
**APRENDIZADO DE MÁQUINA APLICADO NA CLASSIFICAÇÃO DE RESOLUTIVIDADE
DE RECLAMAÇÕES DE SERVIÇOS FINANCEIROS**

LUIZ HENRIQUE ROWAN PEIXOTO
RICARDO MARCONDES MARCACINI
SOLANGE OLIVEIRA REZENDE

Nº 431

RELATÓRIOS TÉCNICOS



São Carlos – SP
Dez./2020

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

Luiz Henrique Rowan Peixoto¹, Ricardo Marcondes Marcacini¹, Solange Oliveira Rezende¹

¹ICMC - Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
Av. Trab. São Carlense, 400 - Centro, São Carlos - SP – Brazil

henriqueroan@usp.br, {ricardo.marcacini, solange}@icmc.usp.br

Resumo. *Este trabalho propõe a utilização de técnicas de MT (Mineração de textos) em conjunto com AM (aprendizado de máquina) na classificação de resolução de reclamações de clientes insatisfeitos com serviços financeiros dos principais bancos brasileiros. A abordagem proposta é avaliada por meio de três métodos de classificação textual: BoW com NBSVM, FastText com MLP e BERT. Os resultados mostraram que os métodos podem gerar valor em sistemas de apoio a tomada de decisão na área operacional das empresas e podem auxiliar as mesmas no atendimento de seus clientes elevando o nível de satisfação dos consumidores, reduzindo o risco de deterioração de imagem.*

1. Introdução

O advento da internet, a proliferação da digitalização, das mídias sociais e dos dispositivos móveis na sociedade alavancaram nos últimos anos a amplitude e o impacto da expressão de opinião individual. As opiniões caracterizam-se por ser uma manifestação de uma visualização pessoal, representando o estado de espírito e a atitude de um indivíduo ou de um grupo em relação a uma determinada realidade. Por meio da internet, as opiniões de outros consumidores oferecem a possibilidade de avaliação de um produto ou serviço baseado nas experiências da própria comunidade de consumidores, o que é conceituado como *electronic word-of-mouth* (eWOM) [Hennig-Thurau et al. 2004] ou no popular "boca-a-boca eletrônico".

Dentre as opiniões manifestadas na internet para fins de consumo, destacam-se as reclamações de consumidores, que são resultados 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, 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.

Dado a insatisfação do consumidor, a reclamação é uma ação garantida pelo Código de Defesa do Consumidor (CDC) por meio de um canal de atendimento. Os canais de atendimento para reclamações de consumo podem ser agrupadas em algumas categorias. O Serviço de Atendimento ao Consumidor (SAC) é um canal de atendimento obrigatório por lei, disponibilizado pelo prestador do serviço. Geralmente é exigido que o cliente apresente sua reclamação no SAC antes de qualquer outro canal atendimento. As agências reguladoras são entidades da administração pública responsáveis pela regulação de setores específicos. Estas disponibilizam canais de atendimento ao consumidor para setores como transportes terrestre, aquaviário e aéreo, gestão de recursos naturais, produção

audiovisual, planos de saúde e vigilância sanitária [Guerra and Salinas 2020]. O Programa de Proteção e Defesa do Consumidor (PROCON) é um órgão administrativo extrajudicial do Poder Executivo, que tem como funções o acompanhamento e fiscalização das relações ocorridas entre fornecedores e consumidores garantindo a correta execução do CDC, evitando a atuação do Poder Judiciário [Oliveira et al. 2016]. As *online dispute resolution* (ODR) são meios eletrônicos de resolução de conflitos que podem ter funções jurisdicionais e extrajurisdicionais [Guerra and Salinas 2020], e estes podem ser geridos por agências reguladoras, PROCON ou por empresas do setor privado.

No Brasil, as plataformas ODR têm ganhado relevância e influência [Sousa et al. 2020], e entre as plataformas com maior popularidade destaca-se o [ReclameAQUI 2020] que entre 20/06/2020 e 20/09/2020 foi o 11º site brasileiro mais acessado, onde os usuários fizeram em média 2,9 visitas em páginas distintas com um tempo médio de 3 minutos e 44 segundos nas interações [Alexa 2020]. O ReclameAqui possui uma base de 15 milhões de consumidores e 120 mil empresas cadastradas com uma média de 42 milhões de visualizações mensais. A plataforma do ReclameAQUI não divulga a quantidade de reclamações que o seu serviço recebe ao todo, porém divulga por empresa, e levantando a quantidade de reclamações dos 5 principais bancos do Brasil nos últimos 12 meses chega-se em aproximadamente 237 mil reclamações recebidas.

1.1. Objetivos

Considerando que a reclamação é uma ação garantida por entidades de proteção ao consumidor, a digitalização constante da sociedade e a popularização das plataformas de ODR, resultam em aumento constante das reclamações de clientes em ambientes digitais. Para as empresas analisar, direcionar e tratar essas reclamações manualmente demanda recursos de atendimento e o aumento constante torna este processo insustentável. A escalada das reclamações em ambientes digitais não-tratadas corretamente impactam a imagem das empresas e gera uma necessidade cada vez mais latente de capturar, processar, analisar e gerar conhecimento usando técnicas de Mineração de Textos (MT) e Aprendizado de Máquina (AM) ([Joung et al. 2019]; [Sousa et al. 2020]).

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 análise de sentimentos, sumarização de reclamações, tratamento especial para problemas complexos, identificação de problemas generalizados, desenvolvimento de novos produtos, entre outros ([Krishna et al. 2019]; [Gonçalves 2016]; [Carosia et al. 2020]); [Joung et al. 2019]; [Sousa et al. 2020]; [Hu and Liu 2004a]; [Hu and Liu 2004b]; [Yang et al. 2019]).

Este trabalho tem como objetivo avaliar o uso técnicas de MT com AM na predição preemptiva de resolução de reclamações de clientes. Estas técnicas podem ser empregadas em sistemas de apoio a tomada de decisão da área operacional das empresas e podem auxiliar as mesmas no atendimento de seus clientes elevando o nível de satisfação dos consumidores, reduzindo o risco de deterioração da reputação da empresa perante a sociedade.

2. Fundamentos e Trabalhos Relacionados

2.1. 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. Essa extração é feita por meio de uma variedade de algoritmos estatísticos e de AM gerando aplicações práticas a partir da informação gerada ([Rezende 2003]; [Aggarwal 2018]). Numerosas aplicações já existem no cotidiano da sociedade, entre elas os motores de busca da internet, filtros de spam, categorização e organização de documentos, sistemas de recomendação baseado em conteúdo, análise de opiniões e reclamações para tomada de decisão, sendo o ultimo o foco deste trabalho.

Este trabalho é baseado no processo de MT definido por [Rezende 2003] (Figura 1) 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. Processo de Mineração de Textos. Fonte: [Rezende 2003]



1. Identificação do Problema: com a ajuda de especialistas do domínio, nesta etapa ocorre a definição do problema e seu escopo que serão tratado com textos. Questões a respeito dos dados devem ser esclarecidos, como por exemplo, quais os conjuntos de dados disponíveis, seus atributos, sua periodicidade, qualidade, completude, privacidade dos dados e direitos de uso. Deve ser estabelecido o escopo de técnicas que podem ser empregadas baseando-se no critério de interpretabilidade, pois técnicas mais avançadas podem ter baixa interpretabilidade. E por ultimo, quais as métricas serão utilizadas para avaliar a qualidade da solução.
2. Pré-Processamento: esta etapa diferencia a MT da mineração de dados tradicional, pois agrega um fator de complexidade devido a necessidade de estruturação dos textos. Podem ser empregados nesta etapa o tratamento de ruídos, remoção de *stopwords*, redução de dimensionalidade com *stemming* ou *lemmatization*, remoção de caracteres inválidos, tratamento de dados pessoais ou sensíveis e normalização de frequência dos termos [Aggarwal 2018]. Após os tratamentos é necessário utilizar uma representação que será utilizada pelo algoritmo para extração de padrões, visto que não é possível apresentar o texto puro a estes algoritmos. A qualidade dos resultados dos algoritmos de extração de padrões é proporcional a qualidade da representação de textos e quais características eles

conseguem capturar [Rossi 2016]. Neste trabalho são apresentados duas formas muito comuns de representação de textos: modelo espaço vetorial e modelo de *embeddings*.

3. Extração de Padrões: nesta etapa a representação de texto é apresentada aos algoritmos de extração de padrões, por exemplo os algoritmos de AM. Esses algoritmos podem agrupar textos com mesmo tema, classificar textos baseando em rótulos previamente definidos, extrair regras de associação entre os textos, sumarização de textos e também executar regressão que diferencia a classificação pois a variável resposta é contínua [Aggarwal 2018].
4. Pós-Processamento: os resultados obtidos na extração de padrões são avaliados por meio de métricas definidas na etapa de identificação do problema. A métrica varia bastante do tipo de aplicação empregada seja ela um agrupamento, associação, regressão ou classificação. Para problemas de classificação as métricas mais comuns são derivadas da matriz de confusão: acurácia, precisão, revocação, f1-score, ROC, AUC, entre outros.
5. Utilização do Conhecimento: a utilização do conhecimento nesta etapa pode ser válido e útil se as etapas anteriores forem devidamente aplicadas. Como neste trabalho, o conhecimento extraído é utilizado em processos de tomada de decisão e devem ser constantemente monitorados para evitar que o algoritmo empregado perca a assertividade da métrica testada no pós-processamento. Nesse caso, o processo pode voltar para extração de padrões ou pré-processamento. Se houver mudança de escopo em relação ao objetivo, o processo deve retornar para a etapa de identificação do problema.

2.2. Representação de Textos

A representação de textos conhecida como *bag-of-words* (BoW) é considerada uma técnica inicial para representação de textos em trabalhos de mineração de textos. Tem como método verificar a presença ou não dos termos nos textos, podendo considerar também a quantidade. BoW apresenta bom custo benefício para representação de conteúdo dos documentos para utilização de algoritmos de AM [Rossi 2016]. Entre as suas limitações destacam-se a alta dimensionalidade e alta esparsidade, além de não considerar semântica e contextos dos termos. A representação usando a abordagem BoW pode utilizar uma estrutura tabular onde cada li pode conter um termo representado por uma coluna $t(i)$. O valor $a(i,j)$, do termo $t(j)$ da linha $l(i)$ pode ser calculado utilizando diferentes medidas, como: indicador booleano, contador de termos ou *term frequency-inverse document frequency* (TFIDF).

Para contornar a ausência de semântica dos termos e também dos documentos gerados por representações simples como BoW, novos modelos de linguagem conhecidos como *Word Embedding Models* foram construídos nos últimos anos. Esses modelos de linguagem geralmente são construídas usando redes neurais conhecidas como *encoders*, que possuem como entrada uma camada de tamanho V , onde V é o domínio de termos possíveis, uma camada intermediária de tamanho N , onde N é a dimensão das *embeddings* construídas e uma camada de saída com dimensão V novamente. O foco deste modelo é a utilização da camada intermediária de dimensão N , pois a mesma foi capaz de comprimir o espaço de termos altamente dimensional em um espaço vetorial pré-definido. Entre as principais implementações de *embeddings* publicas destacam-se a Word2Vec (Google)

[Mikolov et al. 2013] que foi desenvolvido por meio da base de dados do Google News Dataset utilizando 100 bilhões de tokens, o Glove (Stanford) [Pennington et al. 2014] que utilizou 6 bilhões de tokens da wikipédia e o Fasttext (Facebook) [Bojanowski et al. 2017] que utilizou 16 bilhões de tokens da wikipédia e notícias.

2.3. Aprendizado de Máquina e Modelos de Linguagem

As variações de *Naive Bayes* (NB) e *Support Vector Machine* (SVM) são frequentemente utilizadas como modelos *baseline* para classificação de textos, porém suas performances quando usado em separado, podem variar consideravelmente dependendo da abordagem e da base de dados utilizada [Wang and Manning 2012]. A combinação das duas técnicas (NBSVM) demonstra ser muito útil em representações esparsas como BoW, apresentando resultados robustos em uma grande variedade de tarefas de classificação de texto [Li et al. 2016]. O NBSVM proposto por [Wang and Manning 2012], usa a abordagem N-gram que consegue capturar a ordem das palavras em contextos curtos para que o NB atribua maiores pesos a palavras mais importantes por meio de logaritmo da razão entre classes. A partir das variáveis NB é aplicado SVM com *kernel* liblinear que possui mais flexibilidade na escolha de penalidades e funções de perda, escala melhor para um grande número de amostras e suporta entrada densa e esparsa. As classificações multiclasse são tratadas em um esquema um-contra-o-resto.

Redes Neurais Artificiais (RNA) é uma classe de métodos de AM inspirados na biologia dos neurônios. Propostas de aprendizado usando métodos inspirados nos neurônios datam dos anos 40, porém o Perceptron de Rosenblatt ([Rosenblatt 1958]) é considerado a proposta mais robusta e famosa no início da história das RNA. As RNA tem sido amplamente aplicadas em problemas de classificação de textos, principalmente por meio dos conceitos que compreendem o aprendizado profundo, onde as redes possuem múltiplas camadas, possuem formatos diferenciados de acordo com o tipo de aprendizado e podem transferir conhecimento entre elas.

O *Multilayer Perceptron* (MLP) é o tipo mais antigo de RNA e se enquadra na categoria de *feedforward*, onde o fluxo de informações que geram o aprendizado ocorrem do início para o fim da rede sem gerar ciclos. Uma MLP consiste de pelo menos 3 camadas de neurônios, sendo uma camada de entrada, uma camada oculta e uma camada de saída. Exceto para os nós de entrada que tem função apenas de propagar os dados, os neurônios utilizam uma função de ativação não-linear que recebe como entrada os valores da camada anterior ponderados por um vetor de pesos exclusivo de cada neurônio. O MLP utiliza como técnica de aprendizado, o *backpropagation* que a partir do erro na saída da RNA ajusta os pesos dos vetores w das camadas de neurônios utilizando o algoritmo Gradiente Descendente e uma taxa de aprendizado, minimizando assim o erro. Esses ajustes ocorrem partindo da última camada até a camada de entrada, onde a retropropagação do erro ajusta os pesos das camadas de neurônios.

Modelos de linguagem tem como conceito base uma distribuição de probabilidade sobre sequências de palavras, associando uma probabilidade para cada sentença em um idioma. Esses modelos mostram-se mais efetivo e amplamente utilizado para tarefas de processamento de linguagem natural [Devlin et al. 2018]. Esses modelos possuem diversas aplicações como reconhecimento de fala (*speech recognition*), tradução (*machine translation*), marcação de parte da fala (*Part of Speech (PoS) tagging*), análise sintática

(*parsing*), reconhecimento óptico de caracteres (*Optical Character Recognition*), reconhecimento de caligrafia (*handwriting recognition*), recuperação de informação (*information retrieval*), entre outras.

Diferentemente de outros modelos de linguagem onde as palavras são apresentadas sequencialmente, os *Transformers* permitem que todas as palavras de uma mesma sequência sejam apresentadas simultaneamente e as *embeddings* são construídas simultaneamente no domínio da sentença. Essa diferença permite paralelização e pode alcançar um novo estado da arte em qualidade da em aplicações de tradução depois de ser treinado por apenas doze horas em oito GPUs P100 [Vaswani et al. 2017]. Outra vantagem dos *Transformers* é a implementação do mecanismo de atenção no *decoder*, onde cada palavra possui um vetor de valores que indicam quais das outras palavras da sentença são mais relevantes para a palavra que está sendo processada. Uma característica importante nos *Transformers* é que uma mesma palavra em diferentes posições da sentença podem ter diferentes *embeddings*. A limitação dos *Transformers* é ser unidirecional restringindo assim a escolha de arquiteturas que podem ser usadas durante o pré-treinamento.

O BERT (*Bidirectional Encoder Representations from Transformers*) é uma evolução dos transformers, onde sua principal contribuição está no processamento bidirecional dos *Transformers*. Esse método permite ao BERT ter vetores de atenção da esquerda para direita e da direita para esquerda para cada palavra, aproximando-se de como o cérebro humano constrói e interpreta as sentenças. O BERT está sediado no conceito de AM conhecido como *Transfer learning*, que é a reutilização de um modelo pré-treinado com aplicação em um novo problema. Esta abordagem já é muito aplicada em tarefas com imagens e é muito útil pois geralmente são treinados em bases de dados grandes e complexas, durante dias ou semanas, usando matrizes de recursos de processamento não disponível para a maior parte comunidade científica. O BERT é um *framework* com 2 passos, *pre-training* e *fine-tuning*. No *pre-training* o modelo é treinado em dados não categorizados em diferentes tarefas. No *fine-tuning* o BERT é inicializado com os parâmetros do *pre-training* e apresentado para os dados categorizados, onde suas últimas camadas de RNA *Feed-forward* com ativação *Softmax* são ajustadas [Devlin et al. 2018].

2.4. Trabalhos Relacionados

Em [Yang et al. 2019], os autores tem como objetivo detectar a intenção do cliente de escalar sua reclamação ao órgão regulador. Como contexto, o trabalho usa dados de texto de chat da chinesa JD.com que é o principal e-commerce no país. A utilização do conhecimento obtido foi testado em uma ferramenta que monitora chats de reclamações de clientes em tempo real e direciona os clientes propensos a escalar sua reclamação para uma área especializada para tratamento de problemas complexos. Foi utilizado como proposta um modelo híbrido de redes neurais com variáveis projetadas manualmente.

Em [Sezgen et al. 2019], os autores definem um processo baseado em *Latent Semantic Analysis* LSA para minerar mais de 5 mil avaliações da plataforma TripAdvisor e diferenciar os fatores de (in)satisfação de clientes do setor aéreo. Os resultados foram segmentados por classe de cabine e modelo de negócios da companhia aérea.

Em [Liu et al. 2019], os autores propõem um método para detecção de expectativas dos clientes, pois quando suas expectativas são atendidas a satisfação aumenta. As expectativas foram categorizadas em resolução de problemas, compensação finan-

ceira e compensação psicológica. Os dados foram obtidos por meio de uma plataforma de reclamações automobilísticas. Foram testados 8 tipos de classificadores de AM em cenários usando BoW com TF-IDF e *embeddings* Word2vec e GloVe.

Em [Iglesias 2020], os autores exploram a contribuição de MT na mineração de dados de SAC de uma empresa do setor automobilístico. Um processo de mineração é apresentado desde os dados brutos até na extração de conhecimento. Os algoritmos *Naive Bayes*, *Sequential minimal optimization* (SMO) e árvores de decisão foram empregados na mensuração da (in)satisfação dos consumidores.

Em [Krishna et al. 2019], os autores aplicaram técnicas de AM para classificação de sentimentos em dados de reclamações de 4 bancos indianos. Análises foram feitas com e sem técnicas de pré-processamento de textos. As representações textos avaliadas foram *document-term matrix* (DTM) com TF-IDF, *Word2Vec* e *Linguistic Inquiry and Word Count* (LIWC). Os resultados das abordagens de representação LIWC com *Random Forest* e *Naive Bayes* apresentaram os melhores resultados.

Em [Joung et al. 2019], os autores exploram a aplicação de MT em dados de reclamações de clientes de empresas de ar de condicionado coreanas para identificar as verdadeiras necessidades dos clientes usando o método *Outcome-Driven Innovation* (ODI) em combinação com análise por agrupamentos hierárquicos. A partir das necessidades levantadas, o objetivo é orientar o desenvolvimento de produtos voltados para o mercado.

Em [Sousa et al. 2020], os autores usam dados do ReclameAQUI com técnicas de MT para encontrar dentro do conjunto de reclamações grandes conjuntos de problemas no setor de telecomunicações, onde os resultados são apresentados por empresa e região geográfica. A principal contribuição do trabalho está na proposta de uma abordagem que auxilia no planejamento estratégico das empresas, levando em consideração situações reportadas pelos consumidores.

Nos trabalhos clássicos de [Hu and Liu 2004a] e [Hu and Liu 2004b], os autores exploram a extração de características relevantes nas avaliações de produtos adquiridos por e-commerce além de sumarizar as sentenças e classificar os sentimentos dos clientes.

Tabela 1. Comparativo entre trabalhos

Trabalho	Consumidor		Representação		Extração de padrões			Utilização do conhecimento		
	Avaliação	Reclamação	Modelo Espaço-Vetorial	Embeddings	AM	Modelos de Linguagem	Outros	Análise de Sentimentos	Resolução de Conflitos	Necessidades de Clientes
[Yang et al. 2019]		X	X	X	X				X	
[Sezgen et al. 2019]	X	X	X				X			X
[Liu et al. 2019]		X	X	X	X				X	X
[Iglesias 2020]	X	X	X		X		X	X		
[Krishna et al. 2019]		X	X	X	X			X		
[Joung et al. 2019]		X	X		X					X
[Sousa et al. 2020]		X	X				X			X
[Hu and Liu 2004a]	X		X				X	X		
[Hu and Liu 2004b]	X		X				X	X		

Na Tabela 1 é apresentado um resumo dos trabalhos citados. Na coluna Consumidor é feita uma divisão pelo tipo de dado explorado: avaliação e/ou reclamação de produtos ou serviços, sendo que os trabalhos que utilizam ambos geralmente separam eles. A coluna Representação exibe qual técnica foi empregada para representação dos textos. Na coluna Extração de padrões mostra que tipo de técnicas foram empregadas.

A coluna outros são geralmente técnicas de análises algébricas que não se enquadram em AM ou Aprendizado profundo. Na utilização do conhecimento a coluna análise de sentimentos indica que o trabalho focou apenas nesse tema sem abordar a resolução de conflitos ou na extração de necessidades dos clientes.

Os trabalhos relacionados possuem aplicações praticas com dados de satisfação ou reclamações de consumidores e os mesmos não exploraram o uso de modelos de linguagem, tais como BERT. O presente trabalho utiliza modelos de linguagem baseados em BERT em comparação com um cenário base utilizando NBSVM e um cenário intermediário utilizando *embedding* livre de contexto FastText. Esses cenários comparativos visam entender a aplicabilidade de modelos simples e complexos nos cenários reais de previsão de resolutividade e satisfação de reclamações de clientes, simulando a aplicação dos modelos na esteira de atendimento das empresas.

3. Abordagem proposta e resultados

3.1. Identificação do problema

Diante do contexto apresentado, este trabalho emprega técnicas de MT com AM em dados não-estruturados de reclamações coletados na plataforma ReclameAQUI de consumidores de serviços dos maiores bancos do sistema financeiro brasileiro. Para obter uma maior variabilidade de reclamações, porém mantendo o mesmo domínio, foram obtidas aleatoriamente 25000 reclamações finalizadas em 2020, sendo 5000 de cada um dos 5 maiores bancos. As reclamações finalizadas possuem indicador de resolução, avaliações preenchidas pelos clientes e não permitem mais réplicas por parte dos consumidores/empresas.

Essa abordagem utiliza como variável resposta, o indicador que mostra se a reclamação foi resolvida ou não. Na base coletada, esta variável tem 35,5% das reclamações com valor igual a não-resolvida. A avaliação comparativa utiliza as métricas de acurácia, precisão, revocação e f1-score. Modelos de representação de textos e técnicas de AM foram testadas em 3 cenários de complexidade de técnicas, sendo o primeiro um *baseline*, o segundo um intermediário e o terceiro um método estado da arte:

1. BoW com NBSVM
2. *Embeddings FastText* com previsor MLP
3. BERT (*embeddings* e previsor)

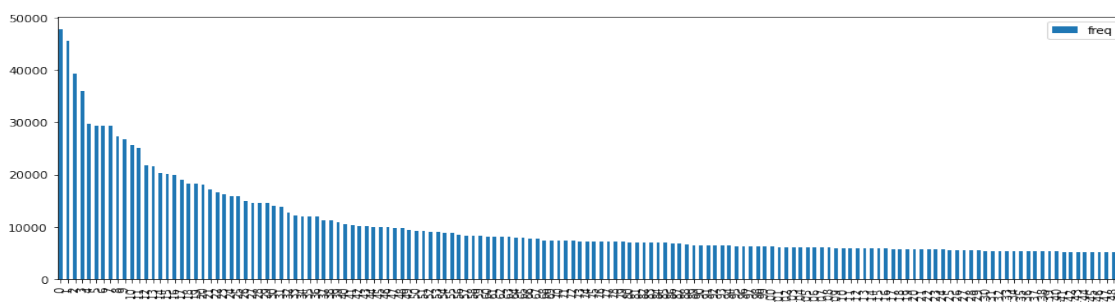
Os experimentos são construídos com linguagem de programação Python utilizando a biblioteca ktrain que empacota e simplifica a utilização da biblioteca Keras que implementa entre outras, os algoritmos testados neste trabalho. O processamento foi feito por meio da plataforma Google Colab, que disponibiliza gratuitamente recursos computacionais para processamento de projetos de ciência de dados.

3.2. Pré-processamento

Os métodos de pré-processamento empregados fazem parte do pacote ktrain. Foram empregados 2 tipos de pré-processamento, sendo o padrão para os cenários 1 e 2 e um pré-processamento especial para o BERT. O pré-processamento empregado remove as *stopwords* dos textos e cria um vetor para cada documento, onde cada posição indica o id da palavra naquela posição. Os vetores foram definidos com tamanho fixo de 128. Foi definido também a utilização de 1-gram, ou seja, as palavras foram consideradas sozinhas. Foi

utilizado 10% dos dados para testes. O algoritmo considerou e gerou ID's para as 50.000 palavras mais relevantes do corpus. O base de dados contém 57.424 palavras distintas e cada reclamação possui em média 241 palavras, sendo os percentis 95 e 99 iguais a 511 e 767 palavras respectivamente. Na figura 2 é mostrado a frequência das 150 palavras mais comuns no texto.

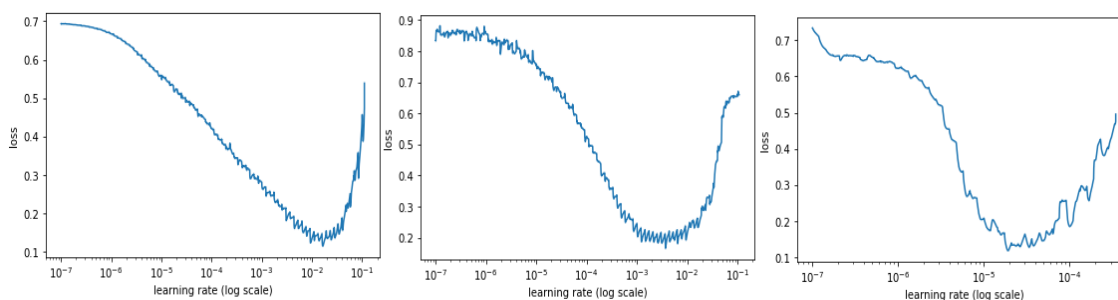
Figura 2. Frequência das 150 palavras mais comuns



3.3. Extração de padrões

Um passo importante na etapa de etapa de aprendizagem para os modelos implementados no ktrain é a escolha da taxa de aprendizagem. As taxas de aprendizagem foram obtidas por meio da função *find learning rate* que executa o treinamento por até 100 épocas ou até a convergência, o que ocorrer primeiro, para sugerir uma taxa de aprendizagem adequada. A taxa de aprendizagem para cada cenário é escolhida no gráfico como sendo um valor próximo do vale, porém antes do mínimo. A evolução do estimador durante as épocas é mostrado na Figura 3 para os cenários 1, 2 e 3 respectivamente.

Figura 3. Evolução da taxa de aprendizagem para os cenários 1, 2 e 3



Para o treinamento dos modelos as taxas de aprendizagem usadas para os cenários 1, 2 e 3 foram respectivamente: 0.005, 0.005 e 0.00001. A função *autofit* foi empregada no treinamento dos cenários 1 e 2, pois tem como vantagens alguns ajustes de taxa durante o aprendizado. Para o BERT foi utilizado o *fit_onecycle* que é um método de custo computacional menor, indicado para métodos pesados como o BERT.

3.4. Pós-processamento

Nessa etapa são avaliadas métricas definidas na etapa de identificação do problema. Os resultados na Tabela 2 mostram a acurácia geral e as demais métricas para a classe de interesse que é a identificação de reclamação NÃO-resolvida. É possível notar que houve

um empate entre os classificadores NBSVM e MLP na métrica acurácia, ambos com 0.81. Na métrica precisão o classificador NBSVM teve um resultado superior com 0.77. Nas métricas de revocação e f1-score, o classificador MLP obteve resultado superior com 0.8 e 0.75 respectivamente. Inesperadamente o BERT não conseguiu bons resultados neste problema e ainda serão investigadas mudanças como treinamento usando um pacote de *embeddings* pré-treinadas para o português e o uso do Distilbert.

Tabela 2. Resultados dos cenários 1, 2 e 3

Cenário	Acurácia	Precisão	Revocação	F1-Score
Bow com NBSVM	0.81	0.77	0.64	0.70
FastText + MLP	0.81	0.71	0.80	0.75
BERT	0.67	0.54	0.45	0.49

3.5. Utilização do conhecimento

Este trabalho é de caráter experimental, porém tem como objetivo avaliar metodologias que podem ser empregadas em sistemas de apoio a tomada de decisão na área operacional das empresas e podem auxiliar as mesmas no atendimento de seus clientes elevando o nível de satisfação dos consumidores, reduzindo o risco de deterioração da reputação da empresa perante a sociedade de consumidores.

4. Conclusões

O presente trabalho propôs a utilização de técnicas de MT com AM na classificação de resolutividade de reclamações de clientes insatisfeitos com serviços financeiros dos principais bancos brasileiros. A abordagem proposta foi avaliada em 3 cenários: BoW com NBSVM, FastText com MLP e BERT. O cenário de *embeddings* FastText com MLP apresentou melhores resultados neste problema que os demais cenários. Apesar da complexidade do BERT, o mesmo obteve resultados inferiores em todas as métricas avaliadas. Algumas mudanças na abordagem do BERT serão investigadas como treinamento usando um pacote de *embeddings* pré-treinadas para o português e o uso do Distilbert. Os resultados mostraram que os métodos podem gerar valor em sistemas de apoio a tomada de decisão na área operacional das empresas e podem auxiliar as mesmas no atendimento de seus clientes elevando o nível de satisfação dos consumidores, reduzindo o risco de imagem da empresa.

Referências

- Aggarwal, C. C. (2018). *Machine learning for text*. Springer.
- Alexa (2020). @Alexa an amazon.com company. <https://www.alexa.com/topsites/countries/BR>. Acessado em: 2020-09-20.
- Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Carosia, A., Coelho, G. P., and Silva, A. (2020). Analyzing the brazilian financial market through portuguese sentiment analysis in social media. *Applied Artificial Intelligence*, 34(1):1–19.

- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Gonçalves, C. d. A. (2016). *Análise de sentimentos em reclamações: uma aplicação no maior site de reclamações do Brasil*. PhD thesis.
- Guerra, S. and Salinas, N. S. C. (2020). Resolução eletrônica de conflitos em agências reguladoras. *Revista Direito GV*, 16(1).
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1):38–52.
- Hu, M. and Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Hu, M. and Liu, B. (2004b). Mining opinion features in customer reviews. In *AAAI*, volume 4, pages 755–760.
- Iglesias, B. L. S. (2020). A mineração de opinião em mídias sociais como ferramenta para medir a (in) satisfação do consumidor.
- Joung, J., Jung, K., Ko, S., and Kim, K. (2019). Customer complaints analysis using text mining and outcome-driven innovation method for market-oriented product development. *Sustainability*, 11(1):40.
- Krishna, G. J., Ravi, V., Reddy, B. V., Zaheeruddin, M., Jaiswal, H., Teja, P. S. R., and Gavval, R. (2019). Sentiment classification of indian banks’ customer complaints. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pages 429–434. IEEE.
- Li, B., Zhao, Z., Liu, T., Wang, P., and Du, X. (2016). 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*, pages 1591–1600.
- Liu, Y., Wan, Y., and Su, X. (2019). Identifying individual expectations in service recovery through natural language processing and machine learning. *Expert Systems with Applications*, 131:288–298.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Oliveira, L. V. C., Moreira, R. V., Ribeiro, L. M. T. B., Cabral, H. L. T. B., and de Souza, C. H. M. (2016). Procon virtual efetivando o direito do consumidor.
- Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- ReclameAQUI (2020). *ReclameAQUI*. <https://www.reclameaqui.com.br/institucional/>. Acessado em: 2020-10-24.
- Rezende, S. O. (2003). *Sistemas Inteligentes: Fundamentos e Aplicações*. Editora Manole Ltda, Barueri, SP.

- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- Rossi, R. G. (2016). *Classificação automática de textos por meio de aprendizado de máquina baseado em redes*. PhD thesis, Universidade de São Paulo.
- Sezgen, E., Mason, K. J., and Mayer, R. (2019). Voice of airline passenger: A text mining approach to understand customer satisfaction. *Journal of Air Transport Management*, 77:65–74.
- Sousa, G. N. d., Guimarães, I. d. S., Viana, J. A. N., Reinhold, O., Jacob Junior, A. F. L., and Lobato, F. M. F. (2020). Análise do setor de telecomunicação brasileiro: Uma visão sobre reclamações. *RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação*, (37):31–48.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Wang, S. I. and Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 90–94.
- Yang, W., Tan, L., Lu, C., Cui, A., Li, H., Chen, X., Xiong, K., Wang, M., Li, M., Pei, J., et al. (2019). Detecting customer complaint escalation with recurrent neural networks and manually-engineered features. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)*, pages 56–63.