



Aprendizado de Máquina Aplicado na Classificação de Resolutividade de Reclamações de Serviços Financeiros

Luiz Henrique Rowan Peixoto,¹ Solange Oliveira Rezende²

ICMC - Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)

1 Introdução

As reclamações de clientes estão entre as muitas manifestações de opiniões feitas na internet atualmente. Estas reclamações são resultado de sentimentos de insatisfação em relação a produtos e serviços contratados. A insatisfação é decorrente de uma expectativa não atendida em uma relação de consumo, que pode ser gerada pela ausência de qualidade do produto, por uma cobrança indevida, falhas na entrega ou habilitação do serviço, falta de entendimento dos termos contratuais, suporte inadequado, entre outras.

As *Online Dispute Resolution* (ODR) são plataformas digitais de resolução de conflitos nas relações de consumo [4]. Estas plataformas podem ser geridas pelo poder público ou pela iniciativa privada, além disso, caracterizam-se por disponibilizar um serviço na internet, pelo acompanhamento das reclamações desde a originação até a conclusão. Dentre as plataformas brasileiras de ODR, destaca-se o ReclameAqui que entre 13/01/2021 e 12/04/2021 foi o 11º site brasileiro mais visitado [2]. A plataforma tem como função o cadastramento e a condução das reclamações, porém o maior serviço que este tipo de plataforma fornece não é individual e sim coletivo. No ReclameAqui mais de 600 mil pessoas pesquisam mensalmente reputações de empresas antes de realizar uma compra ou contratar um serviço. O objetivo é, portanto, persuadir as empresas a corrigir suas falhas como forma de melhorar os indicadores que podem levar a um risco de imagem perante aos consumidores propensos a consumir seus produtos.

Diante do contexto da importância do atendimento adequado das reclamações pelas empresas, este trabalho tem como objetivo avaliar o uso técnicas de Mineração de Textos (MT) com Aprendizado de Máquina (AM) para classificação em dados não-estruturados de reclamações de serviços do sistema financeiro coletados na plataforma ReclameAQUI. A utilização de técnicas de MT para extração e representação textual em conjunto com algoritmos de AM para extração de padrões em textos tem-se mostrado uma abordagem bastante utilizada para resolução de problemas dos consumidores e de conflitos entre consumidores e empresas. Essas técnicas tem sido empregadas para

¹henriquerowan@usp.br

²solange@icmc.usp.br

análise de sentimentos, sumarização de reclamações, tratamento especial para problemas complexos, identificação de problemas generalizados, desenvolvimento de novos produtos, entre outros.

2 Processo de Mineração de Textos

A MT é uma sub-área da mineração de dados, e pode ser definida como um processo de extração de informações úteis a partir de documentos textuais escritos em linguagem natural [6]. Este processamento é feito por meio de uma variedade de algoritmos estatísticos e de AM gerando aplicações práticas a partir da informação gerada [1]. Este trabalho utiliza o processo de MT definido por [6] que contém 5 etapas genéricas aplicáveis a diversos tipos de problemas: Identificação do Problema, Pré-Processamento, Extração de Padrões, Pós-processamento e Utilização do Conhecimento (Figura 1).



Figura 1: Processo de MT [6]

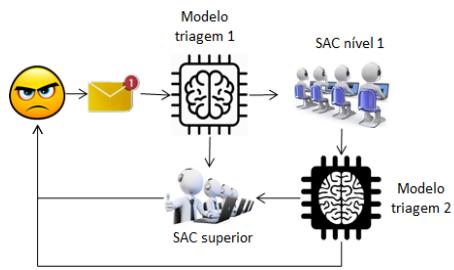


Figura 2: Fluxo de atendimento com modelos de triagem.

Para capturar os dados utilizados no trabalho foi escolhido entre os 5 principais bancos brasileiros o que apresentou a maior variabilidade de respostas, portanto com foco nas narrativas das respostas que apresentam a solução do problema de forma detalhada. A variabilidade textual da resposta da empresa é imprescindível no processo de aprendizagem e generalização. A base de dados textuais foi montada a partir da captura de 6000 reclamações finalizadas entre abril/2020 e janeiro/2021, capturadas de forma aleatória na plataforma ReclameAqui. As variáveis capturadas foram: título da reclamação, narrativa da reclamação, narrativa da resposta da empresa e também se o cliente entende que a reclamação foi resolvida ou não (variável dependente). A variável dependente de resolutividade apresenta distribuição de 54,7% e 45,3% para as classes resolvido e não-resolvido, ou seja, é uma variável binária sem comportamento de classe rara.

Em um processo de atendimento de reclamação tradicional o reclamante envia sua reclamação por meio da plataforma de ODR e o Serviço de Atendimento ao Cliente (SAC) nível 1 coordena o atendimento e formaliza a solução na plataforma. A resolução dada pelo atendimento pode ou não atender a necessidade do cliente, e o mesmo irá formalizar na plataforma ao finalizar a reclamação. Neste cenário este trabalho propõe a utilização de modelos de triagem para direcionamento de um conjunto de reclamações com baixa probabilidade de resolução ao nível de atendimento superior (Figura 2). Um nível de atendimento superior pode envolver mais acessos a informação e tomada de decisão, um canal especial de resposta com custo maior e/ou concessão de benefícios ao cliente.

Os modelos de triagem no processo de atendimento (Figura 2), tem como objetivo classificar a resolutividade das reclamações de forma preemptiva, direcionando dentro da política de atendi-

mento os casos com menor probabilidade de resolutividade para o atendimento superior. O modelo de triagem 1 foi treinado com dados textuais das narrativas dos clientes, enquanto que o modelo de triagem 2 foi treinado também com as narrativas da empresa.

Os experimentos foram executados utilizando 3 abordagens de aprendizado de máquina muito utilizadas para mineração de textos, porém com grau de complexidade distintas. Como modelo base utilizou-se o Naive-Bayes SVM [5], para o intermediário foi utilizado as *embeddings* FastText [3] com classificador *Multi-layer Perceptron* e como técnica desafiante foi utilizado o modelo de linguagem Distilbert [7]. Para treinamento dos modelos foi utilizado *holdout* com 90% dos dados para treino e 10% para testes.

Na etapa de pré-processamento dos dados foi mantido apenas os códigos alfanuméricos com remoção de pontuações, símbolos e múltiplos espaçamentos. Para obter o modelo de representação usado em cada um dos modelos é necessário primeiramente separar os textos em *tokens*. Nas representações usadas pelo NBSVM e FastText, os tokens foram separados pelo caracter de espaço, enquanto que no Distilbert o *tokenizer* é específico do modelo de linguagem. Devido a restrição do Distilbert em utilizar no máximo 512 *tokens* por documento, as reclamações que excederam este limite foram tratadas com a estratégia *head and tail* onde o texto foi reduzido considerando 33% dos *tokens* no inicio e 67% dos *tokens* ao final.

3 Resultados

Os resultados mostrados na Tabela 1 derivados da matriz de confusão (Precisão, Revocação, F1-Score e Acurácia) foram otimizados por F1-Score, ou seja, o limiar de probabilidade escolhido maximizava o parâmetro. A métrica KS é derivada do teste de Kolmogorov–Smirnov, sendo esta muito utilizada na indústria financeira para avaliação de performance de modelos. No KS é avaliado qual é o poder de separação das classes binárias ordenado pela probabilidade dada pelo modelo. O valor de KS usado é o ponto onde as classes obtiveram a maior distância de separação percentual, sendo que a performance de um modelo por KS é aceitável a partir de 20%.

Tabela 1: Resultados dos modelos nos cenários de modelagem de triagens 1 e 2

	NBSVM					FastText					Distilbert				
	Precisão	Revocação	F1	Acurácia	KS	Precisão	Revocação	F1	Acurácia	KS	Precisão	Revocação	F1	Acurácia	KS
Mod1	0.47	0.98	0.63	0.48	0.18	0.51	0.9	0.65	0.55	0.22	0.5	0.91	0.65	0.54	0.14
Mod2	0.64	0.85	0.73	0.71	0.43	0.6	0.9	0.72	0.68	0.43	0.62	0.87	0.72	0.7	0.43

Para implantação de um modelo em produção um requisito não-funcional importante a ser considerado é o tempo de processamento do modelo escolhido. Esse requisito pode variar bastante de acordo com a arquitetura computacional empregada, porém nos testes realizados o Distilbert levou em média 100 vezes mais tempo para execução que os demais. Dependendo do cenário na empresa esse custo elevado de processamento pode inviabilizar a implantação.

Analisando apenas os resultados das métricas consideradas, no cenário de modelagem de triagem 1 (Mod1) o modelo FastText apresentou resultados similares para as métricas de matriz de confusão, porém apresenta um resultado superior de discriminação quando olhamos o KS. Para o cenário de modelagem de triagem 2 (Mod2) todos os modelos apresentaram resultados equivalentes tanto em métricas de matriz de confusão quanto em KS. É importante destacar que resultados

de Mod2 são substancialmente mais relevantes que os resultados de Mod1, chegando a ter 29 pontos a mais de KS no Distilbert. A escolha do melhor modelo de predição depende da estratégia de atuação da operação de atendimento, porém olhando o KS dos modelos em Mod2 é possível concluir que a aplicação de MT com AM nessa abordagem de atendimento é factível. A adoção de modelos treinados em Mod1 não é conclusiva e depende bastante da capacidade de atendimento em nível superior e do apetite de risco da empresa.

4 Conclusões

Os resultados mostraram que a capacidade de compreensão semântica do Distilbert não agregou melhores resultados ao problema. Os modelos NBSVM e FastText depois de treinados executam em média 100 vezes mais rápido que o Distilbert e isso é um fator importante na escolha do modelo. A utilização apenas da narrativa da reclamação do consumidor apresentou resultado pouco relevante na discriminação da resolutividade, porém quando agregado a narrativa de atendimento da empresa os modelos performaram consideravelmente bem e podem ser usadas em uma estratégia de atendimento. A escolha de um bom modelo de predição depende da estratégia de atuação da operação de atendimento, sendo que a capacidade de ordenação do risco torna-se preponderante nessa escolha.

Referências

- [1] AGGARWAL, C. C. Machine learning for text. [S.l.]: Springer, 2018.
- [2] ALEXA. @Alexa an amazon.com company. [S.l.], 2021. <https://www.alexa.com/topsites/countries/BR>. Acessado em: 2021-04-13.
- [3] BOJANOWSKI, P.; GRAVE, E.; JOULIN, A.; MIKOLOV, T. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, MIT Press, v. 5, p. 135–146, 2017.
- [4] KAUFMANN-KOHLER, G.; SCHULTZ, T. Online dispute resolution: challenges for contemporary justice. [S.l.]: Kluwer Law International BV, 2004.
- [5] LI, B.; ZHAO, Z.; LIU, T.; WANG, P.; DU, X. Weighted neural bag-of-n-grams model: New baselines for text classification. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. [S.l.: s.n.], 2016. p. 1591–1600.
- [6] REZENDE, S. O. Sistemas Inteligentes: Fundamentos e Aplicações. Barueri, SP: Editora Manole Ltda, 2003. ISBN 8520416837.
- [7] SANH, Victor et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108, 2019