.116
SMO-Poly- 10^{-2}	63.419	36.581	0.829	0.634	0.634	0.722	0.289	0.634	0.413
SMO-Poly- 10^{-1}	84.640	15.360	0.332	0.846	0.846	0.860	0.637	0.846	0.732
SMO-Poly- 10^0	88.567	11.433	0.421	0.886	0.886	0.862	0.733	0.886	0.792
SMO-Poly- 10^1	89.056	10.944	0.397	0.891	0.891	0.853	0.748	0.891	0.797
SMO-Poly- 10^2	88.973	11.027	0.480	0.890	0.890	0.828	0.758	0.890	0.791
SMO-Poly- 10^3	88.967	11.033	0.462	0.890	0.890	0.828	0.758	0.890	0.791
SMO-Poly- 10^4	88.964	11.036	0.462	0.890	0.890	0.828	0.758	0.890	0.791
SMO-Poly- 10^5	88.967	11.033	0.465	0.890	0.890	0.828	0.758	0.890	0.791
SMO-RBF-0	3.647	96.353	0.358	0.036	0.036	0.003	0.083	0.036	0.006
SMO-RBF- 10^{-5}	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^{-4}	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^{-3}	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^{-2}	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^{-1}	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^0	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^1	-	-	-	-	-	-	-	-	-
SMO-RBF- 10^2	95.627	4.373	0.262	0.956	0.956	0.931	0.887	0.956	0.908
SMO-RBF- 10^3	95.152	4.848	0.330	0.952	0.952	0.921	0.883	0.952	0.901
SMO-RBF- 10^4	95.140	4.860	0.324	0.951	0.951	0.921	0.883	0.951	0.901
SMO-RBF- 10^5	95.134	4.866	0.321	0.951	0.951	0.920	0.883	0.951	0.901

⁶The complete results will be available at http://sites.labc.icmc.usp.br/text_collections/

Table 3.58: Classification Results for Pubmed-Cancer collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	-	-	-	-	-	-	-	-	-
IB3	-	-	-	-	-	-	-	-	-
IB5	-	-	-	-	-	-	-	-	-
IB7	-	-	-	-	-	-	-	-	-
IB9	-	-	-	-	-	-	-	-	-
IB11	-	-	-	-	-	-	-	-	-
IB13	-	-	-	-	-	-	-	-	-
IB15	-	-	-	-	-	-	-	-	-
IB17	-	-	-	-	-	-	-	-	-
IB19	-	-	-	-	-	-	-	-	-
IB21	-	-	-	-	-	-	-	-	-
IB25	-	-	-	-	-	-	-	-	-
IB29	-	-	-	-	-	-	-	-	-
IB35	-	-	-	-	-	-	-	-	-
IB41	-	-	-	-	-	-	-	-	-
IB49	-	-	-	-	-	-	-	-	-
IB57	-	-	-	-	-	-	-	-	-
IB73	-	-	-	-	-	-	-	-	-
IB89	-	-	-	-	-	-	-	-	-
IB1-WV	-	-	-	-	-	-	-	-	-
IB3-WV	-	-	-	-	-	-	-	-	-
IB5-WV	-	-	-	-	-	-	-	-	-
IB7-WV	-	-	-	-	-	-	-	-	-
IB9-WV	-	-	-	-	-	-	-	-	-
IB11-WV	-	-	-	-	-	-	-	-	-
IB13-WV	-	-	-	-	-	-	-	-	-
IB15-WV	-	-	-	-	-	-	-	-	-
IB17-WV	-	-	-	-	-	-	-	-	-
IB19-WV	-	-	-	-	-	-	-	-	-
IB21-WV	-	-	-	-	-	-	-	-	-
IB25-WV	-	-	-	-	-	-	-	-	-
IB29-WV	-	-	-	-	-	-	-	-	-
IB35-WV	-	-	-	-	-	-	-	-	-
IB41-WV	-	-	-	-	-	-	-	-	-
IB49-WV	-	-	-	-	-	-	-	-	-
IB57-WV	-	-	-	-	-	-	-	-	-
IB73-WV	-	-	-	-	-	-	-	-	-
IB89-WV	-	-	-	-	-	-	-	-	-

3.29 Re0 Collection

Table 3.59 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.60 presents the classification results obtained by IBk for Re0 collection.

Table 3.59: Classification Results for Re0 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	57.052	42.948	3.622	0.570	0.570	0.448	0.552	0.570	0.495
Multinomial NB	79.925	20.076	3.875	0.799	0.799	0.808	0.691	0.799	0.745
J48	75.268	24.732	2.699	0.753	0.753	0.652	0.616	0.753	0.633
SMO-Linear-0	2.526	97.474	1.352	0.025	0.025	0.002	0.077	0.025	0.004
SMO-Linear- 10^{-5}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-Linear- 10^{-4}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-Linear- 10^{-3}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-Linear- 10^{-2}	47.077	52.923	5.017	0.471	0.471	0.183	0.101	0.471	0.130
SMO-Linear- 10^{-1}	67.890	32.110	5.726	0.679	0.679	0.842	0.407	0.679	0.549
SMO-Linear- 10^0	75.200	24.801	4.143	0.752	0.752	0.779	0.609	0.752	0.683
SMO-Linear- 10^1	76.265	23.735	1.857	0.763	0.763	0.717	0.717	0.763	0.717
SMO-Linear- 10^2	77.196	22.805	2.064	0.772	0.772	0.699	0.759	0.772	0.728
SMO-Linear- 10^3	75.934	24.066	2.897	0.759	0.759	0.693	0.749	0.759	0.720
SMO-Linear- 10^4	75.471	24.529	3.378	0.755	0.755	0.687	0.746	0.755	0.715
SMO-Linear- 10^5	75.471	24.529	3.378	0.755	0.755	0.687	0.746	0.755	0.715
SMO-Poly-0	2.526	97.474	1.352	0.025	0.025	0.002	0.077	0.025	0.004
SMO-Poly- 10^{-5}	40.430	59.570	4.392	0.404	0.404	0.070	0.077	0.404	0.073
SMO-Poly- 10^{-4}	40.430	59.570	4.351	0.404	0.404	0.070	0.077	0.404	0.073
SMO-Poly- 10^{-3}	42.755	57.245	4.501	0.428	0.428	0.180	0.086	0.428	0.116
SMO-Poly- 10^{-2}	50.070	49.930	5.293	0.501	0.501	0.793	0.177	0.501	0.289
SMO-Poly- 10^{-1}	56.918	43.082	5.332	0.569	0.569	0.680	0.315	0.569	0.430
SMO-Poly- 10^0	60.507	39.493	5.316	0.605	0.605	0.648	0.416	0.605	0.507
SMO-Poly- 10^1	62.235	37.765	3.602	0.622	0.622	0.595	0.491	0.622	0.538
SMO-Poly- 10^2	65.161	34.839	2.121	0.652	0.652	0.582	0.622	0.652	0.602
SMO-Poly- 10^3	65.094	34.907	2.956	0.651	0.651	0.547	0.641	0.651	0.590
SMO-Poly- 10^4	64.428	35.572	3.190	0.644	0.644	0.539	0.641	0.644	0.586
SMO-Poly- 10^5	64.428	35.572	3.190	0.644	0.644	0.539	0.641	0.644	0.586
SMO-RBF-0	2.526	97.474	1.352	0.025	0.025	0.002	0.077	0.025	0.004
SMO-RBF- 10^{-5}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-RBF- 10^{-4}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-RBF- 10^{-3}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-RBF- 10^{-2}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-RBF- 10^{-1}	40.430	59.570	4.392	0.404	0.404	0.031	0.077	0.404	0.044
SMO-RBF- 10^0	50.738	49.262	6.071	0.507	0.507	0.323	0.120	0.507	0.175
SMO-RBF- 10^1	72.212	27.788	5.504	0.722	0.722	0.839	0.475	0.722	0.607
SMO-RBF- 10^2	76.729	23.271	2.999	0.767	0.767	0.812	0.639	0.767	0.716
SMO-RBF- 10^3	77.796	22.204	2.830	0.778	0.778	0.753	0.726	0.778	0.739
SMO-RBF- 10^4	77.395	22.605	1.310	0.774	0.774	0.740	0.736	0.774	0.738
SMO-RBF- 10^5	76.863	23.137	1.610	0.769	0.769	0.736	0.736	0.769	0.736

Table 3.60: Classification Results for Re0 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	80.723	19.277	3.938	0.807	0.807	0.733	0.727	0.807	0.730
IB3	79.330	20.670	3.562	0.793	0.793	0.715	0.647	0.793	0.679
IB5	81.653	18.348	2.423	0.816	0.816	0.756	0.682	0.816	0.717
IB7	80.656	19.344	2.845	0.807	0.807	0.760	0.639	0.807	0.694
IB9	80.921	19.080	2.032	0.809	0.809	0.788	0.667	0.809	0.722
IB11	80.656	19.344	3.256	0.807	0.807	0.787	0.652	0.807	0.713
IB13	80.189	19.811	3.802	0.802	0.802	0.775	0.617	0.802	0.687
IB15	80.588	19.412	3.560	0.806	0.806	0.785	0.619	0.806	0.692
IB17	80.588	19.412	3.411	0.806	0.806	0.785	0.616	0.806	0.691
IB19	80.189	19.811	2.935	0.802	0.802	0.782	0.591	0.802	0.673
IB21	79.924	20.076	3.547	0.799	0.799	0.791	0.581	0.799	0.670
IB25	79.257	20.743	2.862	0.793	0.793	0.788	0.550	0.793	0.648
IB29	78.989	21.012	2.599	0.790	0.790	0.826	0.534	0.790	0.649
IB35	78.325	21.676	2.740	0.783	0.783	0.838	0.514	0.783	0.637
IB41	77.594	22.406	2.526	0.776	0.776	0.760	0.495	0.776	0.600
IB49	77.260	22.740	2.708	0.773	0.773	0.682	0.472	0.773	0.557
IB57	76.395	23.605	2.716	0.764	0.764	0.673	0.439	0.764	0.531
IB73	74.997	25.003	2.515	0.750	0.750	0.542	0.389	0.750	0.453
IB89	74.066	25.934	2.793	0.741	0.741	0.481	0.372	0.741	0.420
IB1-WV	80.723	19.277	3.938	0.807	0.807	0.733	0.727	0.807	0.730
IB3-WV	81.388	18.612	3.191	0.814	0.814	0.726	0.698	0.814	0.712
IB5-WV	83.379	16.621	2.609	0.834	0.834	0.773	0.716	0.834	0.743
IB7-WV	82.981	17.019	2.644	0.830	0.830	0.781	0.700	0.830	0.739
IB9-WV	83.513	16.487	2.158	0.835	0.835	0.806	0.712	0.835	0.756
IB11-WV	83.249	16.752	2.993	0.832	0.832	0.815	0.699	0.832	0.752
IB13-WV	82.317	17.683	3.817	0.823	0.823	0.789	0.662	0.823	0.720
IB15-WV	82.583	17.417	3.408	0.826	0.826	0.810	0.660	0.826	0.728
IB17-WV	82.450	17.550	3.322	0.824	0.824	0.790	0.664	0.824	0.722
IB19-WV	82.052	17.948	3.572	0.820	0.820	0.799	0.640	0.820	0.711
IB21-WV	81.519	18.481	3.611	0.815	0.815	0.822	0.632	0.815	0.714
IB25-WV	81.120	18.880	2.897	0.811	0.811	0.802	0.602	0.811	0.687
IB29-WV	80.917	19.084	2.841	0.809	0.809	0.839	0.584	0.809	0.689
IB35-WV	80.117	19.883	2.250	0.801	0.801	0.842	0.556	0.801	0.670
IB41-WV	79.255	20.745	2.377	0.793	0.793	0.774	0.529	0.793	0.629
IB49-WV	78.923	21.077	2.520	0.789	0.789	0.769	0.520	0.789	0.621
IB57-WV	77.925	22.076	2.434	0.779	0.779	0.681	0.487	0.779	0.568
IB73-WV	76.991	23.009	2.335	0.770	0.770	0.694	0.444	0.770	0.541
IB89-WV	76.061	23.939	2.850	0.761	0.761	0.636	0.417	0.761	0.504

3.30 Re1 Collection

Table 3.61 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.62 presents the classification results obtained by IBk for Re1 collection.

Table 3.61: Classification Results for Re1 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	66.740	33.260	4.319	0.667	0.667	0.548	0.581	0.667	0.564
Multinomial NB	83.341	16.659	2.347	0.833	0.833	0.779	0.646	0.833	0.706
J48	79.604	20.396	2.101	0.796	0.796	0.727	0.669	0.796	0.697
SMO-Linear-0	1.871	98.129	0.873	0.019	0.019	0.001	0.040	0.019	0.001
SMO-Linear- 10^{-5}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-Linear- 10^{-4}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-Linear- 10^{-3}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-Linear- 10^{-2}	32.587	67.413	2.838	0.326	0.326	0.226	0.068	0.326	0.104
SMO-Linear- 10^{-1}	64.033	35.967	2.946	0.640	0.640	0.716	0.355	0.640	0.474
SMO-Linear- 10^0	72.724	27.276	3.065	0.727	0.727	0.745	0.560	0.727	0.639
SMO-Linear- 10^1	72.180	27.820	2.312	0.722	0.722	0.708	0.584	0.722	0.640
SMO-Linear- 10^2	71.154	28.846	2.402	0.712	0.712	0.674	0.583	0.712	0.625
SMO-Linear- 10^3	70.972	29.028	2.402	0.710	0.710	0.670	0.582	0.710	0.623
SMO-Linear- 10^4	70.972	29.028	2.402	0.710	0.710	0.670	0.582	0.710	0.623
SMO-Linear- 10^5	70.972	29.028	2.402	0.710	0.710	0.670	0.582	0.710	0.623
SMO-Poly-0	1.871	98.129	0.873	0.019	0.019	0.001	0.040	0.019	0.001
SMO-Poly- 10^{-5}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-Poly- 10^{-4}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-Poly- 10^{-3}	24.321	75.679	2.395	0.243	0.243	0.202	0.050	0.243	0.080
SMO-Poly- 10^{-2}	37.779	62.221	2.206	0.378	0.378	0.647	0.142	0.378	0.232
SMO-Poly- 10^{-1}	50.636	49.364	2.302	0.506	0.506	0.719	0.289	0.506	0.412
SMO-Poly- 10^0	53.169	46.831	2.696	0.532	0.532	0.624	0.323	0.532	0.425
SMO-Poly- 10^1	53.350	46.650	2.927	0.533	0.533	0.592	0.353	0.533	0.442
SMO-Poly- 10^2	53.953	46.047	3.215	0.540	0.540	0.556	0.367	0.540	0.442
SMO-Poly- 10^3	53.653	46.347	2.983	0.537	0.537	0.546	0.371	0.537	0.442
SMO-Poly- 10^4	53.170	46.830	3.235	0.532	0.532	0.530	0.371	0.532	0.436
SMO-Poly- 10^5	53.109	46.891	3.215	0.531	0.531	0.530	0.370	0.531	0.436
SMO-RBF-0	1.871	98.129	0.873	0.019	0.019	0.001	0.040	0.019	0.001
SMO-RBF- 10^{-5}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-RBF- 10^{-4}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-RBF- 10^{-3}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-RBF- 10^{-2}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-RBF- 10^{-1}	22.389	77.611	2.001	0.224	0.224	0.009	0.040	0.224	0.015
SMO-RBF- 10^0	39.045	60.955	2.592	0.390	0.390	0.262	0.091	0.390	0.136
SMO-RBF- 10^1	69.825	30.175	3.010	0.698	0.698	0.800	0.463	0.698	0.586
SMO-RBF- 10^2	71.698	28.302	2.878	0.717	0.717	0.772	0.529	0.717	0.628
SMO-RBF- 10^3	71.636	28.364	2.553	0.716	0.716	0.745	0.552	0.716	0.634
SMO-RBF- 10^4	71.154	28.846	2.285	0.712	0.712	0.731	0.551	0.712	0.628
SMO-RBF- 10^5	70.972	29.028	2.209	0.710	0.710	0.727	0.550	0.710	0.626

Table 3.62: Classification Results for Re1 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	80.144	19.856	1.662	0.801	0.801	0.722	0.679	0.801	0.700
IB3	80.927	19.073	1.579	0.809	0.809	0.723	0.637	0.809	0.677
IB5	79.901	20.099	1.968	0.799	0.799	0.719	0.600	0.799	0.654
IB7	78.336	21.665	1.702	0.783	0.783	0.723	0.564	0.783	0.634
IB9	77.550	22.450	2.680	0.775	0.775	0.718	0.537	0.775	0.614
IB11	76.041	23.959	2.093	0.760	0.760	0.707	0.506	0.760	0.590
IB13	75.922	24.078	2.060	0.759	0.759	0.716	0.501	0.759	0.590
IB15	75.016	24.984	1.691	0.750	0.750	0.684	0.475	0.750	0.561
IB17	75.016	24.984	2.151	0.750	0.750	0.720	0.467	0.750	0.567
IB19	74.533	25.467	2.192	0.745	0.745	0.724	0.465	0.745	0.567
IB21	74.595	25.405	2.224	0.746	0.746	0.700	0.459	0.746	0.555
IB25	74.111	25.889	2.313	0.741	0.741	0.709	0.446	0.741	0.547
IB29	73.146	26.854	2.313	0.731	0.731	0.697	0.428	0.731	0.530
IB35	72.421	27.580	2.251	0.724	0.724	0.628	0.409	0.724	0.496
IB41	71.818	28.182	2.484	0.718	0.718	0.584	0.396	0.718	0.472
IB49	71.156	28.844	2.674	0.712	0.712	0.584	0.379	0.712	0.460
IB57	70.189	29.811	2.546	0.702	0.702	0.562	0.364	0.702	0.442
IB73	67.775	32.225	2.905	0.678	0.678	0.525	0.326	0.678	0.402
IB89	66.266	33.734	3.312	0.663	0.663	0.481	0.296	0.663	0.366
IB1-WV	80.144	19.856	1.662	0.801	0.801	0.722	0.679	0.801	0.700
IB3-WV	81.894	18.106	1.190	0.819	0.819	0.749	0.666	0.819	0.705
IB5-WV	81.592	18.408	1.456	0.816	0.816	0.766	0.638	0.816	0.696
IB7-WV	80.869	19.131	1.666	0.809	0.809	0.753	0.612	0.809	0.675
IB9-WV	79.843	20.157	1.695	0.798	0.798	0.761	0.593	0.798	0.667
IB11-WV	79.421	20.579	1.303	0.794	0.794	0.766	0.583	0.794	0.662
IB13-WV	78.455	21.545	1.747	0.785	0.785	0.734	0.558	0.785	0.634
IB15-WV	78.334	21.666	1.702	0.783	0.783	0.750	0.548	0.783	0.633
IB17-WV	78.214	21.786	1.754	0.782	0.782	0.752	0.533	0.782	0.623
IB19-WV	77.671	22.329	1.817	0.777	0.777	0.750	0.525	0.777	0.618
IB21-WV	77.189	22.811	1.982	0.772	0.772	0.751	0.501	0.772	0.601
IB25-WV	76.585	23.415	2.145	0.766	0.766	0.765	0.490	0.766	0.598
IB29-WV	75.982	24.018	2.394	0.760	0.760	0.729	0.469	0.760	0.571
IB35-WV	74.594	25.406	2.235	0.746	0.746	0.700	0.447	0.746	0.546
IB41-WV	74.413	25.587	2.529	0.744	0.744	0.660	0.439	0.744	0.528
IB49-WV	73.449	26.551	2.878	0.734	0.734	0.667	0.419	0.734	0.514
IB57-WV	72.484	27.516	3.244	0.725	0.725	0.632	0.398	0.725	0.488
IB73-WV	70.673	29.327	2.513	0.707	0.707	0.647	0.367	0.707	0.468
IB89-WV	69.104	30.896	2.713	0.691	0.691	0.617	0.341	0.691	0.440

3.31 Re8 Collection

Table 3.63 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.64 presents the classification results obtained by IBk for Re8 collection.

Table 3.63: Classification Results for Re8 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	81.275	18.725	1.355	0.813	0.813	0.642	0.720	0.813	0.679
Multinomial NB	95.335	4.665	0.941	0.953	0.953	0.910	0.892	0.953	0.901
J48	90.735	9.265	0.831	0.907	0.907	0.825	0.814	0.907	0.819
SMO-Linear-0	3.818	96.182	0.849	0.038	0.038	0.005	0.125	0.038	0.009
SMO-Linear- 10^{-5}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-Linear- 10^{-4}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-Linear- 10^{-3}	51.199	48.801	1.922	0.512	0.512	0.189	0.125	0.512	0.151
SMO-Linear- 10^{-2}	81.249	18.751	1.556	0.812	0.812	0.753	0.333	0.812	0.462
SMO-Linear- 10^{-1}	91.960	8.040	1.180	0.920	0.920	0.907	0.695	0.920	0.787
SMO-Linear- 10^0	93.328	6.672	0.805	0.933	0.933	0.878	0.777	0.933	0.824
SMO-Linear- 10^1	93.954	6.046	0.607	0.940	0.940	0.832	0.815	0.940	0.824
SMO-Linear- 10^2	93.758	6.242	0.543	0.938	0.938	0.826	0.811	0.938	0.818
SMO-Linear- 10^3	93.771	6.229	0.532	0.938	0.938	0.827	0.811	0.938	0.819
SMO-Linear- 10^4	93.771	6.229	0.532	0.938	0.938	0.827	0.811	0.938	0.819
SMO-Linear- 10^5	93.771	6.229	0.532	0.938	0.938	0.827	0.811	0.938	0.819
SMO-Poly-0	3.818	96.182	0.849	0.038	0.038	0.005	0.125	0.038	0.009
SMO-Poly- 10^{-5}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-Poly- 10^{-4}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-Poly- 10^{-3}	52.489	47.511	1.922	0.525	0.525	0.547	0.145	0.525	0.229
SMO-Poly- 10^{-2}	64.282	35.718	1.556	0.643	0.643	0.825	0.294	0.643	0.434
SMO-Poly- 10^{-1}	81.418	18.582	1.097	0.814	0.814	0.850	0.498	0.814	0.628
SMO-Poly- 10^0	84.298	15.702	0.658	0.843	0.843	0.798	0.541	0.843	0.645
SMO-Poly- 10^1	86.709	13.292	1.434	0.867	0.867	0.717	0.609	0.867	0.658
SMO-Poly- 10^2	87.112	12.888	0.993	0.871	0.871	0.686	0.641	1.742	0.663
SMO-Poly- 10^3	87.620	12.380	1.090	0.876	0.876	0.691	0.651	0.876	0.670
SMO-Poly- 10^4	87.607	12.393	1.126	0.876	0.876	0.690	0.651	0.876	0.670
SMO-Poly- 10^5	87.607	12.393	1.126	0.876	0.876	0.690	0.651	0.876	0.670
SMO-RBF-0	3.818	96.182	0.849	0.038	0.038	0.005	0.125	0.038	0.009
SMO-RBF- 10^{-5}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-RBF- 10^{-4}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-RBF- 10^{-3}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-RBF- 10^{-2}	51.121	48.879	1.888	0.511	0.511	0.064	0.125	0.511	0.085
SMO-RBF- 10^{-1}	57.428	42.572	1.973	0.574	0.574	0.140	0.154	0.574	0.147
SMO-RBF- 10^0	83.933	16.067	1.441	0.839	0.839	0.746	0.424	0.839	0.541
SMO-RBF- 10^1	83.933	16.067	1.441	0.839	0.839	0.746	0.424	0.839	0.541
SMO-RBF- 10^2	92.494	7.506	0.825	0.925	0.925	0.821	0.768	0.925	0.793
SMO-RBF- 10^3	92.911	7.089	0.777	0.929	0.929	0.789	0.788	0.929	0.788
SMO-RBF- 10^4	92.690	7.310	0.732	0.927	0.927	0.786	0.784	0.927	0.785
SMO-RBF- 10^5	92.703	7.297	0.710	0.927	0.927	0.786	0.784	0.927	0.785

Table 3.64: Classification Results for Re8 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	93.954	6.046	0.688	0.940	0.940	0.895	0.869	0.940	0.882
IB3	93.784	6.216	0.570	0.938	0.938	0.905	0.859	0.938	0.881
IB5	93.706	6.294	0.500	0.937	0.937	0.908	0.848	0.937	0.877
IB7	93.250	6.750	0.447	0.932	0.932	0.908	0.844	0.932	0.874
IB9	93.276	6.724	0.467	0.933	0.933	0.905	0.837	0.933	0.870
IB11	92.872	7.128	0.485	0.929	0.929	0.901	0.824	0.929	0.861
IB13	92.624	7.376	0.519	0.926	0.926	0.904	0.823	0.926	0.861
IB15	92.364	7.636	0.696	0.924	0.924	0.906	0.818	0.924	0.860
IB17	92.259	7.741	0.891	0.923	0.923	0.906	0.808	0.923	0.854
IB19	92.129	7.871	0.815	0.921	0.921	0.915	0.805	0.921	0.856
IB21	92.090	7.910	0.827	0.921	0.921	0.919	0.800	0.921	0.856
IB25	91.882	8.119	0.727	0.919	0.919	0.919	0.787	0.919	0.848
IB29	91.582	8.418	0.775	0.916	0.916	0.914	0.778	0.916	0.840
IB35	91.126	8.874	0.814	0.911	0.911	0.920	0.769	0.911	0.838
IB41	90.813	9.187	0.718	0.908	0.908	0.918	0.763	0.908	0.833
IB49	90.370	9.630	0.927	0.904	0.904	0.917	0.752	0.904	0.826
IB57	89.966	10.034	0.859	0.900	0.900	0.916	0.739	0.900	0.818
IB73	89.380	10.620	0.986	0.894	0.894	0.914	0.722	0.894	0.807
IB89	88.780	11.220	1.129	0.888	0.888	0.911	0.701	0.888	0.792
IB1-WV	93.954	6.046	0.688	0.940	0.940	0.895	0.869	0.940	0.882
IB3-WV	94.149	5.851	0.478	0.941	0.941	0.912	0.866	0.941	0.888
IB5-WV	94.123	5.877	0.455	0.941	0.941	0.919	0.858	0.941	0.888
IB7-WV	93.667	6.333	0.580	0.937	0.937	0.914	0.850	0.937	0.881
IB9-WV	93.602	6.398	0.499	0.936	0.936	0.913	0.847	0.936	0.879
IB11-WV	93.224	6.776	0.531	0.932	0.932	0.908	0.832	0.932	0.868
IB13-WV	92.976	7.024	0.635	0.930	0.930	0.906	0.828	0.930	0.865
IB15-WV	92.664	7.337	0.636	0.927	0.927	0.908	0.822	0.927	0.863
IB17-WV	92.664	7.337	0.737	0.927	0.927	0.909	0.819	0.927	0.861
IB19-WV	92.416	7.584	0.766	0.924	0.924	0.915	0.811	0.924	0.860
IB21-WV	92.390	7.610	0.747	0.924	0.924	0.919	0.809	0.924	0.860
IB25-WV	92.325	7.675	0.682	0.923	0.923	0.921	0.796	0.923	0.854
IB29-WV	91.999	8.001	0.732	0.920	0.920	0.920	0.794	0.920	0.852
IB35-WV	91.569	8.431	0.839	0.916	0.916	0.922	0.776	0.916	0.843
IB41-WV	91.139	8.861	0.811	0.911	0.911	0.918	0.771	0.911	0.838
IB49-WV	90.722	9.278	0.917	0.907	0.907	0.917	0.760	0.907	0.831
IB57-WV	90.474	9.526	0.897	0.905	0.905	0.918	0.750	0.905	0.826
IB73-WV	89.705	10.295	1.072	0.897	0.897	0.913	0.732	0.897	0.813
IB89-WV	89.419	10.581	1.080	0.894	0.894	0.914	0.716	0.894	0.803

3.32 Review-Polarity Collection

Table 3.65 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.66 presents the classification results obtained by IBk for Review-Polarity collection.

Table 3.65: Classification Results for Review-Polarity collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	66.800	33.200	2.088	0.668	0.668	0.670	0.668	0.668	0.669
Multinomial NB	80.100	19.900	3.048	0.801	0.801	0.802	0.801	0.801	0.801
J48	68.250	31.750	3.600	0.683	0.683	0.683	0.683	0.683	0.683
SMO-Linear-0	50.000	50.000	2.470	0.500	0.500	0.250	0.500	0.500	0.333
SMO-Linear- 10^{-5}	48.000	52.000	1.449	0.480	0.480	0.480	0.480	0.480	0.480
SMO-Linear- 10^{-4}	53.800	46.200	9.854	0.538	0.538	0.541	0.538	0.538	0.540
SMO-Linear- 10^{-3}	53.850	46.150	12.145	0.539	0.539	0.542	0.539	0.539	0.540
SMO-Linear- 10^{-2}	83.150	16.850	2.873	0.832	0.832	0.838	0.832	0.832	0.835
SMO-Linear- 10^{-1}	83.200	16.800	2.694	0.832	0.832	0.832	0.832	0.832	0.832
SMO-Linear- 10^0	82.650	17.350	2.490	0.827	0.827	0.827	0.827	0.827	0.827
SMO-Linear- 10^1	82.650	17.350	2.490	0.827	0.827	0.827	0.827	0.827	0.827
SMO-Linear- 10^2	82.650	17.350	2.490	0.827	0.827	0.827	0.827	0.827	0.827
SMO-Linear- 10^3	82.650	17.350	2.490	0.827	0.827	0.827	0.827	0.827	0.827
SMO-Linear- 10^4	82.650	17.350	2.490	0.827	0.827	0.827	0.827	0.827	0.827
SMO-Linear- 10^5	82.650	17.350	2.490	0.827	0.827	0.827	0.827	0.827	0.827
SMO-Poly-0	50.000	50.000	2.470	0.500	0.500	0.250	0.500	0.500	0.333
SMO-Poly- 10^{-5}	51.800	48.200	9.247	0.518	0.518	0.520	0.518	0.518	0.519
SMO-Poly- 10^{-4}	57.350	42.650	10.428	0.574	0.574	0.615	0.574	0.574	0.593
SMO-Poly- 10^{-3}	72.150	27.850	3.668	0.722	0.722	0.778	0.722	0.722	0.749
SMO-Poly- 10^{-2}	77.050	22.950	3.189	0.771	0.771	0.793	0.771	0.771	0.782
SMO-Poly- 10^{-1}	74.550	25.450	3.446	0.746	0.746	0.774	0.746	0.746	0.759
SMO-Poly- 10^0	71.200	28.800	4.112	0.712	0.712	0.745	0.712	0.712	0.728
SMO-Poly- 10^1	71.200	28.800	4.112	0.712	0.712	0.745	0.712	0.712	0.728
SMO-Poly- 10^2	71.200	28.800	4.112	0.712	0.712	0.745	0.712	0.712	0.728
SMO-Poly- 10^3	71.200	28.800	4.112	0.712	0.712	0.745	0.712	0.712	0.728
SMO-Poly- 10^4	71.200	28.800	4.112	0.712	0.712	0.745	0.712	0.712	0.728
SMO-Poly- 10^5	71.200	28.800	4.112	0.712	0.712	0.745	0.712	0.712	0.728
SMO-RBF-0	50.000	50.000	2.470	0.500	0.500	0.250	0.500	0.500	0.333
SMO-RBF- 10^{-5}	48.000	52.000	1.449	0.480	0.480	0.480	0.480	0.480	0.480
SMO-RBF- 10^{-4}	48.000	52.000	1.449	0.480	0.480	0.480	0.480	0.480	0.480
SMO-RBF- 10^{-3}	48.000	52.000	1.449	0.480	0.480	0.480	0.480	0.480	0.480
SMO-RBF- 10^{-2}	49.450	50.550	5.037	0.495	0.495	0.494	0.495	0.495	0.494
SMO-RBF- 10^{-1}	57.700	42.300	2.977	0.577	0.577	0.577	0.577	0.577	0.577
SMO-RBF- 10^0	80.300	19.700	2.804	0.803	0.803	0.803	0.803	0.803	0.803
SMO-RBF- 10^1	83.600	16.400	2.835	0.836	0.836	0.836	0.836	0.836	0.836
SMO-RBF- 10^2	83.650	16.350	3.147	0.837	0.837	0.837	0.837	0.837	0.837
SMO-RBF- 10^3	83.650	16.350	3.147	0.837	0.837	0.837	0.837	0.837	0.837
SMO-RBF- 10^4	83.650	16.350	3.147	0.837	0.837	0.837	0.837	0.837	0.837
SMO-RBF- 10^5	83.650	16.350	3.147	0.837	0.837	0.837	0.837	0.837	0.837

Table 3.66: Classification Results for Review-Polarity collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	67.600	32.400	3.441	0.676	0.676	0.676	0.676	0.676	0.676
IB3	67.950	32.050	3.622	0.680	0.680	0.680	0.680	0.680	0.680
IB5	66.950	33.050	4.574	0.670	0.670	0.670	0.670	0.670	0.670
IB7	68.900	31.100	3.089	0.689	0.689	0.689	0.689	0.689	0.689
IB9	69.000	31.000	3.486	0.690	0.690	0.690	0.690	0.690	0.690
IB11	69.500	30.500	4.062	0.695	0.695	0.695	0.695	0.695	0.695
IB13	69.500	30.500	3.376	0.695	0.695	0.695	0.695	0.695	0.695
IB15	69.850	30.150	3.654	0.699	0.699	0.699	0.698	0.699	0.698
IB17	69.800	30.200	4.605	0.698	0.698	0.698	0.698	0.698	0.698
IB19	70.050	29.950	4.590	0.701	0.701	0.701	0.701	0.701	0.701
IB21	69.400	30.600	4.241	0.694	0.694	0.694	0.694	0.694	0.694
IB25	69.850	30.150	3.668	0.699	0.699	0.699	0.698	0.699	0.698
IB29	69.300	30.700	3.607	0.693	0.693	0.693	0.693	0.693	0.693
IB35	69.250	30.750	4.125	0.693	0.693	0.693	0.693	0.693	0.693
IB41	70.000	30.000	4.410	0.700	0.700	0.700	0.700	0.700	0.700
IB49	69.900	30.100	3.583	0.699	0.699	0.699	0.699	0.699	0.699
IB57	69.450	30.550	3.539	0.695	0.695	0.695	0.695	0.695	0.695
IB73	69.700	30.300	3.600	0.697	0.697	0.697	0.697	0.697	0.697
IB89	69.700	30.300	3.938	0.697	0.697	0.697	0.697	0.697	0.697
IB1-WV	67.600	32.400	3.441	0.676	0.676	0.676	0.676	0.676	0.676
IB3-WV	68.300	31.700	3.458	0.683	0.068	0.683	0.683	0.124	0.683
IB5-WV	67.350	32.650	4.433	0.674	0.674	0.674	0.674	0.674	0.674
IB7-WV	69.300	30.700	2.917	0.693	0.693	0.693	0.693	0.693	0.693
IB9-WV	69.450	30.550	3.312	0.695	0.695	0.695	0.695	0.695	0.695
IB11-WV	69.850	30.150	3.982	0.699	0.699	0.699	0.698	0.699	0.698
IB13-WV	69.900	30.100	3.404	0.699	0.699	0.699	0.699	0.699	0.699
IB15-WV	70.200	29.800	3.716	0.702	0.702	0.702	0.702	0.702	0.702
IB17-WV	70.150	29.850	4.405	0.702	0.702	0.702	0.702	0.702	0.702
IB19-WV	70.450	29.550	4.321	0.705	0.705	0.705	0.705	0.705	0.705
IB21-WV	69.850	30.150	4.050	0.699	0.699	0.699	0.699	0.699	0.699
IB25-WV	70.250	29.750	3.723	0.703	0.703	0.703	0.702	0.703	0.702
IB29-WV	69.950	30.050	3.510	0.700	0.700	0.700	0.700	0.700	0.700
IB35-WV	69.650	30.350	3.988	0.697	0.697	0.697	0.697	0.697	0.697
IB41-WV	70.500	29.500	4.410	0.705	0.705	0.705	0.705	0.705	0.705
IB49-WV	70.350	29.650	3.479	0.704	0.704	0.704	0.704	0.704	0.704
IB57-WV	70.050	29.950	3.574	0.701	0.701	0.701	0.701	0.701	0.701
IB73-WV	69.850	30.150	3.435	0.699	0.699	0.699	0.699	0.699	0.699
IB89-WV	69.950	30.050	3.643	0.700	0.700	0.700	0.700	0.700	0.700

3.33 Reviews Collection

Table 3.67 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.68 presents the classification results obtained by IBk for Reviews collection.

Table 3.67: Classification Results for Reviews collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
Naïve Bayes	85.230	14.771	1.248	0.852	0.852	0.769	0.815	0.852	0.791
Multinomial NB	93.340	6.660	1.051	0.933	0.933	0.922	0.890	0.933	0.906
J48	88.350	11.650	1.704	0.884	0.884	0.845	0.828	0.884	0.836
SMO-Linear-0	3.367	96.633	0.737	0.034	0.034	0.007	0.200	0.034	0.013
SMO-Linear- 10^{-5}	34.112	65.888	2.500	0.341	0.341	0.068	0.200	0.341	0.102
SMO-Linear- 10^{-4}	34.112	65.888	2.500	0.341	0.341	0.068	0.200	0.341	0.102
SMO-Linear- 10^{-3}	38.487	61.513	2.524	0.385	0.385	0.657	0.241	0.385	0.353
SMO-Linear- 10^{-2}	83.215	16.785	2.017	0.832	0.832	0.906	0.676	0.832	0.774
SMO-Linear- 10^{-1}	91.546	8.454	1.731	0.915	0.915	0.922	0.851	0.915	0.885
SMO-Linear- 10^0	91.644	8.356	1.792	0.916	0.916	0.915	0.861	0.916	0.887
SMO-Linear- 10^1	91.546	8.454	1.842	0.915	0.915	0.914	0.860	0.915	0.886
SMO-Linear- 10^2	91.546	8.454	1.842	0.915	0.915	0.914	0.860	0.915	0.886
SMO-Linear- 10^3	91.546	8.454	1.842	0.915	0.915	0.914	0.860	0.915	0.886
SMO-Linear- 10^4	91.546	8.454	1.842	0.915	0.915	0.914	0.860	0.915	0.886
SMO-Linear- 10^5	91.546	8.454	1.842	0.915	0.915	0.914	0.860	0.915	0.886
SMO-Poly-0	3.367	96.633	0.737	0.034	0.034	0.007	0.200	0.034	0.013
SMO-Poly- 10^{-5}	34.579	65.421	2.505	0.346	0.346	0.469	0.208	0.346	0.288
SMO-Poly- 10^{-4}	35.685	64.315	2.422	0.357	0.357	0.633	0.219	0.357	0.325
SMO-Poly- 10^{-3}	51.659	48.341	2.461	0.517	0.517	0.854	0.356	0.517	0.502
SMO-Poly- 10^{-2}	70.927	29.073	2.741	0.709	0.709	0.830	0.565	0.709	0.672
SMO-Poly- 10^{-1}	78.447	21.554	2.400	0.784	0.784	0.843	0.645	0.784	0.731
SMO-Poly- 10^0	80.511	19.489	3.095	0.805	0.805	0.834	0.660	0.805	0.737
SMO-Poly- 10^1	78.865	21.135	3.811	0.789	0.789	0.819	0.650	0.789	0.724
SMO-Poly- 10^2	78.865	21.135	3.811	0.789	0.789	0.819	0.650	0.789	0.724
SMO-Poly- 10^3	78.865	21.135	3.811	0.789	0.789	0.819	0.650	0.789	0.724
SMO-Poly- 10^4	78.865	21.135	3.811	0.789	0.789	0.819	0.650	0.789	0.724
SMO-Poly- 10^5	78.865	21.135	3.811	0.789	0.789	0.819	0.650	0.789	0.724
SMO-RBF-0	3.367	96.633	0.737	0.034	0.034	0.007	0.200	0.034	0.013
SMO-RBF- 10^{-5}	34.112	65.888	0.737	0.341	0.341	0.068	0.200	0.341	0.102
SMO-RBF- 10^{-4}	34.112	65.888	2.500	0.341	0.341	0.068	0.200	0.341	0.102
SMO-RBF- 10^{-3}	34.112	65.888	2.500	0.341	0.341	0.068	0.200	0.341	0.102
SMO-RBF- 10^{-2}	34.112	65.888	2.500	0.341	0.341	0.068	0.200	0.341	0.102
SMO-RBF- 10^{-1}	38.044	61.956	2.500	0.380	0.380	0.500	0.237	0.380	0.322
SMO-RBF- 10^0	83.141	16.859	2.555	0.831	0.831	0.896	0.725	0.831	0.801
SMO-RBF- 10^1	89.654	10.347	2.746	0.897	0.897	0.924	0.821	0.897	0.869
SMO-RBF- 10^2	89.973	10.027	1.259	0.900	0.900	0.923	0.823	0.900	0.870
SMO-RBF- 10^3	89.998	10.002	1.590	0.900	0.900	0.923	0.823	0.900	0.870
SMO-RBF- 10^4	89.998	10.002	1.643	0.900	0.900	0.923	0.823	0.900	0.870
SMO-RBF- 10^5	89.998	10.002	1.643	0.900	0.900	0.923	0.823	0.900	0.870

Table 3.68: Classification Results for Reviews collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	91.227	8.774	1.341	0.912	0.912	0.907	0.857	0.912	0.881
IB3	91.669	8.331	1.480	0.917	0.917	0.900	0.857	0.917	0.878
IB5	92.258	7.742	1.404	0.923	0.923	0.926	0.862	0.923	0.893
IB7	92.209	7.791	1.454	0.922	0.922	0.929	0.859	0.922	0.893
IB9	91.865	8.135	1.352	0.919	0.919	0.928	0.847	0.919	0.886
IB11	91.865	8.135	1.528	0.919	0.919	0.927	0.842	0.919	0.883
IB13	91.546	8.454	1.534	0.915	0.915	0.926	0.830	0.915	0.875
IB15	91.570	8.430	1.592	0.916	0.916	0.925	0.836	0.916	0.878
IB17	91.398	8.602	1.592	0.914	0.914	0.923	0.831	0.914	0.874
IB19	91.398	8.602	1.495	0.914	0.914	0.922	0.829	0.914	0.873
IB21	91.374	8.626	1.548	0.914	0.914	0.920	0.829	0.914	0.872
IB25	90.759	9.241	1.180	0.908	0.908	0.914	0.822	0.908	0.865
IB29	90.587	9.413	1.442	0.906	0.906	0.912	0.823	0.906	0.865
IB35	90.071	9.929	1.288	0.901	0.901	0.905	0.814	0.901	0.857
IB41	89.653	10.347	1.128	0.897	0.897	0.901	0.807	0.897	0.851
IB49	89.137	10.863	1.322	0.891	0.891	0.896	0.800	0.891	0.845
IB57	88.843	11.158	1.412	0.888	0.888	0.892	0.796	0.888	0.841
IB73	88.253	11.747	1.670	0.883	0.883	0.888	0.779	0.883	0.830
IB89	88.401	11.600	1.703	0.884	0.884	0.890	0.781	0.884	0.832
IB1-WV	91.227	8.774	1.341	0.912	0.912	0.907	0.857	0.912	0.881
IB3-WV	91.743	8.258	1.809	0.917	0.917	0.915	0.853	0.917	0.883
IB5-WV	92.308	7.692	1.305	0.923	0.923	0.929	0.856	0.923	0.891
IB7-WV	92.062	7.938	1.503	0.921	0.921	0.928	0.851	0.921	0.888
IB9-WV	92.037	7.963	1.454	0.920	0.920	0.928	0.849	0.920	0.887
IB11-WV	92.160	7.840	1.563	0.922	0.922	0.931	0.844	0.922	0.885
IB13-WV	91.915	8.086	1.496	0.919	0.919	0.929	0.834	0.919	0.879
IB15-WV	92.111	7.889	1.590	0.921	0.921	0.930	0.840	0.921	0.883
IB17-WV	92.160	7.840	1.664	0.922	0.922	0.929	0.836	0.922	0.880
IB19-WV	92.111	7.889	1.574	0.921	0.921	0.928	0.834	0.921	0.879
IB21-WV	91.841	8.159	1.525	0.918	0.918	0.924	0.832	0.918	0.876
IB25-WV	91.349	8.651	1.208	0.913	0.913	0.918	0.827	0.913	0.870
IB29-WV	90.931	9.069	1.379	0.909	0.909	0.915	0.826	0.909	0.868
IB35-WV	90.415	9.585	1.287	0.904	0.904	0.909	0.816	0.904	0.860
IB41-WV	89.948	10.052	1.231	0.899	0.899	0.903	0.810	0.899	0.854
IB49-WV	89.555	10.445	1.264	0.896	0.896	0.900	0.804	0.896	0.849
IB57-WV	89.137	10.863	1.266	0.891	0.891	0.895	0.797	0.891	0.843
IB73-WV	88.720	11.280	1.394	0.887	0.887	0.892	0.784	0.887	0.835
IB89-WV	88.671	11.329	1.502	0.887	0.887	0.892	0.787	0.887	0.836

3.34 SpamAssassin Collection

Table 3.69 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.70 presents the classification results obtained by IBk for SpamAssassin collection.

Table 3.69: Classification Results for SpamAssassin collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	87.805	12.195	0.933	0.878	0.878	0.857	0.809	0.878	0.832
Multinomial NB	96.513	3.487	0.502	0.965	0.965	0.977	0.932	0.965	0.954
J48	96.609	3.391	0.721	0.966	0.966	0.958	0.952	0.966	0.955
SMO-Linear-0	25.642	74.358	1.426	0.256	0.256	0.128	0.500	0.256	0.204
SMO-Linear- 10^{-5}	75.118	24.882	1.388	0.751	0.751	0.870	0.515	0.751	0.647
SMO-Linear- 10^{-4}	90.982	9.018	0.828	0.910	0.910	0.941	0.826	0.910	0.880
SMO-Linear- 10^{-3}	98.042	1.958	0.509	0.980	0.980	0.983	0.965	0.980	0.974
SMO-Linear- 10^{-2}	98.962	1.038	0.481	0.990	0.990	0.989	0.984	0.990	0.986
SMO-Linear- 10^{-1}	74.465	25.535	1.381	0.745	0.745	0.722	0.502	0.745	0.592
SMO-Linear- 10^0	74.358	25.642	1.426	0.744	0.744	0.372	0.500	0.744	0.426
SMO-Linear- 10^1	98.952	1.048	0.320	0.990	0.990	0.988	0.985	0.990	0.986
SMO-Linear- 10^2	98.609	1.391	0.321	0.986	0.986	0.984	0.980	0.986	0.982
SMO-Linear- 10^3	98.566	1.434	0.325	0.986	0.986	0.983	0.979	0.986	0.981
SMO-Linear- 10^4	98.566	1.434	0.325	0.986	0.986	0.983	0.979	0.986	0.981
SMO-Linear- 10^5	98.566	1.434	0.325	0.986	0.986	0.983	0.979	0.986	0.981
SMO-Poly-0	25.642	74.358	1.426	0.256	0.256	0.128	0.500	0.256	0.204
SMO-Poly- 10^{-5}	80.702	19.298	1.540	0.807	0.807	0.889	0.625	0.807	0.734
SMO-Poly- 10^{-4}	91.773	8.227	0.974	0.918	0.918	0.943	0.843	0.918	0.890
SMO-Poly- 10^{-3}	96.341	3.659	0.483	0.963	0.963	0.970	0.934	0.963	0.952
SMO-Poly- 10^{-2}	97.935	2.065	0.441	0.979	0.979	0.976	0.969	0.979	0.973
SMO-Poly- 10^{-1}	76.316	23.684	1.284	0.763	0.763	0.877	0.538	0.763	0.667
SMO-Poly- 10^0	74.882	25.118	1.448	0.749	0.749	0.857	0.510	0.749	0.640
SMO-Poly- 10^1	98.032	1.968	0.332	0.980	0.980	0.974	0.974	0.980	0.974
SMO-Poly- 10^2	97.721	2.279	0.371	0.977	0.977	0.970	0.970	0.977	0.970
SMO-Poly- 10^3	97.614	2.386	0.365	0.976	0.976	0.968	0.969	0.976	0.969
SMO-Poly- 10^4	97.614	2.386	0.365	0.976	0.976	0.968	0.969	0.976	0.969
SMO-Poly- 10^5	97.614	2.386	0.365	0.976	0.976	0.968	0.969	0.976	0.969
SMO-RBF-0	25.642	74.358	1.426	0.256	0.256	0.128	0.500	0.256	0.204
SMO-RBF- 10^{-5}	74.358	25.642	1.426	0.744	0.744	0.372	0.500	0.744	0.426
SMO-RBF- 10^{-4}	74.358	25.642	1.426	0.744	0.744	0.372	0.500	0.744	0.426
SMO-RBF- 10^{-3}	74.690	25.310	1.364	0.747	0.747	0.873	0.506	0.747	0.641
SMO-RBF- 10^{-2}	93.699	6.301	0.937	0.937	0.937	0.951	0.883	0.937	0.916
SMO-RBF- 10^{-1}	74.358	25.642	1.426	0.744	0.744	0.372	0.500	0.744	0.426
SMO-RBF- 10^0	74.358	25.642	1.426	0.744	0.744	0.372	0.500	0.744	0.426
SMO-RBF- 10^1	98.470	1.530	0.533	0.985	0.985	0.987	0.973	0.985	0.980
SMO-RBF- 10^2	98.770	1.230	0.375	0.988	0.988	0.989	0.979	0.988	0.984
SMO-RBF- 10^3	98.502	1.498	0.262	0.985	0.985	0.986	0.975	0.985	0.980
SMO-RBF- 10^4	98.235	1.765	0.332	0.982	0.982	0.982	0.971	0.982	0.977
SMO-RBF- 10^5	98.235	1.765	0.332	0.982	0.982	0.982	0.971	0.982	0.977

Table 3.70: Classification Results for SpamAssassin collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	98.727	1.273	0.404	0.987	0.987	0.985	0.982	0.987	0.983
IB3	97.636	2.364	0.475	0.976	0.976	0.971	0.966	0.976	0.969
IB5	97.700	2.300	0.484	0.977	0.977	0.974	0.965	0.977	0.970
IB7	97.176	2.824	0.404	0.972	0.972	0.970	0.955	0.972	0.963
IB9	97.080	2.920	0.441	0.971	0.971	0.969	0.954	0.971	0.961
IB11	96.769	3.231	0.568	0.968	0.968	0.966	0.948	0.968	0.957
IB13	96.502	3.498	0.590	0.965	0.965	0.964	0.943	0.965	0.954
IB15	96.448	3.552	0.666	0.964	0.964	0.963	0.943	0.964	0.953
IB17	-	-	-	-	-	-	-	-	-
IB19	95.924	4.076	0.812	0.959	0.959	0.959	0.933	0.959	0.946
IB21	95.785	4.215	0.819	0.958	0.958	0.957	0.931	0.958	0.944
IB25	95.314	4.686	0.821	0.953	0.953	0.953	0.923	0.953	0.938
IB29	95.101	4.899	0.803	0.951	0.951	0.951	0.919	0.951	0.935
IB35	94.951	5.049	0.838	0.950	0.950	0.949	0.917	0.950	0.933
IB41	94.598	5.402	0.759	0.946	0.946	0.946	0.910	0.946	0.928
IB49	94.416	5.584	0.691	0.944	0.944	0.947	0.905	0.944	0.925
IB57	94.234	5.766	0.625	0.942	0.942	0.947	0.900	0.942	0.923
IB73	93.860	6.140	0.794	0.939	0.939	0.944	0.893	0.939	0.918
IB89	93.614	6.386	0.672	0.936	0.936	0.942	0.888	0.936	0.915
IB1-WV	98.716	1.284	0.400	0.987	0.987	0.985	0.981	0.987	0.983
IB3-WV	98.427	1.573	0.332	0.984	0.984	0.983	0.976	0.984	0.979
IB5-WV	98.085	1.915	0.494	0.981	0.981	0.980	0.969	0.981	0.975
IB7-WV	97.903	2.097	0.377	0.979	0.979	0.979	0.965	0.979	0.972
IB9-WV	97.604	2.396	0.423	0.976	0.976	0.977	0.960	0.976	0.968
IB11-WV	97.400	2.600	0.537	0.974	0.974	0.974	0.958	0.974	0.966
IB13-WV	97.133	2.867	0.555	0.971	0.971	0.972	0.952	0.971	0.962
IB15-WV	93.033	6.967	14.537	0.930	0.930	0.956	0.931	0.930	0.943
IB17-WV	93.050	6.950	14.538	0.931	0.931	0.956	0.931	0.931	0.944
IB19-WV	96.737	3.263	0.678	0.967	0.967	0.969	0.945	0.967	0.957
IB21-WV	96.555	3.445	0.657	0.966	0.966	0.966	0.943	0.966	0.954
IB25-WV	96.138	3.862	0.654	0.961	0.961	0.963	0.935	0.961	0.949
IB29-WV	95.839	4.161	0.679	0.958	0.958	0.960	0.930	0.958	0.945
IB35-WV	95.764	4.236	0.737	0.958	0.958	0.960	0.928	0.958	0.944
IB41-WV	95.389	4.611	0.694	0.954	0.954	0.955	0.923	0.954	0.939
IB49-WV	95.165	4.835	0.695	0.952	0.952	0.954	0.918	0.952	0.936
IB57-WV	94.929	5.071	0.692	0.949	0.949	0.953	0.912	0.949	0.932
IB73-WV	94.437	5.563	0.738	0.944	0.944	0.950	0.903	0.944	0.926
IB89-WV	94.095	5.905	0.813	0.941	0.941	0.947	0.897	0.941	0.921

3.35 SpamTrec-3000 Collection

Table 3.71 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.72 presents the classification results obtained by IBk for SpamTrec-3000 collection.

Table 3.71: Classification Results for SpamTrec-3000 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	81.417	18.583	1.687	0.814	0.814	0.838	0.814	0.814	0.826
Multinomial NB	97.417	2.583	0.479	0.974	0.974	0.975	0.974	0.974	0.975
J48	96.883	3.117	0.796	0.969	0.969	0.969	0.969	0.969	0.969
SMO-Linear-0	50.000	50.000	2.257	0.500	0.500	0.250	0.500	0.500	0.333
SMO-Linear- 10^{-5}	52.450	47.550	2.963	0.525	0.525	0.756	0.525	0.525	0.620
SMO-Linear- 10^{-4}	55.050	44.950	2.333	0.551	0.551	0.759	0.551	0.551	0.638
SMO-Linear- 10^{-3}	64.617	35.383	2.390	0.646	0.646	0.786	0.646	0.646	0.709
SMO-Linear- 10^{-2}	89.167	10.833	1.295	0.892	0.892	0.909	0.892	0.892	0.900
SMO-Linear- 10^{-1}	95.850	4.150	0.612	0.959	0.959	0.960	0.958	0.959	0.959
SMO-Linear- 10^0	97.717	2.283	0.768	0.977	0.977	0.977	0.977	0.977	0.977
SMO-Linear- 10^1	97.367	2.633	0.323	0.974	0.974	0.974	0.974	0.974	0.974
SMO-Linear- 10^2	96.100	3.900	0.416	0.961	0.961	0.961	0.961	0.961	0.961
SMO-Linear- 10^3	96.550	3.450	0.460	0.966	0.966	0.966	0.965	0.966	0.966
SMO-Linear- 10^4	96.433	3.567	0.461	0.964	0.964	0.964	0.964	0.964	0.964
SMO-Linear- 10^5	96.450	3.550	0.448	0.965	0.965	0.964	0.965	0.965	0.965
SMO-Poly-0	50.000	50.000	2.257	0.500	0.500	0.250	0.500	0.500	0.333
SMO-Poly- 10^{-5}	56.467	43.533	2.117	0.565	0.565	0.758	0.565	0.565	0.647
SMO-Poly- 10^{-4}	60.983	39.017	2.265	0.610	0.610	0.777	0.610	0.610	0.683
SMO-Poly- 10^{-3}	71.133	28.867	1.142	0.711	0.711	0.815	0.711	0.711	0.759
SMO-Poly- 10^{-2}	84.000	16.000	1.147	0.840	0.840	0.876	0.840	0.840	0.857
SMO-Poly- 10^{-1}	90.800	9.200	1.251	0.908	0.908	0.920	0.908	0.908	0.914
SMO-Poly- 10^0	93.833	6.167	0.888	0.938	0.938	0.942	0.938	0.938	0.940
SMO-Poly- 10^1	88.417	11.583	10.063	0.884	0.884	0.904	0.886	0.884	0.895
SMO-Poly- 10^2	90.050	9.950	0.946	0.901	0.901	0.903	0.901	0.901	0.902
SMO-Poly- 10^3	86.850	13.150	9.797	0.869	0.869	0.888	0.870	0.869	0.879
SMO-Poly- 10^4	90.950	9.050	0.785	0.910	0.910	0.910	0.909	0.910	0.910
SMO-Poly- 10^5	92.150	7.850	0.603	0.922	0.922	0.922	0.921	0.922	0.922
SMO-RBF-0	50.000	50.000	2.257	0.500	0.500	0.250	0.500	0.500	0.333
SMO-RBF- 10^{-5}	48.067	51.933	1.165	0.481	0.481	0.240	0.500	0.481	0.325
SMO-RBF- 10^{-4}	48.067	51.933	1.165	0.481	0.481	0.240	0.500	0.481	0.325
SMO-RBF- 10^{-3}	48.067	51.933	1.165	0.481	0.481	0.240	0.500	0.481	0.325
SMO-RBF- 10^{-2}	59.650	40.350	1.883	0.597	0.597	0.659	0.596	0.597	0.626
SMO-RBF- 10^{-1}	70.400	29.600	1.393	0.704	0.704	0.748	0.704	0.704	0.725
SMO-RBF- 10^0	89.633	10.367	1.035	0.896	0.896	0.900	0.896	0.896	0.898
SMO-RBF- 10^1	94.250	5.750	1.106	0.943	0.943	0.943	0.942	0.943	0.942
SMO-RBF- 10^2	96.517	3.483	0.456	0.965	0.965	0.967	0.965	0.965	0.966
SMO-RBF- 10^3	96.300	3.700	0.600	0.963	0.963	0.965	0.963	0.963	0.964
SMO-RBF- 10^4	95.783	4.217	0.448	0.958	0.958	0.960	0.958	0.958	0.959
SMO-RBF- 10^5	95.733	4.267	0.517	0.957	0.957	0.959	0.957	0.957	0.958

Table 3.72: Classification Results for SpamTrec-3000 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	98.317	1.683	0.411	0.983	0.983	0.983	0.983	0.983	0.983
IB3	97.883	2.117	0.553	0.979	0.979	0.979	0.979	0.979	0.979
IB5	97.650	2.350	0.540	0.977	0.977	0.977	0.977	0.977	0.977
IB7	97.483	2.517	0.669	0.975	0.975	0.975	0.975	0.975	0.975
IB9	97.533	2.467	0.752	0.975	0.975	0.976	0.975	0.975	0.976
IB11	97.483	2.517	0.783	0.975	0.975	0.975	0.975	0.975	0.975
IB13	97.433	2.567	0.870	0.974	0.974	0.975	0.974	0.974	0.975
IB15	97.267	2.733	0.883	0.973	0.973	0.973	0.973	0.973	0.973
IB17	97.133	2.867	0.865	0.971	0.971	0.972	0.971	0.971	0.972
IB19	92.183	7.817	14.420	0.922	0.922	0.898	0.923	0.922	0.910
IB21	97.067	2.933	0.860	0.971	0.971	0.971	0.971	0.971	0.971
IB25	95.248	4.752	4.534	0.952	0.952	0.951	0.958	0.952	0.955
IB29	96.550	3.450	0.782	0.965	0.965	0.966	0.966	0.965	0.966
IB35	96.595	3.405	0.780	0.966	0.966	0.967	0.964	0.966	0.966
IB41	96.250	3.750	0.750	0.963	0.963	0.963	0.963	0.963	0.963
IB49	94.950	5.050	2.646	0.950	0.950	0.951	0.952	0.950	0.951
IB57	95.550	4.450	0.949	0.956	0.956	0.957	0.956	0.956	0.956
IB73	84.646	15.354	5.090	0.846	0.846	0.854	0.877	0.846	0.866
IB89	93.883	6.117	1.126	0.939	0.939	0.942	0.939	0.939	0.940
IB1-WV	98.617	1.383	0.325	0.986	0.986	0.986	0.986	0.986	0.986
IB3-WV	98.383	1.617	0.533	0.984	0.984	0.984	0.984	0.984	0.984
IB5-WV	98.200	1.800	0.470	0.982	0.982	0.982	0.982	0.982	0.982
IB7-WV	98.083	1.917	0.544	0.981	0.981	0.981	0.981	0.981	0.981
IB9-WV	98.033	1.967	0.748	0.980	0.980	0.981	0.980	0.980	0.980
IB11-WV	97.950	2.050	0.771	0.980	0.980	0.980	0.980	0.980	0.980
IB13-WV	93.083	6.917	14.548	0.931	0.931	0.956	0.932	0.931	0.944
IB15-WV	93.033	6.967	14.537	0.930	0.930	0.956	0.931	0.930	0.943
IB17-WV	93.050	6.950	14.538	0.931	0.931	0.956	0.931	0.931	0.944
IB19-WV	92.933	7.067	14.499	0.929	0.929	0.955	0.930	0.929	0.942
IB21-WV	92.900	7.100	14.490	0.929	0.929	0.954	0.930	0.929	0.942
IB25-WV	92.345	7.655	15.644	0.923	0.923	0.945	0.941	0.923	0.943
IB29-WV	92.667	7.333	14.401	0.927	0.927	0.952	0.928	0.927	0.940
IB35-WV	90.883	9.117	19.194	0.909	0.909	0.941	0.929	0.909	0.935
IB41-WV	50.300	49.700	2.377	0.503	0.503	0.751	0.503	0.503	0.602
IB49-WV	50.350	49.650	2.365	0.504	0.504	0.748	0.513	0.504	0.608
IB57-WV	50.300	49.700	2.377	0.503	0.503	0.751	0.503	0.503	0.602
IB73-WV	47.267	52.733	10.162	0.473	0.473	0.710	0.579	0.473	0.638
IB89-WV	50.300	49.700	2.377	0.503	0.503	0.751	0.503	0.503	0.602

3.36 SyskillWebert Collection

Table 3.73 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.74 presents the classification results obtained by IBk for SyskillWebert collection.

Table 3.73: Classification Results for SyskillWebert collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	72.513	27.487	5.136	0.725	0.725	0.685	0.694	0.725	0.689
Multinomial NB	90.758	9.243	5.701	0.907	0.907	0.893	0.890	0.907	0.892
J48	95.820	4.180	1.968	0.958	0.958	0.972	0.948	0.958	0.960
SMO-Linear-0	21.266	78.735	3.431	0.213	0.213	0.053	0.250	0.213	0.088
SMO-Linear- 10^{-5}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-Linear- 10^{-4}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-Linear- 10^{-3}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-Linear- 10^{-2}	52.460	47.540	7.630	0.524	0.524	0.721	0.404	0.524	0.518
SMO-Linear- 10^{-1}	76.979	23.021	4.279	0.769	0.769	0.803	0.712	0.769	0.755
SMO-Linear- 10^0	77.852	22.148	6.269	0.778	0.778	0.764	0.756	0.778	0.760
SMO-Linear- 10^1	73.084	26.916	5.887	0.731	0.731	0.725	0.724	0.731	0.725
SMO-Linear- 10^2	71.551	28.449	5.639	0.716	0.716	0.722	0.708	0.716	0.715
SMO-Linear- 10^3	71.551	28.449	5.639	0.716	0.716	0.722	0.708	0.716	0.715
SMO-Linear- 10^4	71.551	28.449	5.639	0.716	0.716	0.722	0.708	0.716	0.715
SMO-Linear- 10^5	71.551	28.449	5.639	0.716	0.716	0.722	0.708	0.716	0.715
SMO-Poly-0	21.266	78.735	3.431	0.213	0.213	0.053	0.250	0.213	0.088
SMO-Poly- 10^{-5}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-Poly- 10^{-4}	41.658	58.342	5.392	0.416	0.416	0.478	0.258	0.416	0.335
SMO-Poly- 10^{-3}	45.258	54.742	7.072	0.452	0.452	0.808	0.305	0.452	0.443
SMO-Poly- 10^{-2}	52.433	47.567	5.238	0.524	0.524	0.700	0.396	0.524	0.506
SMO-Poly- 10^{-1}	56.595	43.405	7.634	0.566	0.566	0.601	0.524	0.566	0.560
SMO-Poly- 10^0	46.676	53.325	8.982	0.467	0.467	0.651	0.490	0.467	0.559
SMO-Poly- 10^1	44.875	55.125	9.549	0.449	0.449	0.614	0.479	0.449	0.538
SMO-Poly- 10^2	43.084	56.916	8.310	0.431	0.431	0.564	0.446	0.431	0.498
SMO-Poly- 10^3	45.196	54.804	6.170	0.452	0.452	0.597	0.470	0.452	0.526
SMO-Poly- 10^4	45.196	54.804	6.170	0.452	0.452	0.597	0.470	0.452	0.526
SMO-Poly- 10^5	45.196	54.804	6.170	0.452	0.452	0.597	0.470	0.452	0.526
SMO-RBF-0	21.266	78.735	3.431	0.213	0.213	0.053	0.250	0.213	0.088
SMO-RBF- 10^{-5}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-RBF- 10^{-4}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-RBF- 10^{-3}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-RBF- 10^{-2}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-RBF- 10^{-1}	41.061	58.939	5.821	0.410	0.410	0.103	0.250	0.410	0.145
SMO-RBF- 10^0	51.292	48.708	8.318	0.512	0.512	0.766	0.427	0.512	0.548
SMO-RBF- 10^1	71.649	28.351	9.965	0.716	0.716	0.702	0.683	0.716	0.692
SMO-RBF- 10^2	70.401	29.599	5.982	0.704	0.704	0.676	0.682	0.704	0.679
SMO-RBF- 10^3	68.324	31.676	7.432	0.683	0.683	0.658	0.658	0.683	0.658
SMO-RBF- 10^4	68.619	31.382	7.935	0.686	0.686	0.669	0.661	0.686	0.665
SMO-RBF- 10^5	68.619	31.382	7.935	0.686	0.686	0.669	0.661	0.686	0.665

Table 3.74: Classification Results for SyskillWebert collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	92.531	7.469	2.719	0.925	0.925	0.919	0.922	0.925	0.920
IB3	94.626	5.374	3.192	0.946	0.946	0.943	0.941	0.946	0.942
IB5	95.223	4.777	2.368	0.952	0.952	0.959	0.943	0.952	0.951
IB7	94.929	5.071	2.652	0.949	0.949	0.959	0.938	0.949	0.948
IB9	94.635	5.366	3.165	0.946	0.946	0.958	0.934	0.946	0.945
IB11	94.029	5.972	3.240	0.940	0.940	0.952	0.927	0.940	0.939
IB13	94.626	5.374	2.574	0.946	0.946	0.960	0.934	0.946	0.947
IB15	88.734	11.266	15.958	0.886	0.886	0.929	0.858	0.886	0.892
IB17	91.676	8.325	8.162	0.916	0.916	0.939	0.895	0.916	0.917
IB19	91.373	8.628	7.345	0.913	0.913	0.935	0.894	0.913	0.914
IB21	91.373	8.628	7.345	0.913	0.913	0.940	0.894	0.913	0.916
IB25	89.893	10.107	10.007	0.898	0.898	0.929	0.873	0.898	0.900
IB29	88.717	11.284	13.371	0.886	0.886	0.927	0.858	0.886	0.891
IB35	88.119	11.881	14.070	0.880	0.880	0.922	0.851	0.880	0.885
IB41	88.717	11.284	13.439	0.886	0.886	0.928	0.859	0.886	0.892
IB49	46.649	53.351	14.575	0.467	0.467	0.818	0.322	0.467	0.462
IB57	88.431	11.569	15.285	0.883	0.883	0.928	0.854	0.883	0.890
IB73	87.834	12.166	15.941	0.877	0.877	0.926	0.847	0.877	0.885
IB89	86.319	13.681	15.191	0.862	0.862	0.912	0.829	0.862	0.868
IB1-WV	92.825	7.175	2.715	0.928	0.928	0.921	0.925	0.928	0.923
IB3-WV	95.232	4.768	2.714	0.952	0.952	0.949	0.943	0.952	0.946
IB5-WV	95.526	4.474	2.737	0.955	0.955	0.954	0.947	0.955	0.951
IB7-WV	95.820	4.180	3.028	0.958	0.958	0.958	0.949	0.958	0.954
IB9-WV	95.223	4.777	2.708	0.955	0.955	0.958	0.945	0.955	0.952
IB11-WV	94.029	5.972	2.961	0.943	0.943	0.949	0.930	0.943	0.940
IB13-WV	94.332	5.669	2.443	0.946	0.946	0.958	0.934	0.946	0.946
IB15-WV	88.734	11.266	15.958	0.889	0.889	0.930	0.861	0.889	0.894
IB17-WV	90.499	9.501	10.716	0.907	0.907	0.930	0.885	0.907	0.907
IB19-WV	90.196	9.804	10.658	0.904	0.904	0.926	0.882	0.904	0.903
IB21-WV	90.196	9.804	10.658	0.904	0.904	0.926	0.882	0.904	0.903
IB25-WV	89.305	10.695	12.395	0.895	0.895	0.917	0.870	0.895	0.893
IB29-WV	88.717	11.284	13.241	0.889	0.889	0.928	0.862	0.889	0.893
IB35-WV	88.414	11.587	13.082	0.886	0.886	0.923	0.859	0.886	0.890
IB41-WV	89.011	10.989	12.456	0.892	0.892	0.929	0.866	0.892	0.896
IB49-WV	47.237	52.763	15.500	0.476	0.476	0.822	0.333	0.476	0.474
IB57-WV	74.590	25.410	10.951	0.749	0.749	0.809	0.696	0.749	0.748
IB73-WV	74.884	25.116	10.138	0.751	0.751	0.820	0.700	0.751	0.755
IB89-WV	73.993	26.007	11.693	0.743	0.743	0.824	0.688	0.743	0.750

3.37 Tr11 Collection

Table 3.75 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.76 presents the classification results obtained by IBk for Tr11 collection.

Table 3.75: Classification Results for Tr11 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	54.065	45.935	7.626	0.541	0.541	0.470	0.543	0.541	0.504
Multinomial NB	85.006	14.994	3.637	0.850	0.850	0.724	0.703	0.850	0.713
J48	78.984	21.016	6.115	0.790	0.790	0.680	0.647	0.790	0.663
SMO-Linear-0	5.070	94.930	4.488	0.051	0.051	0.006	0.111	0.051	0.011
SMO-Linear- 10^{-5}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-Linear- 10^{-4}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-Linear- 10^{-3}	33.839	66.162	10.316	0.338	0.338	0.147	0.124	0.338	0.135
SMO-Linear- 10^{-2}	37.712	62.288	11.103	0.377	0.377	0.361	0.154	0.377	0.216
SMO-Linear- 10^{-1}	70.790	29.210	6.416	0.708	0.708	0.642	0.511	0.708	0.569
SMO-Linear- 10^0	77.067	22.933	5.444	0.771	0.771	0.671	0.622	0.771	0.645
SMO-Linear- 10^1	76.591	23.409	5.500	0.766	0.766	0.659	0.615	0.766	0.636
SMO-Linear- 10^2	76.591	23.409	5.500	0.766	0.766	0.659	0.615	0.766	0.636
SMO-Linear- 10^3	76.591	23.409	5.500	0.766	0.766	0.659	0.615	0.766	0.636
SMO-Linear- 10^4	76.591	23.409	5.500	0.766	0.766	0.659	0.615	0.766	0.636
SMO-Linear- 10^5	76.591	23.409	5.500	0.766	0.766	0.659	0.615	0.766	0.636
SMO-Poly-0	5.070	94.930	4.488	0.051	0.051	0.006	0.111	0.051	0.011
SMO-Poly- 10^{-5}	33.362	66.638	9.931	0.333	0.333	0.136	0.123	0.333	0.129
SMO-Poly- 10^{-4}	33.844	66.156	9.945	0.338	0.338	0.247	0.131	0.338	0.171
SMO-Poly- 10^{-3}	36.266	63.734	9.814	0.362	0.362	0.456	0.150	0.362	0.225
SMO-Poly- 10^{-2}	56.063	43.937	9.381	0.560	0.560	0.533	0.337	0.560	0.413
SMO-Poly- 10^{-1}	65.714	34.286	6.627	0.657	0.657	0.629	0.464	0.657	0.534
SMO-Poly- 10^0	66.202	33.798	7.671	0.662	0.662	0.608	0.476	0.662	0.534
SMO-Poly- 10^1	66.202	33.798	7.671	0.662	0.662	0.608	0.476	0.662	0.534
SMO-Poly- 10^2	66.202	33.798	7.671	0.662	0.662	0.608	0.476	0.662	0.534
SMO-Poly- 10^3	66.202	33.798	7.671	0.662	0.662	0.608	0.476	0.662	0.534
SMO-Poly- 10^4	66.202	33.798	7.671	0.662	0.662	0.608	0.476	0.662	0.534
SMO-Poly- 10^5	66.202	33.798	7.671	0.662	0.662	0.608	0.476	0.662	0.534
SMO-RBF-0	5.070	94.930	4.488	0.051	0.051	0.006	0.111	0.051	0.011
SMO-RBF- 10^{-5}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-RBF- 10^{-4}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-RBF- 10^{-3}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-RBF- 10^{-2}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-RBF- 10^{-1}	31.916	68.084	8.947	0.319	0.319	0.035	0.111	0.319	0.054
SMO-RBF- 10^0	36.028	63.972	10.170	0.360	0.360	0.468	0.146	0.360	0.222
SMO-RBF- 10^1	69.826	30.174	4.751	0.698	0.698	0.675	0.502	0.698	0.575
SMO-RBF- 10^2	71.289	28.711	5.239	0.713	0.713	0.655	0.535	0.713	0.589
SMO-RBF- 10^3	71.289	28.711	5.239	0.713	0.713	0.655	0.535	0.713	0.589
SMO-RBF- 10^4	71.289	28.711	5.239	0.713	0.713	0.655	0.535	0.713	0.589
SMO-RBF- 10^5	71.289	28.711	5.239	0.713	0.713	0.655	0.535	0.713	0.589

Table 3.76: Classification Results for Tr11 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	85.738	14.263	4.283	0.857	0.857	0.776	0.770	0.857	0.773
IB3	84.286	15.714	4.270	0.843	0.843	0.702	0.713	0.843	0.708
IB5	82.828	17.172	3.435	0.829	0.829	0.676	0.671	0.829	0.673
IB7	83.804	16.196	3.297	0.838	0.838	0.701	0.700	0.838	0.700
IB9	83.322	16.678	3.851	0.833	0.833	0.699	0.677	0.833	0.688
IB11	82.846	17.155	2.782	0.829	0.829	0.698	0.653	0.829	0.675
IB13	82.358	17.642	2.714	0.824	0.824	0.693	0.639	0.824	0.665
IB15	81.632	18.368	3.662	0.816	0.816	0.668	0.623	0.816	0.645
IB17	82.596	17.404	4.206	0.826	0.826	0.699	0.623	0.826	0.659
IB19	81.876	18.124	4.636	0.819	0.819	0.683	0.612	0.819	0.646
IB21	82.602	17.398	4.562	0.826	0.826	0.686	0.618	0.826	0.650
IB25	82.108	17.892	5.151	0.821	0.821	0.669	0.600	0.821	0.633
IB29	80.912	19.088	5.040	0.809	0.809	0.672	0.572	0.809	0.618
IB35	80.180	19.820	4.989	0.802	0.802	0.606	0.549	0.802	0.576
IB41	79.925	20.076	5.834	0.800	0.800	0.609	0.541	0.800	0.573
IB49	80.163	19.837	5.700	0.802	0.802	0.624	0.540	0.802	0.579
IB57	78.467	21.533	6.820	0.785	0.785	0.554	0.502	0.785	0.527
IB73	75.348	24.652	5.721	0.754	0.754	0.443	0.429	0.754	0.436
IB89	75.093	24.907	6.108	0.751	0.751	0.443	0.420	0.751	0.431
IB1-WV	85.738	14.263	4.283	0.857	0.857	0.776	0.770	0.857	0.773
IB3-WV	86.957	13.043	4.212	0.870	0.870	0.790	0.773	0.870	0.781
IB5-WV	84.292	15.709	5.036	0.843	0.843	0.695	0.703	0.843	0.699
IB7-WV	84.286	15.714	2.971	0.843	0.843	0.698	0.714	0.843	0.705
IB9-WV	84.779	15.221	3.237	0.848	0.848	0.707	0.710	0.848	0.708
IB11-WV	84.791	15.209	2.589	0.848	0.848	0.717	0.692	0.848	0.705
IB13-WV	84.309	15.691	1.858	0.843	0.843	0.721	0.685	0.843	0.703
IB15-WV	83.804	16.196	2.506	0.838	0.838	0.729	0.666	0.838	0.696
IB17-WV	83.084	16.916	3.463	0.831	0.831	0.708	0.637	0.831	0.670
IB19-WV	82.358	17.642	3.458	0.824	0.824	0.694	0.623	0.824	0.657
IB21-WV	83.328	16.673	3.513	0.833	0.833	0.745	0.634	0.833	0.685
IB25-WV	83.316	16.684	4.842	0.833	0.833	0.740	0.635	0.833	0.683
IB29-WV	82.114	17.886	5.363	0.821	0.821	0.732	0.596	0.821	0.657
IB35-WV	82.120	17.880	4.754	0.821	0.821	0.741	0.591	0.821	0.657
IB41-WV	82.114	17.886	4.147	0.821	0.821	0.750	0.586	0.821	0.658
IB49-WV	81.127	18.874	5.810	0.812	0.812	0.634	0.554	0.812	0.591
IB57-WV	80.401	19.599	6.073	0.804	0.804	0.618	0.538	0.804	0.575
IB73-WV	77.991	22.009	6.545	0.780	0.780	0.599	0.484	0.780	0.535
IB89-WV	77.021	22.979	6.534	0.771	0.771	0.448	0.459	0.771	0.453

3.38 Tr12 Collection

Table 3.77 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.78 presents the classification results obtained by IBk for Tr12 collection.

Table 3.77: Classification Results for Tr12 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	57.823	42.178	9.397	0.578	0.578	0.584	0.648	0.578	0.614
Multinomial NB	80.151	19.849	7.097	0.802	0.802	0.819	0.791	0.802	0.804
J48	79.234	20.766	7.451	0.792	0.792	0.796	0.763	0.792	0.779
SMO-Linear-0	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-Linear- 10^{-5}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-Linear- 10^{-4}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-Linear- 10^{-3}	29.657	70.343	10.304	0.297	0.297	0.162	0.128	0.297	0.143
SMO-Linear- 10^{-2}	32.571	67.430	8.152	0.326	0.326	0.336	0.155	0.326	0.212
SMO-Linear- 10^{-1}	67.369	32.631	7.560	0.674	0.674	0.774	0.595	0.674	0.673
SMO-Linear- 10^0	69.617	30.383	7.477	0.696	0.696	0.734	0.632	0.696	0.680
SMO-Linear- 10^1	69.617	30.383	7.477	0.696	0.696	0.734	0.632	0.696	0.680
SMO-Linear- 10^2	69.617	30.383	7.477	0.696	0.696	0.734	0.632	0.696	0.680
SMO-Linear- 10^3	69.617	30.383	7.477	0.696	0.696	0.734	0.632	0.696	0.680
SMO-Linear- 10^4	69.617	30.383	7.477	0.696	0.696	0.734	0.632	0.696	0.680
SMO-Linear- 10^5	69.617	30.383	7.477	0.696	0.696	0.734	0.632	0.696	0.680
SMO-Poly-0	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-Poly- 10^{-5}	29.345	70.655	9.774	0.294	0.294	0.120	0.126	0.294	0.123
SMO-Poly- 10^{-4}	29.990	70.010	9.943	0.300	0.300	0.137	0.130	0.300	0.133
SMO-Poly- 10^{-3}	34.476	65.524	8.587	0.345	0.345	0.523	0.177	0.345	0.264
SMO-Poly- 10^{-2}	49.163	50.837	9.026	0.492	0.492	0.733	0.377	0.492	0.498
SMO-Poly- 10^{-1}	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-Poly- 10^0	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-Poly- 10^1	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-Poly- 10^2	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-Poly- 10^3	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-Poly- 10^4	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-Poly- 10^5	55.252	44.748	7.442	0.553	0.553	0.710	0.464	0.553	0.561
SMO-RBF-0	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-RBF- 10^{-5}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-RBF- 10^{-4}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-RBF- 10^{-3}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-RBF- 10^{-2}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-RBF- 10^{-1}	29.657	70.343	10.603	0.297	0.297	0.037	0.125	0.297	0.057
SMO-RBF- 10^0	30.323	69.678	9.385	0.304	0.304	0.083	0.134	0.304	0.103
SMO-RBF- 10^1	65.464	34.536	9.221	0.655	0.655	0.821	0.554	0.655	0.662
SMO-RBF- 10^2	66.099	33.901	8.137	0.661	0.661	0.801	0.569	0.661	0.665
SMO-RBF- 10^3	66.099	33.901	8.137	0.661	0.661	0.801	0.569	0.661	0.665
SMO-RBF- 10^4	66.099	33.901	8.137	0.661	0.661	0.801	0.569	0.661	0.665
SMO-RBF- 10^5	66.099	33.901	8.137	0.661	0.661	0.801	0.569	0.661	0.665

Table 3.78: Classification Results for Tr12 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	80.454	19.546	5.977	0.805	0.805	0.817	0.795	0.805	0.806
IB3	76.613	23.387	8.744	0.767	0.767	0.770	0.737	0.767	0.753
IB5	75.988	24.012	8.725	0.760	0.760	0.773	0.713	0.760	0.742
IB7	75.978	24.022	7.355	0.760	0.760	0.748	0.690	0.760	0.718
IB9	74.708	25.292	9.927	0.748	0.748	0.766	0.671	0.748	0.715
IB11	75.665	24.335	8.583	0.757	0.757	0.778	0.682	0.757	0.727
IB13	75.968	24.032	10.769	0.760	0.760	0.784	0.684	0.760	0.731
IB15	74.063	25.938	10.549	0.741	0.741	0.735	0.648	0.741	0.689
IB17	75.675	24.325	9.452	0.757	0.757	0.778	0.665	0.757	0.717
IB19	75.353	24.647	7.920	0.754	0.754	0.772	0.665	0.754	0.714
IB21	74.385	25.615	8.461	0.744	0.744	0.779	0.658	0.744	0.713
IB25	73.448	26.553	9.125	0.735	0.735	0.782	0.637	0.735	0.702
IB29	73.438	26.563	9.346	0.735	0.735	0.767	0.633	0.735	0.693
IB35	73.448	26.553	8.162	0.735	0.735	0.635	0.621	0.735	0.628
IB41	73.115	26.885	9.286	0.732	0.732	0.686	0.613	0.732	0.647
IB49	72.177	27.823	7.913	0.722	0.722	0.636	0.602	0.722	0.618
IB57	70.262	29.738	8.612	0.703	0.703	0.616	0.584	0.703	0.599
IB73	66.421	33.579	9.553	0.665	0.665	0.632	0.540	0.665	0.582
IB89	57.157	42.843	10.758	0.572	0.572	0.624	0.435	0.572	0.513
IB1-WV	80.454	19.546	5.977	0.805	0.805	0.817	0.795	0.805	0.806
IB3-WV	80.474	19.526	6.737	0.805	0.805	0.810	0.787	0.805	0.798
IB5-WV	81.119	18.881	6.838	0.812	0.812	0.825	0.779	0.812	0.801
IB7-WV	81.744	18.256	7.529	0.818	0.818	0.836	0.769	0.818	0.801
IB9-WV	79.183	20.817	9.589	0.792	0.792	0.813	0.736	0.792	0.773
IB11-WV	79.829	20.171	9.983	0.799	0.799	0.828	0.722	0.799	0.771
IB13-WV	79.506	20.494	9.253	0.796	0.796	0.817	0.713	0.796	0.762
IB15-WV	78.851	21.149	9.737	0.789	0.789	0.807	0.710	0.789	0.755
IB17-WV	79.506	20.494	8.430	0.796	0.796	0.815	0.717	0.796	0.763
IB19-WV	78.881	21.119	8.991	0.789	0.789	0.811	0.713	0.789	0.759
IB21-WV	77.601	22.399	9.763	0.776	0.776	0.805	0.700	0.776	0.749
IB25-WV	76.008	23.992	9.230	0.760	0.760	0.817	0.674	0.760	0.738
IB29-WV	75.363	24.637	9.649	0.754	0.754	0.808	0.666	0.754	0.730
IB35-WV	75.363	24.637	8.862	0.754	0.754	0.809	0.654	0.754	0.723
IB41-WV	75.040	24.960	9.642	0.751	0.751	0.707	0.636	0.751	0.669
IB49-WV	74.093	25.907	8.939	0.741	0.741	0.700	0.623	0.741	0.659
IB57-WV	72.480	27.520	8.472	0.725	0.725	0.678	0.606	0.725	0.640
IB73-WV	69.637	30.363	8.920	0.696	0.696	0.753	0.578	0.696	0.654
IB89-WV	64.808	35.192	10.206	0.649	0.649	0.770	0.523	0.649	0.623

3.39 Tr21 Collection

Table 3.79 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.80 presents the classification results obtained by IBk for Tr21 collection.

Table 3.79: Classification Results for Tr21 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
Naïve Bayes	47.950	52.050	9.303	0.479	0.479	0.383	0.630	0.479	0.476
Multinomial NB	61.355	38.645	8.657	0.613	0.613	0.722	0.593	0.613	0.651
J48	81.275	18.726	4.303	0.813	0.813	0.774	0.649	0.813	0.706
SMO-Linear-0	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-Linear- 10^{-5}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-Linear- 10^{-4}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-Linear- 10^{-3}	69.029	30.972	8.073	0.690	0.690	0.282	0.177	0.690	0.217
SMO-Linear- 10^{-2}	69.332	30.669	8.035	0.693	0.693	0.282	0.188	0.693	0.225
SMO-Linear- 10^{-1}	78.610	21.390	8.110	0.786	0.786	0.593	0.412	0.786	0.486
SMO-Linear- 10^0	79.777	20.223	7.753	0.798	0.798	0.680	0.477	0.798	0.560
SMO-Linear- 10^1	79.483	20.517	7.592	0.795	0.795	0.652	0.476	0.795	0.550
SMO-Linear- 10^2	79.483	20.517	7.592	0.795	0.795	0.652	0.476	0.795	0.550
SMO-Linear- 10^3	79.483	20.517	7.592	0.795	0.795	0.652	0.476	0.795	0.550
SMO-Linear- 10^4	79.483	20.517	7.592	0.795	0.795	0.652	0.476	0.795	0.550
SMO-Linear- 10^5	79.483	20.517	7.592	0.795	0.795	0.652	0.476	0.795	0.550
SMO-Poly-0	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-Poly- 10^{-5}	68.734	31.266	8.064	0.688	0.688	0.198	0.176	0.688	0.187
SMO-Poly- 10^{-4}	69.037	30.963	7.595	0.690	0.690	0.226	0.187	0.690	0.205
SMO-Poly- 10^{-3}	67.843	32.157	8.027	0.679	0.679	0.170	0.174	0.679	0.172
SMO-Poly- 10^{-2}	72.638	27.362	8.567	0.726	0.726	0.501	0.294	0.726	0.371
SMO-Poly- 10^{-1}	75.312	24.688	8.367	0.753	0.753	0.554	0.346	0.753	0.426
SMO-Poly- 10^0	75.018	24.982	8.269	0.750	0.750	0.554	0.345	0.750	0.425
SMO-Poly- 10^1	75.018	24.982	8.269	0.750	0.750	0.554	0.345	0.750	0.425
SMO-Poly- 10^2	75.018	24.982	8.269	0.750	0.750	0.554	0.345	0.750	0.425
SMO-Poly- 10^3	75.018	24.982	8.269	0.750	0.750	0.554	0.345	0.750	0.425
SMO-Poly- 10^4	75.018	24.982	8.269	0.750	0.750	0.554	0.345	0.750	0.425
SMO-Poly- 10^5	75.018	24.982	8.269	0.750	0.750	0.554	0.345	0.750	0.425
SMO-RBF-0	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^{-5}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^{-4}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^{-3}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^{-2}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^{-1}	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^0	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
SMO-RBF- 10^1	77.968	22.032	6.027	0.780	0.780	0.607	0.313	0.780	0.413
SMO-RBF- 10^2	79.465	20.535	5.727	0.795	0.795	0.515	0.339	0.795	0.409
SMO-RBF- 10^3	79.465	20.535	5.727	0.795	0.795	0.515	0.339	0.795	0.409
SMO-RBF- 10^4	79.465	20.535	5.727	0.795	0.795	0.515	0.339	0.795	0.409
SMO-RBF- 10^5	79.465	20.535	5.727	0.795	0.795	0.515	0.339	0.795	0.409

Table 3.80: Classification Results for Tr21 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	87.193	12.808	6.284	0.872	0.872	0.807	0.644	0.872	0.717
IB3	88.075	11.925	7.598	0.881	0.881	0.717	0.535	0.881	0.613
IB5	88.369	11.631	7.167	0.884	0.884	0.744	0.521	0.884	0.613
IB7	86.586	13.414	7.585	0.866	0.866	0.760	0.463	0.866	0.575
IB9	85.695	14.305	7.448	0.857	0.857	0.638	0.424	0.857	0.509
IB11	85.686	14.314	7.227	0.857	0.857	0.638	0.424	0.857	0.509
IB13	85.107	14.893	7.107	0.851	0.851	0.470	0.409	0.851	0.438
IB15	84.804	15.196	6.915	0.848	0.848	0.465	0.404	0.848	0.432
IB17	84.804	15.196	6.915	0.848	0.848	0.465	0.404	0.848	0.432
IB19	84.216	15.784	6.974	0.842	0.842	0.463	0.396	0.842	0.427
IB21	84.216	15.784	6.974	0.842	0.842	0.463	0.396	0.842	0.427
IB25	83.619	16.382	6.622	0.836	0.836	0.462	0.387	0.836	0.421
IB29	83.324	16.676	6.736	0.833	0.833	0.467	0.383	0.833	0.421
IB35	81.542	18.458	6.293	0.815	0.815	0.465	0.355	0.815	0.403
IB41	79.144	20.856	7.163	0.792	0.792	0.461	0.318	0.792	0.376
IB49	76.462	23.538	7.923	0.765	0.765	0.458	0.279	0.765	0.347
IB57	72.005	27.995	8.046	0.720	0.720	0.452	0.214	0.720	0.291
IB73	69.617	30.383	8.270	0.696	0.696	0.449	0.180	0.696	0.256
IB89	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136
IB1-WV	87.193	12.808	6.284	0.872	0.872	0.807	0.644	0.872	0.717
IB3-WV	87.772	12.228	7.292	0.878	0.878	0.869	0.575	0.878	0.692
IB5-WV	88.663	11.337	7.335	0.887	0.887	0.911	0.563	0.887	0.696
IB7-WV	88.066	11.934	7.117	0.881	0.881	0.780	0.503	0.881	0.612
IB9-WV	86.578	13.423	7.522	0.866	0.866	0.801	0.457	0.866	0.582
IB11-WV	85.989	14.011	7.249	0.860	0.860	0.633	0.428	0.860	0.511
IB13-WV	86.292	13.708	7.384	0.863	0.863	0.634	0.432	0.863	0.514
IB15-WV	85.998	14.002	7.591	0.860	0.860	0.467	0.422	0.860	0.443
IB17-WV	86.007	13.993	6.689	0.860	0.860	0.467	0.423	0.860	0.444
IB19-WV	85.419	14.581	6.854	0.854	0.854	0.471	0.414	0.854	0.440
IB21-WV	84.519	15.481	7.060	0.845	0.845	0.464	0.400	0.845	0.430
IB25-WV	84.822	15.178	6.456	0.848	0.848	0.464	0.405	0.848	0.433
IB29-WV	84.519	15.481	7.060	0.845	0.845	0.464	0.400	0.845	0.430
IB35-WV	82.433	17.567	6.640	0.824	0.824	0.460	0.370	0.824	0.410
IB41-WV	81.836	18.164	6.935	0.818	0.818	0.459	0.359	0.818	0.403
IB49-WV	78.244	21.756	7.666	0.783	0.783	0.452	0.307	0.783	0.365
IB57-WV	74.367	25.633	7.800	0.744	0.744	0.455	0.248	0.744	0.321
IB73-WV	70.517	29.483	7.377	0.705	0.705	0.450	0.193	0.705	0.270
IB89-WV	68.725	31.275	7.869	0.688	0.688	0.115	0.167	0.688	0.136

3.40 Tr23 Collection

Table 3.81 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.82 presents the classification results obtained by IBk for Tr23 collection.

Table 3.81: Classification Results for Tr23 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	56.857	43.143	10.117	0.569	0.569	0.549	0.733	0.569	0.628
Multinomial NB	70.619	29.381	9.671	0.706	0.706	0.793	0.808	0.706	0.800
J48	93.190	6.810	6.131	0.931	0.931	0.930	0.917	0.931	0.923
SMO-Linear-0	17.667	82.333	7.967	0.176	0.176	0.029	0.167	0.176	0.050
SMO-Linear- 10^{-5}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-Linear- 10^{-4}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-Linear- 10^{-3}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-Linear- 10^{-2}	43.690	56.310	7.907	0.436	0.436	0.240	0.167	0.436	0.197
SMO-Linear- 10^{-1}	70.619	29.381	9.550	0.706	0.706	0.745	0.508	0.706	0.604
SMO-Linear- 10^0	72.524	27.476	10.733	0.725	0.725	0.731	0.552	0.725	0.629
SMO-Linear- 10^1	72.524	27.476	10.733	0.725	0.725	0.731	0.552	0.725	0.629
SMO-Linear- 10^2	72.524	27.476	10.733	0.725	0.725	0.731	0.552	0.725	0.629
SMO-Linear- 10^3	72.524	27.476	10.733	0.725	0.725	0.731	0.552	0.725	0.629
SMO-Linear- 10^4	72.524	27.476	10.733	0.725	0.725	0.731	0.552	0.725	0.629
SMO-Linear- 10^5	72.524	27.476	10.733	0.725	0.725	0.731	0.552	0.725	0.629
SMO-Poly-0	17.667	82.333	7.967	0.176	0.176	0.029	0.167	0.176	0.050
SMO-Poly- 10^{-5}	44.190	55.810	6.481	0.441	0.441	0.074	0.165	0.441	0.102
SMO-Poly- 10^{-4}	42.714	57.286	7.161	0.426	0.426	0.073	0.159	0.426	0.100
SMO-Poly- 10^{-3}	44.143	55.857	8.535	0.441	0.441	0.407	0.180	0.441	0.249
SMO-Poly- 10^{-2}	57.357	42.643	11.037	0.574	0.574	0.528	0.336	0.574	0.411
SMO-Poly- 10^{-1}	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-Poly- 10^0	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-Poly- 10^1	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-Poly- 10^2	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-Poly- 10^3	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-Poly- 10^4	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-Poly- 10^5	63.738	36.262	9.971	0.637	0.637	0.595	0.428	0.637	0.498
SMO-RBF-0	17.667	82.333	7.967	0.176	0.176	0.029	0.167	0.176	0.050
SMO-RBF- 10^{-5}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-RBF- 10^{-4}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-RBF- 10^{-3}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-RBF- 10^{-2}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-RBF- 10^{-1}	44.667	55.333	5.401	0.446	0.446	0.074	0.167	0.446	0.103
SMO-RBF- 10^0	49.095	50.905	5.881	0.490	0.490	0.411	0.204	0.490	0.272
SMO-RBF- 10^1	70.595	29.405	5.658	0.706	0.706	0.711	0.438	0.706	0.542
SMO-RBF- 10^2	72.548	27.452	6.567	0.725	0.725	0.716	0.469	0.725	0.566
SMO-RBF- 10^3	72.548	27.452	6.567	0.725	0.725	0.716	0.469	0.725	0.566
SMO-RBF- 10^4	72.548	27.452	6.567	0.725	0.725	0.716	0.469	0.725	0.566
SMO-RBF- 10^5	72.548	27.452	6.567	0.725	0.725	0.716	0.469	0.725	0.566

Table 3.82: Classification Results for Tr23 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	84.333	15.667	8.695	0.843	0.843	0.847	0.740	0.843	0.790
IB3	81.952	18.048	10.204	0.819	0.819	0.880	0.700	0.819	0.780
IB5	80.929	19.072	8.664	0.809	0.809	0.861	0.677	0.809	0.758
IB7	78.024	21.976	9.683	0.779	0.779	0.829	0.585	0.779	0.686
IB9	76.071	23.929	10.130	0.760	0.760	0.837	0.515	0.760	0.637
IB11	73.643	26.357	9.553	0.735	0.735	0.882	0.476	0.735	0.619
IB13	72.143	27.857	9.400	0.721	0.721	0.881	0.440	0.721	0.587
IB15	71.643	28.357	7.751	0.716	0.716	0.720	0.412	0.716	0.524
IB17	69.690	30.310	7.383	0.696	0.696	0.533	0.384	0.696	0.446
IB19	68.690	31.310	7.930	0.686	0.686	0.528	0.377	0.686	0.440
IB21	67.214	32.786	7.892	0.672	0.672	0.363	0.353	0.672	0.358
IB25	67.690	32.310	6.340	0.676	0.676	0.382	0.356	0.676	0.368
IB29	69.190	30.810	7.237	0.691	0.691	0.387	0.370	0.691	0.378
IB35	70.190	29.810	7.966	0.701	0.701	0.405	0.374	0.701	0.389
IB41	69.667	30.333	8.186	0.696	0.696	0.402	0.369	0.696	0.385
IB49	68.714	31.286	7.862	0.686	0.686	0.404	0.361	0.686	0.381
IB57	66.738	33.262	7.885	0.667	0.667	0.399	0.345	0.667	0.370
IB73	59.405	40.595	6.319	0.593	0.593	0.408	0.284	0.593	0.335
IB89	53.000	47.000	4.726	0.529	0.529	0.404	0.231	0.529	0.294
IB1-WV	84.333	15.667	8.695	0.843	0.843	0.847	0.740	0.843	0.790
IB3-WV	82.905	17.095	8.868	0.828	0.828	0.889	0.719	0.828	0.795
IB5-WV	80.929	19.072	8.664	0.809	0.809	0.846	0.677	0.809	0.752
IB7-WV	79.500	20.500	9.131	0.794	0.794	0.848	0.604	0.794	0.706
IB9-WV	78.500	21.500	7.328	0.784	0.784	0.850	0.606	0.784	0.708
IB11-WV	77.024	22.976	7.848	0.770	0.770	0.901	0.536	0.770	0.672
IB13-WV	75.048	24.952	7.089	0.750	0.750	0.895	0.511	0.750	0.651
IB15-WV	74.095	25.905	8.081	0.740	0.740	0.896	0.480	0.740	0.625
IB17-WV	73.119	26.881	7.498	0.730	0.730	0.894	0.448	0.730	0.597
IB19-WV	72.119	27.881	7.206	0.721	0.721	0.724	0.415	0.721	0.528
IB21-WV	71.167	28.833	7.895	0.711	0.711	0.561	0.400	0.711	0.467
IB25-WV	70.643	29.357	5.363	0.706	0.706	0.395	0.383	0.706	0.389
IB29-WV	71.119	28.881	5.617	0.711	0.711	0.400	0.388	0.711	0.394
IB35-WV	69.167	30.833	6.239	0.691	0.691	0.390	0.370	0.691	0.380
IB41-WV	67.667	32.333	6.832	0.676	0.676	0.364	0.356	0.676	0.360
IB49-WV	69.167	30.833	6.239	0.691	0.691	0.381	0.369	0.691	0.374
IB57-WV	67.190	32.810	5.966	0.672	0.672	0.383	0.352	0.672	0.367
IB73-WV	64.738	35.262	3.888	0.647	0.647	0.385	0.331	0.647	0.356
IB89-WV	59.905	40.095	6.146	0.598	0.598	0.386	0.291	0.598	0.331

3.41 Tr31 Collection

Table 3.83 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.84 presents the classification results obtained by IBk for Tr31 collection.

Table 3.83: Classification Results for Tr31 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
Naïve Bayes	79.725	20.275	3.963	0.797	0.797	0.648	0.674	0.797	0.661
Multinomial NB	94.390	5.610	1.348	0.944	0.944	0.781	0.791	0.944	0.786
J48	93.103	6.897	2.721	0.931	0.931	0.767	0.762	0.931	0.765
SMO-Linear-0	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-Linear- 10^{-5}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-Linear- 10^{-4}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-Linear- 10^{-3}	38.842	61.158	5.548	0.388	0.388	0.340	0.154	0.388	0.212
SMO-Linear- 10^{-2}	67.319	32.681	4.245	0.673	0.673	0.534	0.362	0.673	0.431
SMO-Linear- 10^{-1}	88.022	11.978	2.543	0.880	0.880	0.765	0.682	0.880	0.721
SMO-Linear- 10^0	90.614	9.386	3.087	0.906	0.906	0.771	0.721	0.906	0.745
SMO-Linear- 10^1	90.725	9.275	3.823	0.907	0.907	0.773	0.722	0.907	0.746
SMO-Linear- 10^2	90.725	9.275	3.823	0.907	0.907	0.773	0.722	0.907	0.746
SMO-Linear- 10^3	90.725	9.275	3.823	0.907	0.907	0.773	0.722	0.907	0.746
SMO-Linear- 10^4	90.725	9.275	3.823	0.907	0.907	0.773	0.722	0.907	0.746
SMO-Linear- 10^5	90.725	9.275	3.823	0.907	0.907	0.773	0.722	0.907	0.746
SMO-Poly-0	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-Poly- 10^{-5}	38.627	61.373	5.570	0.386	0.386	0.324	0.153	0.386	0.208
SMO-Poly- 10^{-4}	38.627	61.373	5.901	0.386	0.386	0.416	0.156	0.386	0.227
SMO-Poly- 10^{-3}	59.551	40.449	2.964	0.595	0.595	0.491	0.298	0.595	0.371
SMO-Poly- 10^{-2}	79.183	20.817	3.069	0.792	0.792	0.714	0.545	0.792	0.618
SMO-Poly- 10^{-1}	81.452	18.548	3.432	0.814	0.814	0.708	0.608	0.814	0.654
SMO-Poly- 10^0	81.777	18.224	2.946	0.818	0.818	0.705	0.609	0.818	0.653
SMO-Poly- 10^1	81.777	18.224	2.946	0.818	0.818	0.705	0.609	0.818	0.653
SMO-Poly- 10^2	81.777	18.224	2.946	0.818	0.818	0.705	0.609	0.818	0.653
SMO-Poly- 10^3	81.777	18.224	2.946	0.818	0.818	0.705	0.609	0.818	0.653
SMO-Poly- 10^4	81.777	18.224	2.946	0.818	0.818	0.705	0.609	0.818	0.653
SMO-Poly- 10^5	81.777	18.224	2.946	0.818	0.818	0.705	0.609	0.818	0.653
SMO-RBF-0	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-RBF- 10^{-5}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-RBF- 10^{-4}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-RBF- 10^{-3}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-RBF- 10^{-2}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-RBF- 10^{-1}	37.978	62.022	5.366	0.380	0.380	0.054	0.143	0.380	0.079
SMO-RBF- 10^0	67.101	32.899	4.577	0.671	0.671	0.568	0.362	0.671	0.442
SMO-RBF- 10^1	85.436	14.564	2.979	0.854	0.854	0.776	0.648	0.854	0.706
SMO-RBF- 10^2	87.270	12.730	2.844	0.873	0.873	0.781	0.675	0.873	0.724
SMO-RBF- 10^3	87.055	12.945	2.549	0.871	0.871	0.777	0.674	0.871	0.722
SMO-RBF- 10^4	87.055	12.945	2.549	0.871	0.871	0.777	0.674	0.871	0.722
SMO-RBF- 10^5	87.055	12.945	2.549	0.871	0.871	0.777	0.674	0.871	0.722

Table 3.84: Classification Results for Tr31 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	94.606	5.394	1.600	0.946	0.946	0.803	0.794	0.946	0.798
IB3	92.338	7.663	2.671	0.923	0.923	0.780	0.754	0.923	0.767
IB5	91.907	8.093	1.697	0.919	0.919	0.786	0.743	0.919	0.764
IB7	91.477	8.523	2.127	0.915	0.915	0.783	0.740	0.915	0.761
IB9	91.047	8.953	1.799	0.910	0.910	0.790	0.731	0.910	0.759
IB11	90.615	9.385	1.809	0.906	0.906	0.792	0.724	0.906	0.756
IB13	89.970	10.030	2.149	0.900	0.900	0.791	0.715	0.900	0.751
IB15	89.326	10.674	2.208	0.893	0.893	0.786	0.709	0.893	0.745
IB17	89.327	10.673	2.150	0.893	0.893	0.790	0.712	0.893	0.749
IB19	88.898	11.102	2.738	0.889	0.889	0.775	0.706	0.889	0.738
IB21	89.327	10.673	2.206	0.893	0.893	0.790	0.705	0.893	0.745
IB25	88.894	11.106	2.025	0.889	0.889	0.779	0.696	0.889	0.735
IB29	88.464	11.536	2.336	0.885	0.885	0.781	0.685	0.885	0.730
IB35	87.599	12.401	2.143	0.876	0.876	0.778	0.672	0.876	0.721
IB41	87.599	12.401	2.671	0.876	0.876	0.784	0.659	0.876	0.716
IB49	86.414	13.586	2.660	0.864	0.864	0.778	0.605	0.864	0.681
IB57	85.548	14.452	2.801	0.855	0.855	0.774	0.568	0.855	0.655
IB73	84.254	15.746	3.118	0.843	0.843	0.624	0.539	0.843	0.579
IB89	83.069	16.931	3.249	0.831	0.831	0.621	0.522	0.831	0.568
IB1-WV	94.606	5.394	1.600	0.946	0.946	0.803	0.794	0.946	0.798
IB3-WV	93.416	6.584	2.392	0.934	0.934	0.796	0.776	0.934	0.786
IB5-WV	92.661	7.339	1.867	0.927	0.927	0.791	0.749	0.927	0.770
IB7-WV	92.339	7.661	2.289	0.923	0.923	0.792	0.745	0.923	0.768
IB9-WV	91.909	8.091	2.603	0.919	0.919	0.799	0.741	0.919	0.769
IB11-WV	91.369	8.631	1.938	0.914	0.914	0.797	0.732	0.914	0.763
IB13-WV	91.046	8.954	1.936	0.910	0.910	0.795	0.729	0.910	0.761
IB15-WV	90.941	9.059	2.103	0.909	0.909	0.800	0.727	0.909	0.762
IB17-WV	90.404	9.596	2.411	0.904	0.904	0.796	0.723	0.904	0.758
IB19-WV	89.974	10.026	2.387	0.900	0.900	0.782	0.719	0.900	0.750
IB21-WV	90.082	9.918	2.518	0.901	0.901	0.785	0.716	0.901	0.749
IB25-WV	89.647	10.353	2.045	0.896	0.896	0.775	0.707	0.896	0.739
IB29-WV	89.327	10.673	2.210	0.893	0.893	0.786	0.698	0.893	0.739
IB35-WV	88.784	11.216	2.257	0.888	0.888	0.788	0.692	0.888	0.737
IB41-WV	88.353	11.647	2.336	0.883	0.883	0.786	0.676	0.883	0.727
IB49-WV	88.353	11.647	2.745	0.883	0.883	0.787	0.662	0.883	0.719
IB57-WV	87.382	12.618	2.443	0.874	0.874	0.780	0.616	0.874	0.689
IB73-WV	85.658	14.342	2.705	0.857	0.857	0.774	0.564	0.857	0.652
IB89-WV	84.258	15.742	3.211	0.843	0.843	0.626	0.540	0.843	0.580

3.42 Tr41 Collection

Table 3.85 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.86 presents the classification results obtained by IBk for Tr41 collection.

Table 3.85: Classification Results for Tr41 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	84.956	15.045	4.639	0.850	0.850	0.802	0.796	0.850	0.799
Multinomial NB	94.523	5.477	4.198	0.945	0.945	0.932	0.876	0.945	0.903
J48	90.773	9.227	3.777	0.908	0.908	0.842	0.833	0.908	0.838
SMO-Linear-0	2.962	97.038	1.546	0.030	0.030	0.003	0.100	0.030	0.006
SMO-Linear- 10^{-5}	27.665	72.336	5.809	0.277	0.277	0.028	0.100	0.277	0.043
SMO-Linear- 10^{-4}	27.665	72.336	5.809	0.277	0.277	0.028	0.100	0.277	0.043
SMO-Linear- 10^{-3}	30.394	69.606	6.385	0.304	0.304	0.228	0.114	0.304	0.152
SMO-Linear- 10^{-2}	61.834	38.166	6.593	0.618	0.618	0.633	0.342	0.618	0.444
SMO-Linear- 10^{-1}	86.437	13.563	3.998	0.864	0.864	0.917	0.731	0.864	0.814
SMO-Linear- 10^0	87.804	12.196	4.411	0.878	0.878	0.896	0.779	0.878	0.834
SMO-Linear- 10^1	87.803	12.197	4.112	0.878	0.878	0.859	0.793	0.878	0.825
SMO-Linear- 10^2	87.803	12.197	4.112	0.878	0.878	0.859	0.793	0.878	0.825
SMO-Linear- 10^3	87.803	12.197	4.112	0.878	0.878	0.859	0.793	0.878	0.825
SMO-Linear- 10^4	87.803	12.197	4.112	0.878	0.878	0.859	0.793	0.878	0.825
SMO-Linear- 10^5	87.803	12.197	4.112	0.878	0.878	0.859	0.793	0.878	0.825
SMO-Poly-0	2.962	97.038	1.546	0.030	0.030	0.003	0.100	0.030	0.006
SMO-Poly- 10^{-5}	27.778	72.222	5.811	0.278	0.278	0.128	0.101	0.278	0.112
SMO-Poly- 10^{-4}	29.718	70.282	4.792	0.297	0.297	0.262	0.114	0.297	0.159
SMO-Poly- 10^{-3}	52.716	47.285	7.386	0.527	0.527	0.710	0.299	0.527	0.421
SMO-Poly- 10^{-2}	74.813	25.187	5.294	0.748	0.748	0.858	0.564	0.748	0.680
SMO-Poly- 10^{-1}	79.943	20.058	4.955	0.800	0.800	0.867	0.643	0.800	0.738
SMO-Poly- 10^0	80.854	19.146	4.458	0.809	0.809	0.847	0.664	0.809	0.744
SMO-Poly- 10^1	80.969	19.031	4.250	0.810	0.810	0.810	0.666	0.810	0.731
SMO-Poly- 10^2	80.969	19.031	4.250	0.810	0.810	0.810	0.666	0.810	0.731
SMO-Poly- 10^3	80.969	19.031	4.250	0.810	0.810	0.810	0.666	0.810	0.731
SMO-Poly- 10^4	80.969	19.031	4.250	0.810	0.810	0.810	0.666	0.810	0.731
SMO-Poly- 10^5	80.969	19.031	4.250	0.810	0.810	0.810	0.666	0.810	0.731
SMO-RBF-0	2.962	97.038	1.546	0.030	0.030	0.003	0.100	0.030	0.006
SMO-RBF- 10^{-5}	27.665	72.336	5.809	0.277	0.277	0.028	0.100	0.277	0.043
SMO-RBF- 10^{-4}	27.665	72.336	5.809	0.277	0.277	0.028	0.100	0.277	0.043
SMO-RBF- 10^{-3}	27.665	72.336	5.809	0.277	0.277	0.028	0.100	0.277	0.043
SMO-RBF- 10^{-2}	27.665	72.336	5.809	0.277	0.277	0.028	0.100	0.277	0.043
SMO-RBF- 10^{-1}	34.956	65.045	6.365	0.350	0.350	0.127	0.137	0.350	0.132
SMO-RBF- 10^0	65.136	34.864	6.659	0.651	0.651	0.600	0.345	0.651	0.438
SMO-RBF- 10^1	83.593	16.407	4.622	0.836	0.836	0.928	0.657	0.836	0.769
SMO-RBF- 10^2	83.937	16.063	4.586	0.839	0.839	0.917	0.678	0.839	0.780
SMO-RBF- 10^3	83.937	16.063	4.472	0.839	0.839	0.917	0.678	0.839	0.780
SMO-RBF- 10^4	83.937	16.063	4.472	0.839	0.839	0.917	0.678	0.839	0.780
SMO-RBF- 10^5	83.937	16.063	4.472	0.839	0.839	0.917	0.678	0.839	0.780

Table 3.86: Classification Results for Tr41 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	92.026	7.974	3.095	0.920	0.920	0.857	0.868	0.920	0.862
IB3	92.701	7.299	3.851	0.927	0.927	0.926	0.856	0.927	0.890
IB5	91.906	8.094	3.793	0.919	0.919	0.922	0.842	0.919	0.880
IB7	92.019	7.981	4.238	0.920	0.920	0.929	0.832	0.920	0.878
IB9	91.903	8.097	4.253	0.919	0.919	0.828	0.825	0.919	0.826
IB11	91.220	8.780	4.623	0.912	0.912	0.823	0.813	0.912	0.818
IB13	91.789	8.211	4.695	0.918	0.918	0.829	0.819	0.918	0.824
IB15	91.789	8.211	4.411	0.918	0.918	0.835	0.817	0.918	0.826
IB17	91.903	8.097	4.603	0.919	0.919	0.834	0.816	0.919	0.825
IB19	91.334	8.667	4.739	0.913	0.913	0.830	0.806	0.913	0.818
IB21	91.109	8.891	5.144	0.911	0.911	0.826	0.807	0.911	0.817
IB25	90.422	9.578	5.308	0.904	0.904	0.826	0.786	0.904	0.805
IB29	90.311	9.689	5.158	0.903	0.903	0.825	0.769	0.903	0.796
IB35	89.171	10.830	5.484	0.892	0.892	0.827	0.733	0.892	0.777
IB41	88.034	11.966	4.702	0.880	0.880	0.824	0.683	0.880	0.747
IB49	85.866	14.134	5.420	0.859	0.859	0.803	0.643	0.859	0.714
IB57	83.476	16.524	5.510	0.835	0.835	0.677	0.598	0.835	0.635
IB73	80.632	19.368	6.255	0.806	0.806	0.666	0.543	0.806	0.598
IB89	76.415	23.586	6.538	0.764	0.764	0.603	0.463	0.764	0.523
IB1-WV	92.026	7.974	3.095	0.920	0.920	0.857	0.868	0.920	0.862
IB3-WV	92.815	7.185	4.043	0.928	0.928	0.867	0.858	0.928	0.862
IB5-WV	93.043	6.957	4.055	0.931	0.931	0.929	0.859	0.931	0.893
IB7-WV	92.478	7.522	3.946	0.925	0.925	0.927	0.851	0.925	0.887
IB9-WV	91.903	8.097	4.313	0.919	0.919	0.826	0.827	0.919	0.826
IB11-WV	91.676	8.324	4.562	0.917	0.917	0.822	0.825	0.917	0.823
IB13-WV	91.904	8.096	4.058	0.919	0.919	0.830	0.824	0.919	0.827
IB15-WV	92.017	7.983	4.240	0.920	0.920	0.830	0.822	0.920	0.826
IB17-WV	92.017	7.983	4.448	0.920	0.920	0.831	0.817	0.920	0.824
IB19-WV	91.903	8.097	4.822	0.919	0.919	0.831	0.818	0.919	0.824
IB21-WV	91.791	8.209	5.020	0.918	0.918	0.829	0.814	0.918	0.821
IB25-WV	90.650	9.350	5.125	0.907	0.907	0.825	0.793	0.907	0.809
IB29-WV	90.765	9.235	5.191	0.908	0.908	0.827	0.789	0.908	0.808
IB35-WV	90.311	9.689	5.108	0.903	0.903	0.827	0.776	0.903	0.801
IB41-WV	89.061	10.939	4.690	0.891	0.891	0.829	0.722	0.891	0.771
IB49-WV	87.692	12.308	5.673	0.877	0.877	0.818	0.680	0.877	0.742
IB57-WV	85.866	14.134	5.833	0.859	0.859	0.794	0.634	0.859	0.705
IB73-WV	83.476	16.524	5.933	0.835	0.835	0.788	0.594	0.835	0.677
IB89-WV	79.492	20.508	6.953	0.795	0.795	0.607	0.518	0.795	0.559

3.43 Tr45 Collection

Table 3.87 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.88 presents the classification results obtained by IBk for Tr45 collection.

Table 3.87: Classification Results for Tr45 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	66.667	33.333	4.807	0.667	0.667	0.695	0.701	0.667	0.698
Multinomial NB	82.464	17.536	5.401	0.825	0.825	0.821	0.797	0.825	0.809
J48	90.290	9.710	3.782	0.903	0.903	0.851	0.844	0.903	0.848
SMO-Linear-0	18.551	81.449	4.289	0.186	0.186	0.019	0.100	0.186	0.031
SMO-Linear- 10^{-5}	23.188	76.812	5.345	0.232	0.232	0.073	0.101	0.232	0.085
SMO-Linear- 10^{-4}	23.188	76.812	5.345	0.232	0.232	0.073	0.101	0.232	0.085
SMO-Linear- 10^{-3}	23.188	76.812	5.265	0.232	0.232	0.073	0.101	0.232	0.085
SMO-Linear- 10^{-2}	45.797	54.203	5.840	0.458	0.458	0.568	0.291	0.458	0.385
SMO-Linear- 10^{-1}	79.275	20.725	4.852	0.793	0.793	0.774	0.711	0.793	0.741
SMO-Linear- 10^0	81.159	18.841	4.099	0.812	0.812	0.786	0.751	0.812	0.768
SMO-Linear- 10^1	81.304	18.696	4.122	0.813	0.813	0.785	0.754	0.813	0.769
SMO-Linear- 10^2	81.304	18.696	4.122	0.813	0.813	0.785	0.754	0.813	0.769
SMO-Linear- 10^3	81.304	18.696	4.122	0.813	0.813	0.785	0.754	0.813	0.769
SMO-Linear- 10^4	81.304	18.696	4.122	0.813	0.813	0.785	0.754	0.813	0.769
SMO-Linear- 10^5	81.304	18.696	4.122	0.813	0.813	0.785	0.754	0.813	0.769
SMO-Poly-0	18.551	81.449	4.289	0.186	0.186	0.019	0.100	0.186	0.031
SMO-Poly- 10^{-5}	23.478	76.522	5.415	0.235	0.235	0.140	0.103	0.235	0.119
SMO-Poly- 10^{-4}	21.739	78.261	4.936	0.217	0.217	0.202	0.097	0.217	0.131
SMO-Poly- 10^{-3}	41.884	58.116	5.592	0.419	0.419	0.583	0.247	0.419	0.347
SMO-Poly- 10^{-2}	65.072	34.928	3.635	0.651	0.651	0.676	0.539	0.651	0.600
SMO-Poly- 10^{-1}	68.406	31.594	4.190	0.684	0.684	0.689	0.598	0.684	0.640
SMO-Poly- 10^0	68.696	31.304	4.160	0.687	0.687	0.691	0.600	0.687	0.642
SMO-Poly- 10^1	68.696	31.304	4.160	0.687	0.687	0.691	0.600	0.687	0.642
SMO-Poly- 10^2	68.696	31.304	4.160	0.687	0.687	0.691	0.600	0.687	0.642
SMO-Poly- 10^3	68.696	31.304	4.160	0.687	0.687	0.691	0.600	0.687	0.642
SMO-Poly- 10^4	68.696	31.304	4.160	0.687	0.687	0.691	0.600	0.687	0.642
SMO-Poly- 10^5	68.696	31.304	4.160	0.687	0.687	0.691	0.600	0.687	0.642
SMO-RBF-0	18.551	81.449	4.289	0.186	0.186	0.019	0.100	0.186	0.031
SMO-RBF- 10^{-5}	23.188	76.812	5.345	0.232	0.232	0.023	0.100	0.232	0.038
SMO-RBF- 10^{-4}	23.188	76.812	5.345	0.232	0.232	0.023	0.100	0.232	0.038
SMO-RBF- 10^{-3}	23.188	76.812	5.345	0.232	0.232	0.023	0.100	0.232	0.038
SMO-RBF- 10^{-2}	23.188	76.812	5.345	0.232	0.232	0.023	0.100	0.232	0.038
SMO-RBF- 10^{-1}	23.188	76.812	5.345	0.232	0.232	0.023	0.100	0.232	0.038
SMO-RBF- 10^0	44.638	55.362	3.985	0.446	0.446	0.526	0.278	0.446	0.364
SMO-RBF- 10^1	76.087	23.913	4.061	0.761	0.761	0.755	0.668	0.761	0.709
SMO-RBF- 10^2	77.971	22.029	5.680	0.780	0.780	0.869	0.692	0.780	0.770
SMO-RBF- 10^3	77.971	22.029	5.680	0.780	0.780	0.869	0.692	0.780	0.770
SMO-RBF- 10^4	77.971	22.029	5.680	0.780	0.780	0.869	0.692	0.780	0.770
SMO-RBF- 10^5	77.971	22.029	5.680	0.780	0.780	0.869	0.692	0.780	0.770

Table 3.88: Classification Results for Tr45 collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	88.841	11.160	3.612	0.888	0.888	0.866	0.822	0.888	0.843
IB3	86.957	13.044	5.423	0.870	0.870	0.853	0.801	0.870	0.826
IB5	87.391	12.609	3.669	0.874	0.874	0.879	0.817	0.874	0.847
IB7	86.522	13.478	3.244	0.865	0.865	0.880	0.797	0.865	0.837
IB9	85.362	14.638	3.692	0.854	0.854	0.862	0.779	0.854	0.818
IB11	84.203	15.797	4.321	0.842	0.842	0.854	0.757	0.842	0.802
IB13	83.623	16.377	4.398	0.836	0.836	0.842	0.743	0.836	0.789
IB15	83.043	16.957	4.765	0.830	0.830	0.851	0.740	0.830	0.791
IB17	82.464	17.536	5.081	0.825	0.825	0.852	0.727	0.825	0.785
IB19	82.029	17.971	5.029	0.820	0.820	0.748	0.714	0.820	0.730
IB21	81.739	18.261	4.109	0.817	0.817	0.747	0.709	0.817	0.727
IB25	80.870	19.131	4.037	0.809	0.809	0.744	0.696	0.809	0.719
IB29	80.580	19.420	3.790	0.806	0.806	0.844	0.691	0.806	0.760
IB35	79.275	20.725	3.493	0.793	0.793	0.742	0.665	0.793	0.702
IB41	79.855	20.145	4.370	0.799	0.799	0.762	0.676	0.799	0.716
IB49	78.986	21.015	5.072	0.790	0.790	0.763	0.663	0.790	0.709
IB57	78.261	21.739	4.099	0.783	0.783	0.773	0.636	0.783	0.697
IB73	75.942	24.058	4.500	0.759	0.759	0.770	0.605	0.759	0.677
IB89	72.464	27.536	5.225	0.725	0.725	0.663	0.554	0.725	0.603
IB1-WV	88.841	11.160	3.612	0.888	0.888	0.866	0.822	0.888	0.843
IB3-WV	88.406	11.594	4.978	0.884	0.884	0.889	0.820	0.884	0.853
IB5-WV	87.971	12.029	4.153	0.880	0.880	0.883	0.824	0.880	0.852
IB7-WV	87.391	12.609	4.050	0.874	0.874	0.882	0.813	0.874	0.846
IB9-WV	86.812	13.189	4.511	0.868	0.868	0.875	0.802	0.868	0.837
IB11-WV	85.942	14.058	5.105	0.859	0.859	0.877	0.790	0.859	0.831
IB13-WV	85.797	14.203	4.574	0.858	0.858	0.873	0.784	0.858	0.826
IB15-WV	84.928	15.073	4.405	0.849	0.849	0.874	0.771	0.849	0.819
IB17-WV	83.913	16.087	4.173	0.839	0.839	0.856	0.752	0.839	0.801
IB19-WV	83.768	16.232	4.481	0.838	0.838	0.859	0.751	0.838	0.801
IB21-WV	83.478	16.522	4.260	0.835	0.835	0.860	0.746	0.835	0.799
IB25-WV	82.609	17.391	4.348	0.826	0.826	0.858	0.729	0.826	0.788
IB29-WV	81.594	18.406	4.203	0.816	0.816	0.856	0.708	0.816	0.775
IB35-WV	81.159	18.841	4.537	0.812	0.812	0.757	0.704	0.812	0.729
IB41-WV	80.580	19.420	3.790	0.806	0.806	0.764	0.685	0.806	0.722
IB49-WV	80.000	20.000	4.386	0.800	0.800	0.766	0.682	0.800	0.721
IB57-WV	79.855	20.145	4.071	0.799	0.799	0.767	0.674	0.799	0.717
IB73-WV	79.130	20.870	4.058	0.791	0.791	0.778	0.661	0.791	0.715
IB89-WV	76.812	23.189	5.144	0.768	0.768	0.769	0.611	0.768	0.681

3.44 WAP Collection

Table 3.89 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.90 presents the classification results obtained by IBk for Wap collection.

Table 3.89: Classification Results for WAP collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
Naïve Bayes	71.731	28.269	2.830	0.717	0.717	0.606	0.566	0.717	0.585
Multinomial NB	81.026	18.974	2.615	0.810	0.810	0.658	0.566	0.810	0.609
J48	67.051	32.949	4.156	0.671	0.671	0.572	0.546	0.671	0.559
SMO-Linear-0	0.962	99.039	0.717	0.010	0.010	0.000	0.050	0.010	0.001
SMO-Linear- 10^{-5}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-Linear- 10^{-4}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-Linear- 10^{-3}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-Linear- 10^{-2}	55.192	44.808	5.495	0.552	0.552	0.533	0.294	0.552	0.379
SMO-Linear- 10^{-1}	81.795	18.205	3.406	0.818	0.818	0.806	0.605	0.818	0.691
SMO-Linear- 10^0	81.859	18.141	3.825	0.819	0.819	0.778	0.623	0.819	0.692
SMO-Linear- 10^1	81.859	18.141	3.825	0.819	0.819	0.778	0.623	0.819	0.692
SMO-Linear- 10^2	81.859	18.141	3.825	0.819	0.819	0.778	0.623	0.819	0.692
SMO-Linear- 10^3	81.859	18.141	3.825	0.819	0.819	0.778	0.623	0.819	0.692
SMO-Linear- 10^4	81.859	18.141	3.825	0.819	0.819	0.778	0.623	0.819	0.692
SMO-Linear- 10^5	81.859	18.141	3.825	0.819	0.819	0.778	0.623	0.819	0.692
SMO-Poly-0	0.962	99.039	0.717	0.010	0.010	0.000	0.050	0.010	0.001
SMO-Poly- 10^{-5}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-Poly- 10^{-4}	23.333	76.667	2.982	0.233	0.233	0.161	0.065	0.233	0.093
SMO-Poly- 10^{-3}	33.013	66.987	3.572	0.330	0.330	0.450	0.146	0.330	0.220
SMO-Poly- 10^{-2}	52.756	47.244	5.145	0.528	0.528	0.706	0.313	0.528	0.433
SMO-Poly- 10^{-1}	56.987	43.013	3.904	0.570	0.570	0.662	0.360	0.570	0.466
SMO-Poly- 10^0	57.244	42.757	3.942	0.572	0.572	0.665	0.365	0.572	0.471
SMO-Poly- 10^1	57.244	42.757	3.942	0.572	0.572	0.665	0.365	0.572	0.471
SMO-Poly- 10^2	57.244	42.757	3.942	0.572	0.572	0.665	0.365	0.572	0.471
SMO-Poly- 10^3	57.244	42.757	3.942	0.572	0.572	0.665	0.365	0.572	0.471
SMO-Poly- 10^4	57.244	42.757	3.942	0.572	0.572	0.665	0.365	0.572	0.471
SMO-Poly- 10^5	57.244	42.757	3.942	0.572	0.572	0.665	0.365	0.572	0.471
SMO-RBF-0	0.962	99.039	0.717	0.010	0.010	0.000	0.050	0.010	0.001
SMO-RBF- 10^{-5}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-RBF- 10^{-4}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-RBF- 10^{-3}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-RBF- 10^{-2}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-RBF- 10^{-1}	21.859	78.141	2.619	0.219	0.219	0.011	0.050	0.219	0.018
SMO-RBF- 10^0	58.590	41.410	3.816	0.586	0.586	0.481	0.356	0.586	0.409
SMO-RBF- 10^1	74.423	25.577	3.080	0.744	0.744	0.823	0.539	0.744	0.651
SMO-RBF- 10^2	74.487	25.513	3.164	0.745	0.745	0.816	0.543	0.745	0.652
SMO-RBF- 10^3	74.487	25.513	3.164	0.745	0.745	0.816	0.543	0.745	0.652
SMO-RBF- 10^4	74.487	25.513	3.164	0.745	0.745	0.816	0.543	0.745	0.652
SMO-RBF- 10^5	74.487	25.513	3.164	0.745	0.745	0.816	0.543	0.745	0.652

Table 3.90: Classification Results for WAP collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	$\mathbf{F}_1^\mu$	$\mathbf{F}_1^M$
IB1	72.115	27.885	2.738	0.721	0.721	0.650	0.608	0.721	0.628
IB3	72.372	27.628	2.056	0.724	0.724	0.700	0.590	0.724	0.640
IB5	73.462	26.539	2.149	0.735	0.735	0.741	0.574	0.735	0.646
IB7	72.115	27.885	2.333	0.721	0.721	0.723	0.540	0.721	0.619
IB9	72.436	27.564	2.532	0.724	0.724	0.751	0.528	0.724	0.620
IB11	72.564	27.436	2.921	0.726	0.726	0.738	0.521	0.726	0.611
IB13	71.282	28.718	2.561	0.713	0.713	0.693	0.494	0.713	0.576
IB15	70.833	29.167	3.208	0.708	0.708	0.727	0.485	0.708	0.582
IB17	69.551	30.449	2.647	0.696	0.696	0.689	0.465	0.696	0.555
IB19	69.423	30.577	2.938	0.694	0.694	0.692	0.465	0.694	0.556
IB21	68.462	31.539	3.164	0.685	0.685	0.692	0.446	0.685	0.543
IB25	67.692	32.308	3.090	0.677	0.677	0.661	0.442	0.677	0.529
IB29	66.667	33.333	2.938	0.667	0.667	0.556	0.427	0.667	0.483
IB35	64.487	35.513	2.996	0.645	0.645	0.545	0.401	0.645	0.462
IB41	62.756	37.244	2.555	0.628	0.628	0.548	0.384	0.628	0.452
IB49	60.449	39.551	2.952	0.604	0.604	0.550	0.363	0.604	0.438
IB57	58.205	41.795	3.725	0.582	0.582	0.551	0.343	0.582	0.423
IB73	54.808	45.192	4.235	0.548	0.548	0.495	0.312	0.548	0.383
IB89	52.628	47.372	4.732	0.526	0.526	0.422	0.290	0.526	0.344
IB1-WV	72.115	27.885	2.738	0.721	0.721	0.650	0.608	0.721	0.628
IB3-WV	73.590	26.410	1.565	0.736	0.736	0.694	0.608	0.736	0.648
IB5-WV	74.487	25.513	1.919	0.745	0.745	0.755	0.594	0.745	0.665
IB7-WV	73.590	26.410	1.538	0.736	0.736	0.762	0.570	0.736	0.652
IB9-WV	74.167	25.833	1.814	0.742	0.742	0.720	0.560	0.742	0.630
IB11-WV	73.590	26.410	2.235	0.736	0.736	0.753	0.550	0.736	0.635
IB13-WV	72.564	27.436	2.593	0.726	0.726	0.711	0.524	0.726	0.603
IB15-WV	72.244	27.757	2.966	0.722	0.722	0.726	0.509	0.722	0.599
IB17-WV	71.538	28.462	2.884	0.715	0.715	0.726	0.498	0.715	0.591
IB19-WV	71.154	28.846	3.534	0.712	0.712	0.741	0.485	0.712	0.587
IB21-WV	70.705	29.295	2.720	0.707	0.707	0.715	0.479	0.707	0.573
IB25-WV	69.103	30.898	2.671	0.691	0.691	0.675	0.457	0.691	0.545
IB29-WV	68.077	31.923	2.717	0.681	0.681	0.665	0.445	0.681	0.533
IB35-WV	67.115	32.885	2.765	0.671	0.671	0.611	0.429	0.671	0.504
IB41-WV	66.410	33.590	3.548	0.664	0.664	0.568	0.420	0.664	0.483
IB49-WV	64.359	35.641	2.940	0.644	0.644	0.568	0.398	0.644	0.468
IB57-WV	62.179	37.821	3.231	0.622	0.622	0.564	0.375	0.622	0.450
IB73-WV	58.654	41.346	4.216	0.587	0.587	0.551	0.339	0.587	0.419
IB89-WV	55.897	44.103	4.718	0.559	0.559	0.431	0.316	0.559	0.364

3.45 WebKb Collection

Table 3.91 presents the classification results obtained by Naïve Bayes, Multinomial Naïve Bayes, J48, and SMO, and Table 3.92 presents the classification results obtained by IBk for WebKb collection.

Table 3.91: Classification Results for WebKb collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
Naïve Bayes	41.391	58.609	1.855	0.414	0.414	0.352	0.428	0.414	0.386
Multinomial NB	60.384	39.616	1.099	0.604	0.604	0.515	0.605	0.604	0.557
J48	69.090	30.910	1.697	0.691	0.691	0.571	0.543	0.691	0.557
SMO-Linear-0	13.572	86.428	1.276	0.136	0.136	0.019	0.143	0.136	0.034
SMO-Linear- 10^{-5}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-Linear- 10^{-4}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-Linear- 10^{-3}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-Linear- 10^{-2}	45.411	54.589	1.728	0.454	0.454	0.255	0.144	0.454	0.184
SMO-Linear- 10^{-1}	57.208	42.793	2.440	0.572	0.572	0.511	0.305	0.572	0.382
SMO-Linear- 10^0	64.223	35.777	2.204	0.642	0.642	0.541	0.489	0.642	0.514
SMO-Linear- 10^1	63.728	36.272	1.983	0.637	0.637	0.529	0.514	0.637	0.521
SMO-Linear- 10^2	61.711	38.289	1.924	0.617	0.617	0.507	0.513	0.617	0.510
SMO-Linear- 10^3	60.190	39.810	1.709	0.602	0.602	0.487	0.504	0.602	0.495
SMO-Linear- 10^4	59.647	40.353	1.816	0.596	0.596	0.480	0.503	0.596	0.491
SMO-Linear- 10^5	59.261	40.739	1.724	0.593	0.593	0.478	0.504	0.593	0.490
SMO-Poly-0	13.572	86.428	1.276	0.136	0.136	0.019	0.143	0.136	0.034
SMO-Poly- 10^{-5}	45.436	54.565	1.696	0.454	0.454	0.065	0.143	0.454	0.089
SMO-Poly- 10^{-4}	45.399	54.601	1.676	0.454	0.454	0.065	0.143	0.454	0.089
SMO-Poly- 10^{-3}	45.037	54.963	1.654	0.450	0.450	0.116	0.142	0.450	0.128
SMO-Poly- 10^{-2}	44.204	55.796	1.619	0.442	0.442	0.231	0.150	0.442	0.182
SMO-Poly- 10^{-1}	45.556	54.444	1.718	0.456	0.456	0.345	0.213	0.456	0.263
SMO-Poly- 10^0	47.222	52.778	2.582	0.472	0.472	0.351	0.274	0.472	0.307
SMO-Poly- 10^1	49.287	50.713	2.499	0.493	0.493	0.373	0.328	0.493	0.349
SMO-Poly- 10^2	50.120	49.880	2.281	0.501	0.501	0.372	0.359	0.501	0.365
SMO-Poly- 10^3	50.663	49.337	1.934	0.507	0.507	0.374	0.379	0.507	0.377
SMO-Poly- 10^4	50.289	49.711	2.018	0.503	0.503	0.368	0.383	0.503	0.375
SMO-Poly- 10^5	50.784	49.216	1.882	0.508	0.508	0.368	0.386	0.508	0.377
SMO-RBF-0	13.572	86.428	1.276	0.136	0.136	0.019	0.143	0.136	0.034
SMO-RBF- 10^{-5}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-RBF- 10^{-4}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-RBF- 10^{-3}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-RBF- 10^{-2}	45.448	54.552	1.680	0.454	0.454	0.065	0.143	0.454	0.089
SMO-RBF- 10^{-1}	45.870	54.130	1.738	0.459	0.459	0.379	0.148	0.459	0.213
SMO-RBF- 10^0	45.870	54.130	1.738	0.459	0.459	0.379	0.148	0.459	0.213
SMO-RBF- 10^1	62.086	37.914	1.814	0.621	0.621	0.584	0.371	0.621	0.454
SMO-RBF- 10^2	65.431	34.569	1.663	0.654	0.654	0.574	0.478	0.654	0.522
SMO-RBF- 10^3	64.259	35.741	1.556	0.643	0.643	0.564	0.488	0.643	0.523
SMO-RBF- 10^4	62.678	37.322	1.406	0.627	0.627	0.542	0.476	0.627	0.507
SMO-RBF- 10^5	62.002	37.998	1.628	0.620	0.620	0.528	0.475	0.620	0.500

Table 3.92: Classification Results for WebKb collection.

Algorithm	CCR	ICR	$\sigma(\text{CCR})$	Pr^μ	Re^μ	Pr^M	Re^M	F_1^μ	F_1^M
IB1	60.915	39.085	1.024	0.609	0.609	0.483	0.530	0.609	0.505
IB3	63.016	36.984	1.544	0.630	0.630	0.522	0.539	0.630	0.531
IB5	65.539	34.461	1.547	0.655	0.655	0.541	0.546	0.655	0.544
IB7	66.735	33.265	1.834	0.667	0.667	0.577	0.559	0.667	0.568
IB9	66.964	33.036	1.190	0.670	0.670	0.564	0.551	0.670	0.557
IB11	67.206	32.794	1.498	0.672	0.672	0.604	0.553	0.672	0.577
IB13	66.880	33.120	1.703	0.669	0.669	0.577	0.545	0.669	0.560
IB15	67.109	32.891	1.668	0.671	0.671	0.580	0.546	0.671	0.562
IB17	66.807	33.193	1.817	0.668	0.668	0.562	0.541	0.668	0.551
IB19	66.880	33.120	1.768	0.669	0.669	0.577	0.541	0.669	0.559
IB21	66.892	33.108	1.879	0.669	0.669	0.506	0.540	0.669	0.522
IB25	66.904	33.096	2.240	0.669	0.669	0.511	0.539	0.669	0.525
IB29	66.614	33.386	2.026	0.666	0.666	0.507	0.536	0.666	0.521
IB35	67.073	32.927	2.494	0.671	0.671	0.519	0.540	0.671	0.529
IB41	66.638	33.362	2.297	0.666	0.666	0.517	0.537	0.666	0.527
IB49	66.832	33.169	2.413	0.668	0.668	0.517	0.537	0.668	0.527
IB57	66.662	33.338	2.207	0.667	0.667	0.667	0.534	0.667	0.593
IB73	66.397	33.603	1.938	0.664	0.664	0.525	0.534	0.664	0.530
IB89	65.878	34.122	1.754	0.659	0.659	0.522	0.531	0.659	0.527
IB1-WV	60.903	39.097	1.011	0.609	0.609	0.482	0.530	0.609	0.505
IB3-WV	63.861	36.139	1.726	0.639	0.639	0.509	0.548	0.639	0.528
IB5-WV	66.530	33.471	1.533	0.665	0.665	0.541	0.556	0.665	0.548
IB7-WV	67.302	32.698	2.094	0.673	0.673	0.561	0.561	0.673	0.561
IB9-WV	67.532	32.468	1.550	0.675	0.675	0.573	0.556	0.675	0.564
IB11-WV	67.954	32.046	1.737	0.680	0.680	0.591	0.558	0.680	0.574
IB13-WV	67.508	32.493	1.577	0.675	0.675	0.579	0.551	0.675	0.564
IB15-WV	67.725	32.275	1.928	0.677	0.677	0.578	0.552	0.677	0.565
IB17-WV	67.483	32.517	1.870	0.675	0.675	0.584	0.548	0.675	0.565
IB19-WV	67.532	32.468	1.896	0.675	0.675	0.608	0.547	0.675	0.576
IB21-WV	67.459	32.541	2.017	0.675	0.675	0.584	0.545	0.675	0.564
IB25-WV	67.664	32.336	2.293	0.677	0.677	0.609	0.546	0.677	0.576
IB29-WV	67.387	32.613	2.153	0.674	0.674	0.657	0.543	0.674	0.595
IB35-WV	67.580	32.420	2.643	0.676	0.676	0.666	0.546	0.676	0.600
IB41-WV	67.133	32.867	2.205	0.671	0.671	0.523	0.540	0.671	0.531
IB49-WV	67.363	32.637	2.440	0.674	0.674	0.527	0.540	0.674	0.533
IB57-WV	67.266	32.734	2.314	0.673	0.673	0.525	0.539	0.673	0.532
IB73-WV	66.892	33.108	2.073	0.669	0.669	0.530	0.537	0.669	0.534
IB89-WV	66.481	33.519	1.587	0.665	0.665	0.532	0.536	0.665	0.534

3.46 Synthesis of the Classification Results

This section presents a synthesis of the classification results. Table 4.1 presents the best classification accuracies obtained by the classification algorithms used in this technical report. We highlight with bold font the highest classification accuracies for each collection.

Considering the results from Table 4.1, we analysed which classification algorithm is advisable for each domain. To do so we generate the average ranking [Demsar, 2006] of the algorithms considering the collections from each domain (see Table 2.91). Figure 3.1 presents the average ranking for each domain. We notice by the diagrams that the most accurate classifier varies for each domain. MNB obtained the best accuracies for most collections from MD, NA, SD, and WP domains. IB k obtained the best accuracies mainly for collections from EM and TREC domains, and SMO was advisable for collections from SA domain.

Table 3.93: Best accuracies for the classification algorithms used in this technical report.

Collection	NB	MNB	J48	SMO	IBk
20ng	64.61	90.08	73.96	84.81	86.71
ACM	73.89	76.04	66.36	77.38	64.36
Classic4	88.48	95.79	90.35	94.53	94.24
CSTR	78.59	84.64	66.85	75.26	82.29
Dmoz-Business-500	58.61	68.28	53.10	65.84	61.49
Dmoz-Health-500	73.15	82.07	73.61	79.90	77.83
Dmoz-Computers-500	59.67	70.13	54.88	66.01	63.70
Dmoz-Science-500	62.11	73.81	57.20	67.38	64.38
Dmoz-Sports-500	75.85	83.76	83.88	85.48	80.09
Enron-Top-20	51.54	72.60	64.47	65.98	66.72
Fbis	61.79	77.18	71.49	78.92	80.99
Hitech	62.92	72.92	56.76	66.44	71.79
Industry-Sector	41.54	76.23	57.56	70.34	77.58
IrishEconomic	59.75	67.65	51.08	65.54	60.90
La1s	75.21	88.17	76.65	84.30	80.55
La2s	75.25	89.91	76.84	86.76	82.79
LATimes	70.76	85.00	75.68	81.11	76.83
Multi-Domain-Sentiment	72.85	78.65	74.95	81.90	71.40
New3	56.72	79.16	70.85	71.93	79.26
NFS	70.86	83.84	70.74	81.87	78.88
Oh0	79.66	89.83	80.95	81.55	81.85
Oh5	78.76	86.27	80.39	77.24	79.41
Oh10	72.38	80.66	72.09	76.00	73.61
Oh15	75.03	83.68	75.78	75.03	75.57
Ohscal	62.78	74.73	71.30	76.69	68.65
Ohsumed-400	34.75	40.51	30.91	35.71	38.16
Opinosis	60.74	59.56	60.83	61.03	62.87
Pubmed-Cancer	69.90	86.61	96.80	95.98	–
Re0	57.05	79.92	75.26	77.79	83.51
Re1	66.73	83.34	79.60	72.72	81.89
Re8	81.27	95.33	90.73	93.95	94.14
Review-Polarity	66.80	80.10	68.25	83.65	70.50
Reviews	85.22	93.33	88.35	91.64	92.30
SpamAssassin	87.80	96.51	96.61	98.96	98.72
SpamTrec-3000	81.42	97.42	96.88	97.72	98.61
SyskillWebbert	72.51	90.75	95.81	77.85	95.81
Tr11	54.06	85.00	78.98	77.06	86.95
Tr12	57.82	80.15	79.23	69.61	81.74
Tr21	47.95	61.35	81.27	79.77	88.66
Tr23	56.85	70.61	93.19	72.54	84.33
Tr31	79.72	94.38	93.10	90.72	94.60
Tr41	84.95	94.52	90.77	87.80	93.04
Tr45	66.66	82.46	90.28	81.30	88.84
WAP	71.73	81.02	67.05	81.85	74.48
WebKb	41.39	60.38	69.08	57.20	67.95

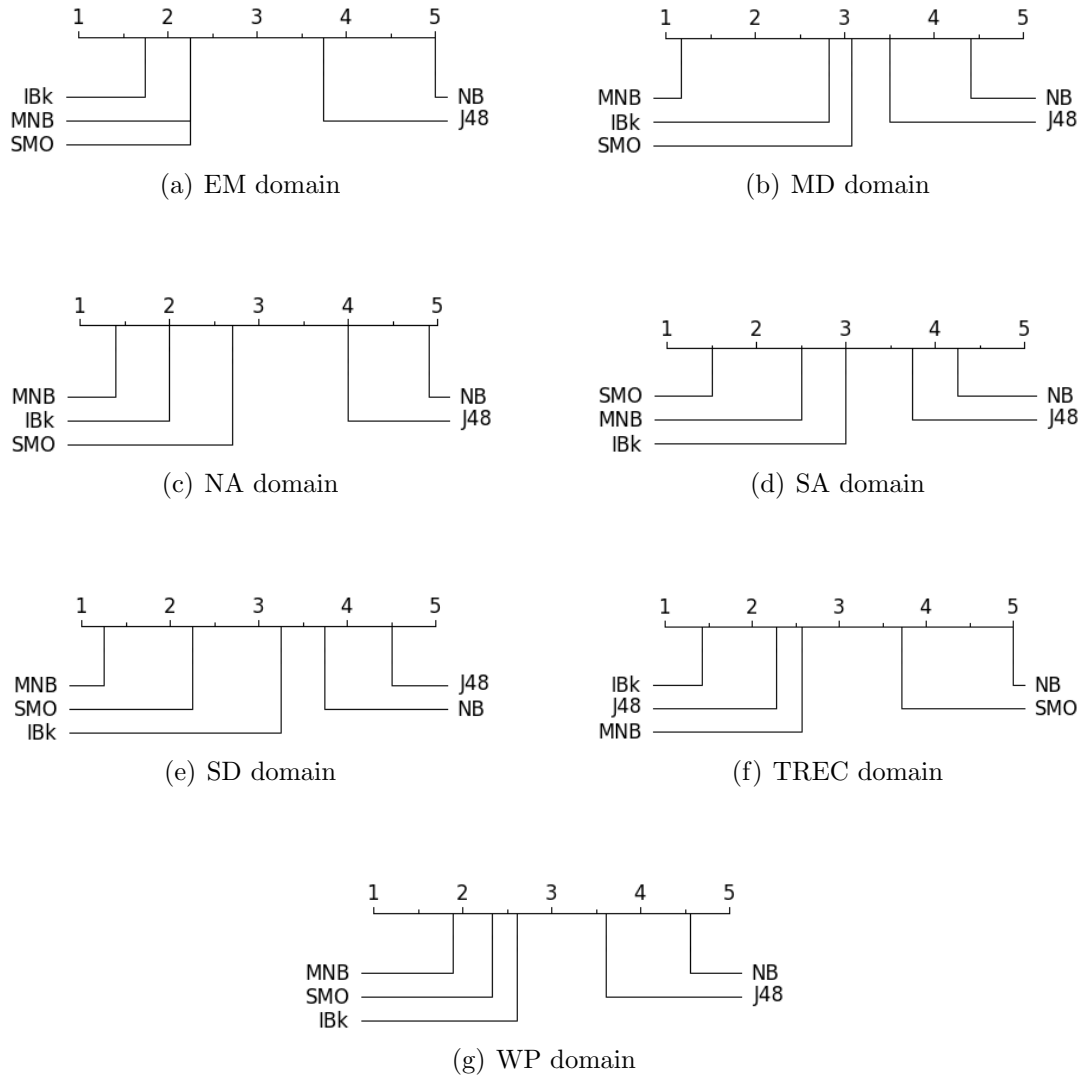


Figure 3.1: Average ranking diagrams for classification results.

Chapter 4

Clustering Results

We used the Torch¹ toolkit to generate the text clustering results. The Torch toolkit was developed in Java language and supports the ARFF file format for documents representation.

We run two traditional clustering algorithms: k -means (partitional clustering) e bi-secting k -means (hierarchical clustering). The number of classes of the text collections was used to define the value of k (number of clusters) and the cosine correlation was used as similarity measure. In all the executions of k -means, we used random initialization of centroids and a maximum of 30 iterations for convergence. Furthermore, we selected the partition with lower quadratic error from 10 different runs of k -means.

We used the F_{SCORE} measure to quantitatively assess the quality of the obtained clustering results. It is essentially an information retrieval measure that computes how much the clustering solution can recover the class information associated with each document. For this purpose, consider the following:

- T is a clustering solution;
- Q_i is a single cluster belonging to T , where Q_i contains a set of documents; and
- L_r is a class of reference and its respective set of documents.

The F measure of a class L_r is calculated by choosing the maximum value obtained in some cluster $Q_i \in T$, according to Eq. (4.1). In this case, $F(L_r, Q_i)$ is a mean between precision $P(L_r, Q_i) = \frac{|L_r \cap Q_i|}{|Q_i|}$ and recall $R(L_r, Q_i) = \frac{|L_r \cap Q_i|}{|L_r|}$. The F_{SCORE} of a clustering solution with n documents and c classes is the sum of the F values of the classes weighted by the size of each class (Eq. 4.2).

$$F(L_r) = \max_{Q_i \in T} F(L_r, Q_i) \quad (4.1)$$

$$F_{SCORE} = \sum_{r=1}^c \frac{|L_r|}{n} F(L_r) \quad (4.2)$$

¹Torch is available at <http://sites.labic.icmc.usp.br/torch/>

Thus, if the clustering solution perfectly recovers the information class of the documents, then the F_{SCORE} measure is equal to 1. In general, the higher the F_{SCORE} values, the better the clustering solution.

Table 4.1 present the F_{SCORE} for the collections analyzed in this technical report. We highlight with bold font the highest F_{SCORE} for each collection.

Considering the results from Table 4.1, we analyzed which clustering algorithm is advisable for each domain. To do so we generated the average ranking of the algorithms considering the collections from each domain. Figure 4.1 presents the average ranking for each domain. We noticed by the diagrams that bisecting k -means is advisable for all the domains.

Table 4.1: F_{SCORE} for the clustering algorithms used in this technical report.

Collection	K-means	Bisecting K-means
20ng	0.419	0.459
ACM	0.411	0.461
Classic3	0.916	0.797
CSTR	0.467	0.718
Dmoz-Business-500	0.216	0.184
Dmoz-Computers-500	0.297	0.289
Dmoz-Health-500	0.386	0.402
Dmoz-Science-500	0.276	0.259
Dmoz-Sports-500	0.439	0.349
Enron-Top-20	0.349	0.427
FBIS	0.563	0.607
Hitech	0.522	0.589
Industry-Sector	0.206	0.259
IrishEcnomic	0.407	0.500
La1s	0.477	0.523
La2s	0.509	0.604
LATimes	0.411	0.505
Multi-Domain-Sentiment	0.606	0.544
New3	0.410	0.527
NFS	0.327	0.445
Oh0	0.525	0.601
Oh10	0.537	0.568
Oh15	0.509	0.513
Oh5	0.519	0.524
Ohscal	0.503	0.426
Ohsumed-400	0.152	0.152
Opinosis	0.545	0.601
PubMed-Cancer	0.372	0.316
Re0	0.411	0.652
Re1	0.463	0.589
Re8	0.515	0.812
Review-Polarity	0.566	0.568
Reviews	0.771	0.804
SpamAssassin	0.619	0.750
SpamTrec-3000	0.661	0.685
SyskillWebert	0.830	0.933
Tr11	0.458	0.671
Tr12	0.342	0.681
Tr21	0.612	0.771
Tr23	0.397	0.619
Tr31	0.630	0.712
Tr41	0.571	0.693
Tr45	0.412	0.620
WAP	0.461	0.608
WebKb	0.364	0.520

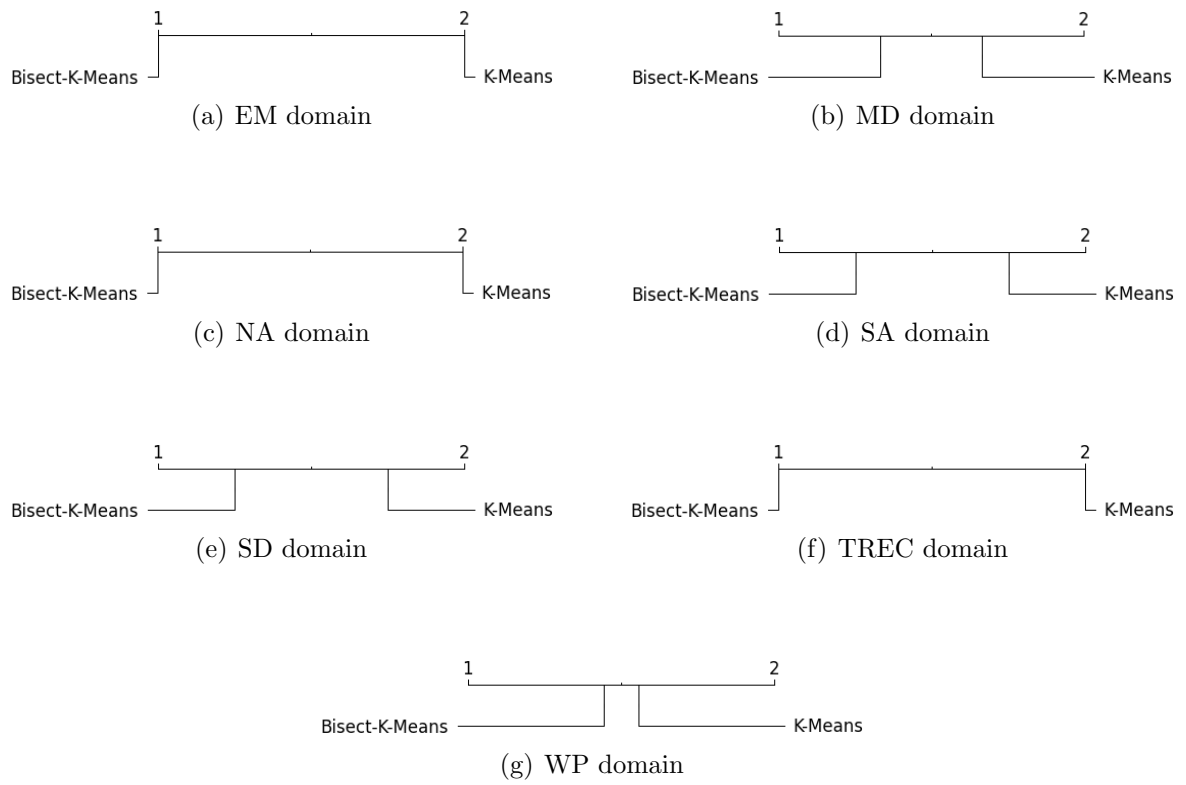


Figure 4.1: Average ranking diagrams for clustering results.

Chapter 5

Final Considerations

This technical report aims to solve the lack of benchmarking collections and results found in literature. To do so we make 45 preprocessed text collections available in an on-line repository: http://sites.labic.icmc.usp.br/text_collections/. The tool developed to preprocess the text collections (Text Preprocessing Tool) is available at <http://sites.labic.icmc.usp.br/tpt/>.

The collections used in this technical report are from different domains: e-mails, medical, news, sentiment analysis, scientific, and TREC. The collections have different characteristics. The number of documents ranges from 204 to 65991, the number of terms from 1762 to 100464, the average number of terms per document from 6.55 to 720.30, and the number of classes from 2 to 51. Besides this traditional characteristics, we extract another characteristic as the matrix sparsity, standard deviation of the number of documents per class, majority class and S-Index (see Section 2).

We also provided classification and clustering results using traditional/state-of-the-art algorithms. The used algorithms for classification were: Naïve Bayes, Multinomial Naïve Bayes, C4.5, Support Vector Machine, and k Nearest Neighbors. The tool developed to generate the classification results (Inductive Classification Tool) is available at <http://sites.labic.icmc.usp.br/ict/>. The experiments carried out in this report demonstrated that Multinomial Naïve Bayes is advisable to be used in the MD, NA, SD, and WP domains, k NN is advisable in EM and TREC domains, and SVM is advisable in SA domain.

For clustering we used k -means and bisecting- k -means algorithms. The tool developed to generate the clustering results (Torch) is available at <http://sites.labic.icmc.usp.br/torch/>. We identified that, in general, bisecting- k -means provided better results than k -means for collections from all the domains.

This technical report provides collections with their characteristic and baseline results to be used as comparison in the proposal of new algorithms. This is also useful to guide users in a choice of algorithms and parameters in practical applications.

Bibliography

- [Apache, 2006] Apache (2006). The apache spamassassin project. <http://spamassassin.apache.org/publiccorpus/>. Last access: November 6 2013. Cited on page 43.
- [Berry, 2004] Berry, M. (2004). *Survey of Text Mining I: Clustering, Classification, and Retrieval*. Number v. 1. Springer. Cited on page 1.
- [Berry et al., 2008] Berry, M., Castellanos, M., for Industrial, S., and Mathematics, A. (2008). *Survey of Text Mining II: Clustering, Classification, and Retrieval*. Number v. 2 in Computer science. Springer. Cited on page 1.
- [Blitzer et al., 2009] Blitzer, J., Dredze, M., and Pereira, F. (2009). Multi-domain sentiment dataset (version 2.0). <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>. Last access: November 6 2013. Cited on page 25.
- [Caruana and Niculescu-Mizil, 2006] Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the International Conference on Machine Learning*, pages 161–168. Cited on page 57.
- [Cohen, 2009] Cohen, W. W. (2009). Enron email dataset. <http://www.cs.cmu.edu/~enron/>. Last access: July 19 2013. Cited on page 16.
- [Cormack and Lynam, 2007] Cormack, G. V. and Lynam, T. R. (2007). 2007 trec public spam corpus. <http://plg.uwaterloo.ca/~gvcormac/treccorpus07/>. Last access: November 6 2013. Cited on page 44.
- [Demsar, 2006] Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30. Cited on page 150.
- [Forman, 2006] Forman, G. (2006). 19MclassTextWc dataset. <http://sourceforge.net/projects/weka/files/datasets/text-datasets/19MclassTextWc.zip/download>. Last access: November 6 2013. Cited on pages 18, 22, 23, 26, 29, 30, 31, 32, 33, 38, 39, 40, 46, 47, 48, 49, 50, 51, 52, and 53.
- [Ganesan, 2012] Ganesan, K. (2012). Opinosis dataset - topic related review sentences. <http://kavita-ganesan.com/opinosis-opinion-dataset>. Last access: November 6 2013. Cited on page 35.

- [Group, 2009] Group, M. L. (2009). Irish economic sentiment dataset. <http://mlg.ucd.ie/sentiment>. Last access: July 12 2012. Cited on page 21.
- [Han et al., 1998] Han, E.-H., Boley, D., Gini, M., Gross, R., Hastings, K., Karypis, G., Kumar, V., Mobasher, B., and Moore, J. (1998). Webace: a web agent for document categorization and exploration. In *Proceedings of the International Conference on Autonomous Agents*, pages 408–415. ACM. Cited on page 53.
- [Hersh et al., 1994] Hersh, W., Buckley, C., Leone, T. J., and Hickam, D. (1994). Ohsumed: an interactive retrieval evaluation and new large test collection for research. In *Proceedings of the Annual International Conference on Research and Development in Information Retrieval*, pages 192–201. Springer-Verlag New York, Inc. Cited on pages 29, 30, 31, 32, 33, and 34.
- [Karypis, 2006] Karypis, G. (2006). Cluto - software for clustering high-dimensional datasets. <http://glaros.dtc.umn.edu/gkhome/cluto/cluto/download>. Last access: November 6 2013. Cited on pages 19 and 42.
- [Kogan, 2007] Kogan, J. (2007). *Introduction to Clustering Large and High-Dimensional Data*. Cambridge University Press. Cited on page 4.
- [Lewis, 2004] Lewis, D. D. (2004). Reuters-21578. <http://www.daviddlewis.com/resources/testcollections/reuters21578/>. Last access: November 6 2013. Cited on pages 38, 39, and 40.
- [Moschitti, 1997] Moschitti, A. (1997). Text categorization corpora. <http://disi.unitn.it/moschitti/corpora/ohsumed-all-docs.tar.gz>. Last access: November 6 2013. Cited on page 34.
- [Netscape, 2013] Netscape (2013). Dmoz - open directory project. <http://www.dmoz.org/>. Last access: July 19 2013. Cited on pages 10, 12, 13, 14, and 15.
- [Nigam, 2000] Nigam, K. (2000). The industry sector dataset. <http://www.cs.cmu.edu/~TextLearning/sector-data.tar.gz>. Last access: July 12 2012. Cited on page 20.
- [Pang and Lee, 2004] Pang, B. and Lee, L. (2004). Movie review data. <http://www.cs.cornell.edu/People/pabo/movie-review-data/>. Last access: November 6 2013. Cited on page 41.
- [Pazzani, 1998] Pazzani, M. (1998). Syskill and webert web page ratings data set. <http://archive.ics.uci.edu/ml/datasets/Syskill+and+Webert+Web+Page+Ratings>. Last access: November 6 2013. Cited on page 45.
- [Pazzani and Meyers, 2003] Pazzani, M. J. and Meyers, A. (2003). National science foundation dataset. <http://archive.ics.uci.edu/ml/databases/nsfabs/>. Last access: November 6 2013. Cited on page 28.

- [Porter, 1980] Porter, M. F. (1980). An algorithm for suffix stripping. *Readings in Information Retrieval*, 14(3):130–137. Cited on page 3.
- [Rennie, 2008] Rennie, J. (2008). 20 newsgroup dataset. <http://qwone.com/~jason/20Newsgroups/>. Last access: July 18 2013. Cited on page 5.
- [Research, 2010] Research, D. M. (2010). Classic3 and classic4 datasets. <http://www.dataminingresearch.com/index.php/2010/09/classic3-classic4-datasets/>. Last access: July 19 2013. Cited on page 8.
- [Salton, 1989] Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co., Inc. Cited on page 3.
- [Sebastiani, 2002] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Survey*, 34(1):1–47. Cited on pages ix and 58.
- [Soares et al., 2008] Soares, M. V. B., Prati, R. C., and Monard, M. C. (2008). PRETEXT II: Descrição da reestruturação da ferramenta de pré-processamento de textos. Technical Report 333, ICMC-USP. http://www.icmc.usp.br/~biblio/BIBLIOTECA/rel_tec/RT_333.pdf. Cited on page 3.
- [Tan et al., 2005] Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley. Cited on page 58.
- [TREC, 2013] TREC (2013). Text REtrieval Conference data. <http://trec.nist.gov/data.html>. Last access: November 6 2013. Cited on pages 1, 18, 19, 22, 23, 46, 47, 48, 49, 50, 51, and 52.
- [Weiss et al., 2010] Weiss, S., Indurkha, N., and Zhang, T. (2010). *Fundamentals of Predictive Text Mining*. Texts in computer science. Springer-Verlag. Cited on page 1.
- [Witten and Frank, 2005] Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2 edition. Cited on pages 3 and 57.