

# **Avaliação das etapas de pré-processamento e de treinamento em algoritmos de classificação de textos no contexto da recuperação da informação**

**Lucas Marques Sathler Guimarães<sup>I</sup>**

**Magali Rezende Gouvêa Meireles<sup>II</sup>**

**Paulo Eduardo Maciel de Almeida<sup>III</sup>**

<sup>I</sup>*Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, MG, Brasil. Graduado em Sistemas de Informação.*

<sup>II</sup>*Doutorado em Ciência da Informação pela UFMG. Professora adjunta IV do Instituto de Ciências Exatas e Informática da PUC Minas. Professora colaboradora do Programa de Pós-graduação em Informática. Editora-chefe da Revista Abakós.*

<sup>III</sup>*Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil. Doutorado pela POLI / USP. Professor associado do Departamento de Computação.*

**<http://dx.doi.org/10.1590/1981-5344/3505>**

*A quantidade de dados não estruturados cresce com a popularização da Internet. Textos em linguagem natural representam um conjunto relevante e significativo para análise e produção de conhecimento. Este trabalho propõe uma análise quantitativa das etapas de pré-processamento e de treinamento de um classificador de textos, que utiliza os sentimentos expressos pelos usuários como atributo. Para realização dos experimentos, foram utilizadas Redes Neurais Artificiais, como algoritmo classificador, e textos provenientes dos sites Amazon, IMDB e Yelp. A base textual permite análise da expressão de sentimentos positivos e negativos dos usuários em avaliações de produtos e serviços em textos não estruturados. Foram realizados dois processos distintos de pré-processamento e diferentes treinamentos das Redes Neurais Artificiais para classificação do*

*conjunto textual. Os resultados confirmam, quantitativamente, a importância das etapas de pré-processamento e de treinamento do classificador, evidenciando a importância do vocabulário selecionado para a representação do texto e para a classificação. As técnicas de classificação disponíveis alcançam resultados satisfatórios. No entanto, mesmo utilizando-se dois processos distintos de pré-processamento e identificando-se o melhor processo de treinamento, não foi possível eliminar, totalmente, as dificuldades de aprendizado e compreensão do modelo para as classificações de sentimentos que envolviam características subjetivas da expressão do sentimento humano.*

**Palavras-chave:** *Classificação. Processamento da Linguagem Natural. Recuperação da Informação. Redes Neurais Artificiais. Treinamento.*

## **Evaluation of the preprocessing and training stages in text classification algorithms in the context of information retrieval**

*The amount of unstructured data grows with the popularization of the Internet. Texts in natural language represent a relevant and significant set for the analysis and production of knowledge. This work proposes a quantitative analysis of the preprocessing and training stages of a text classifier, which uses as an attribute the feelings expressed by the users. Artificial Neural Network, as a classifier algorithm, and texts from Amazon, IMDB and Yelp sites were used for the experiments. The database allows the analysis of the expression of positive and negative feelings of the users in evaluations of products and services in unstructured texts. Two distinct processes of preprocessing and different training of the Artificial Neural Networks were carried out to classify the textual set. The results quantitatively confirm the importance of the preprocessing and training stages of the classifier, highlighting the importance of the vocabulary selected for the text representation and classification. The available classification techniques achieve satisfactory results. However, even by using two distinct processes of*

*preprocessing and identifying the best training process, it was not possible to totally eliminate the learning difficulties and understanding of the model for the classifications of feelings that involved subjective characteristics of the expression of human feeling.*

**Keywords:** *Classification. Natural Language Processing. Information Retrieval. Artificial Neural Networks. Training.*

Recebido em 12.04.2018 Aceito em 14.02.2019

## 1 Introdução

O advento da Internet e sua aceitação pelas mais diversas camadas populacionais em todo o planeta ocasionou um aumento na quantidade de informação disponível, facilitando o processo de disseminação do conhecimento. Em contrapartida, realizar pesquisa em documentos textuais se tornou uma tarefa lenta e dispendiosa, muitas vezes até improdutivo, devido ao grande número de bases de dados disponíveis.

Uma alternativa utilizada rotineiramente pelo ser humano é a identificação de padrões para organizar a informação disponível. Reconhecer padrões é a habilidade de descrever semelhanças entre objetos, ações ou estímulos, processando aquele padrão e atribuindo-o a um grupo dentre um número pré-determinado de classes (HAYKIN, 2001). Diferentemente dos seres humanos, que possuem outras informações associadas às suas experiências de vida, que os permitem compreender e classificar um texto, o Processamento de Linguagem Natural (PLN) pelas máquinas está limitado aos padrões que foram processados para representar um determinado conceito.

Na tentativa de extrair informação de grandes bases textuais, as técnicas que envolvem aprendizado de máquina surgiram como um forte aliado das pesquisas que envolvem a Recuperação da Informação (RI). Segundo Allahyari e outros (2017), a extração de informação é a tarefa de automaticamente extrair informação estruturada a partir de textos não estruturados, que constituem a forma mais simples de dados gerada em diferentes contextos. Vários outros campos de pesquisa como estatística, linguística computacional e mineração de dados se complementam para absorver e gerar conhecimento a partir de bases textuais.

Mukherjee e outros (2017) afirmam que classificação é uma técnica de mineração de dados supervisionada que envolve a atribuição de um rótulo a um conjunto de objetos de entrada não rotulados. A classificação automática de documentos de texto utiliza diferentes técnicas de aprendizado de máquina aplicadas a coleções diferentes de documentos, procurando extrair padrões relevantes para organizar as informações do texto em um formato que facilite o processo de RI (ARANHA; PASSOS, 2006). Este processo encontra sua motivação na necessidade de separar a informação em blocos, permitindo que sejam manipulados e recuperados

de maneira mais intuitiva e simplificada. A complexidade do processo de classificação é, muitas vezes, gerada pela dificuldade para tratar a informação, tornando necessária, em alguns casos, a combinação de diversas técnicas de PLN. Ladeira (2012) ressalta a importância do entendimento das técnicas de PLN para serem aplicadas na Ciência da Informação, nos contextos de representação e de recuperação da informação. Uma das etapas fundamentais do processo de classificação de textos é o pré-processamento, na qual um conjunto de ações transformadoras é aplicado, para garantir que a informação em formato não estruturado (texto) se torne passível de compreensão pela máquina.

Muitos produtos e serviços são avaliados pelo consumidor por meio de textos que expressam seus sentimentos, dando-lhes uma conotação pessoal que representa sua percepção a respeito do que está sendo avaliado. Segundo Samal; Panda (2017), quando um usuário compartilha sua experiência, pode estar auxiliando outro usuário a tomar a decisão correta. O processamento dos textos possibilita que as avaliações sejam classificadas como positivas ou negativas, auxiliando o processo de recuperação desta informação. Para classificar a opinião do consumidor, Lara *et al.* (2017) avaliam, em seu trabalho, o comportamento do consumidor e concluem que, em sua avaliação, o resultado será a sua satisfação ou a sua insatisfação. Uma observação importante, apresentada pelos autores é que “o não atendimento aos requisitos de satisfação, normalmente, leva o consumidor ao estado de dissonância cognitiva”. Uma das hipóteses deste trabalho é que esse estado de desconforto psicológico pode contribuir com o fato de que, insatisfeito, o usuário se expresse de modo irônico ou sarcástico, criando textos que expressem exatamente o oposto do que sente, dificultando sua interpretação e sua posterior classificação.

Esta pesquisa propõe uma avaliação quantitativa das etapas de pré-processamento e de treinamento de um classificador no processo de classificação automática de dados não estruturados. As Redes Neurais Artificiais (RNA) do tipo *Multi-Layer Perceptron* (MLP) foram utilizadas como algoritmo classificador. Foram classificados textos escritos na língua inglesa, utilizando informações linguísticas contextuais para avaliação de sentimentos. Foram utilizados três conjuntos de dados extraídos de *sites* que oferecem produtos ou serviços que podem ser avaliados. Os conjuntos de dados são repositórios disponibilizados pela Universidade da Califórnia, em Irvine, e possuem mais de 3000 avaliações positivas e negativas de usuários sobre os produtos e serviços dos *sites*. A base se encontra balanceada, com percentual de avaliações positivas e negativas equilibradas, e os textos escolhidos foram avaliados pelos próprios autores (usuários) como positivos ou negativos. A base textual foi construída a partir das avaliações de usuários de três *sites*, Amazon, IMDB e Yelp, que oferecem serviços e produtos diferentes. A Amazon é um varejo virtual de diversos produtos. O IMDB é um portal onde usuários podem deixar avaliações sobre filmes e Yelp é um aglomerado de avaliações de estabelecimentos prestadores de serviços. A diferenciação dos *sites* e de

seus produtos é importante pois, tratando-se de linguagem natural, as palavras utilizadas para expressar-se, positivamente ou negativamente, sobre determinado assunto variam de contexto para contexto. Assim sendo, a primeira proposta é verificar se o processo de aprendizado supervisionado do classificador ao ser realizado com uma base de um tipo específico de avaliação de produto ou serviço é capaz de gerar uma classificação eficiente e precisa de uma avaliação que se referia a outro tipo de elemento. A segunda proposta do trabalho é avaliar dois tipos distintos de pré-processamento da base de dados utilizada no trabalho.

Este trabalho está organizado em cinco seções. A Seção 2 apresenta o contexto teórico, discutindo os sentimentos, o processo de pré-processamento e a seleção dos atributos de representação dos textos, as RNA e as métricas utilizadas para a avaliação dos algoritmos de classificação. A Seção 3 apresenta a metodologia, descrevendo a base textual, as ferramentas utilizadas e os experimentos realizados. As últimas seções finalizam o trabalho, apresentando e analisando os resultados obtidos e evidenciando as considerações finais e as propostas de trabalhos futuros.

## **2 Contexto teórico**

Esta seção conta com uma descrição do alicerce teórico utilizado para fundamentar este trabalho. Está dividida em quatro subseções que discutem os conceitos de sentimento, as implicações do processo de pré-processamento e de seleção de características do texto, as características das RNA e as métricas de avaliação utilizadas no processo de classificação.

### **2.1 Sentimentos**

Identificar sensações básicas é uma tarefa não trivial, principalmente quando se associa sensações básicas a comportamentos complexos (SCHERER, 2005). A sensação manifesta-se, então, como a impressão causada em um órgão receptor por meio de um estímulo. Assim, a sensação é puramente perceptual, podendo ser definida como uma atividade dos sentidos (RIES, 2004). Os sentimentos, por sua vez, são fenômenos muito mais complexos do que as sensações porque possuem sempre uma avaliação e uma conotação pessoal. Além disso, os sentimentos existem na relação entre acontecimentos específicos e um modelo amplo da experiência da pessoa. Diferente das sensações, os sentimentos não existem sem essa associação (SCHERER, 2005).

Scherer (2005) ainda constata que, para o sentimento existir, é necessária a manifestação de três componentes expressos pela emoção: representação do estímulo emocional, percepção consciente de estados do corpo e recuperação de significados associados a este estímulo. A conscientização é definida, então, como a condição que diferencia o sentimento. O mesmo autor também explica que as emoções, por sua vez, são mais complexas do que o parecer inicial e são, muitas vezes,

confundidas com os próprios sentimentos por possuírem destaque na experiência do indivíduo. O sentimento é um dos componentes da emoção.

Nesse trabalho, o objeto de análise (produto, serviço ou filme dos *sites* mencionados) gera uma experiência repleta de emoções que tem como componente comum a manifestação de um sentimento positivo ou negativo do usuário.

## 2.2 Pré-processamento e definição de atributos

Segundo Rezende e outros (2011), os conjuntos de documentos de texto são representados por um grande número de atributos ou características. Isto pode ser um fator dificultador na escolha dos atributos que melhor caracterizem o texto. Dentro de um grande número de atributos, um ou mais podem ser irrelevantes ou redundantes no processo de classificação do texto. Por outro lado, técnicas de classificação que utilizam métodos tradicionais necessitam de uma abordagem de recuperação de informação diferenciada para lidar com textos com poucas palavras como os *tweets*, sujeitos à limitação de 140 caracteres (MUKHERJEE *et al.*, 2017).

O objetivo do pré-processamento é extrair, de textos escritos em linguagem não estruturada, uma representação estruturada e manipulável por algoritmos de classificação que identifique o subconjunto mais significativo para a padronização da coleção de textos. Basicamente, a seleção reduz o conjunto de palavras que representará o documento no processo de classificação. São exemplos das etapas de seleção a remoção de palavras irrelevantes, como preposições ou artigos, a formatação dos textos e o cálculo de relevância dos termos para identificar os mais significativos. Por fim, os textos são representados em um formato estruturado que preserve as principais características dos dados (REZENDE *et al.*, 2011).

Para realizar a seleção, neste trabalho, utilizou-se técnicas de pré-processamento linguístico, *score* de relevância e coeficiente de correlação. O pré-processamento linguístico atua realizando a análise sintática dos textos e extraíndo informações a partir dela. A sintaxe representa as regras que regem a formação das frases de uma determinada língua. Determinar a estrutura sintática das frases é um elemento central na interpretação da linguagem natural, tornando a frase passível de ser formalmente analisada (ARANHA; PASSOS, 2006). Neste trabalho, foi utilizado o *parser* da ferramenta *RapidMiner* (MIERSWA *et al.*, 2006) para determinar a estrutura sintática das frases e, também, o método de extração de palavras representativas utilizando informação da Wikipedia (VIDAL *et al.*, 2012). Ambos validam o agrupamento de termos que compõem as frases.

Depois que os textos são pré-processados, é necessário estabelecer um conjunto de termos que melhor represente o assunto. Para fazê-lo, foram utilizados cálculos de relevância. O *score* de relevância (ER) foi

proposto e aplicado inicialmente no estudo de Wiener, Pederson e Weigend (WIENER, *et al.*, 1995) com base no peso de relevância de Salton e Buckley (SALTON *et al.*, 1974). Os autores propõem o *escore* de relevância de acordo com a frequência dos termos em uma dada categoria e também em outras categorias da seleção. Termos que se repetem em várias categorias obtêm valores inferiores, por não possuírem características discriminantes. O contrário acontece com termos que são específicos de poucas categorias, que ganham valores altos. O coeficiente de correlação (CC), desenvolvido por Ng e outros (1997) e utilizado em algumas pesquisas *para complementar* outras técnicas (SHAFIABADY *et al.*, 2016), indica o grau de correlação entre uma palavra e o documento. É utilizado levando em conta a quantidade total de documentos de uma coleção e a quantidade de vezes em que a palavra aparece e não aparece nestes documentos.

O extrator que utiliza termos da Wikipedia (VIDAL *et al.*, 2012) considera títulos de artigos da plataforma e a ocorrência de palavras-chave e categorias associadas com cada definição. Inicialmente concebido para o contexto da publicidade (VIDAL *et al.*, 2012), o extrator apresenta resultados satisfatórios e mais refinados para a etapa de pré-processamento. Foram utilizados, neste trabalho, o método tradicional que considera o ER e o CC e, também, o método da Wikipedia para serem analisadas possibilidades diferentes na etapa de pré-processamento.

## 2.3 Redes Neurais Artificiais

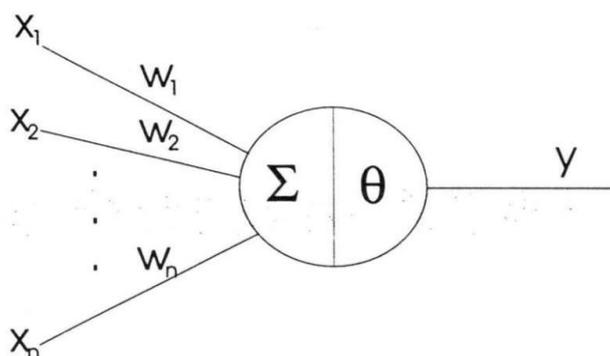
As RNA são abstrações projetadas para simular a estrutura e emular o funcionamento do cérebro humano. São sistemas de processamento, projetados para processar informações paralelamente e de maneira distribuída. São constituídos de unidades chamadas neurônios, que são responsáveis por armazenar informações experimentais e disponibilizá-las para uso nas competências de aprendizado e de generalização (HAYKIN, 2001).

Segundo Haykin (2001), podemos notar a semelhança entre as RNA e os neurônios biológicos em dois aspectos fundamentais: em ambas estruturas o conhecimento é adquirido pela rede por meio de um processo de aprendizagem e as conexões, pesos sinápticos, são responsáveis por armazenar esse conhecimento.

Um neurônio constitui uma unidade de processamento na RNA. Cada neurônio funciona gerando uma saída a partir da combinação de entradas recebidas de sinais externos ou de outros neurônios. As conexões entre dois neurônios, denominadas sinapses, possuem uma relevância (peso) e estas conexões armazenam o conhecimento da rede. A saída  $Y$  de um neurônio, na grande maioria dos modelos aplicados, é o resultado de uma função (função de ativação), aplicada à soma ponderada de suas entradas  $X_i$ , como apresentado na Figura 1. Ajustando-se continuamente a relevância das sinapses (valores dos pesos  $W_i$ ), a RNA assimila padrões e se torna capaz de generalizar. Este processo é o que gera saídas

consistentes para entradas que não foram previamente apresentadas à rede (SCHMIDHUBER, 2015).

Figura 1 – Neurônio de uma RNA



Fonte: BRAGA *et al.* (2007).

A RNA é estruturada em camadas. Os neurônios são agrupados nas camadas de acordo com sua funcionalidade. Funções semelhantes constituem uma mesma camada. Normalmente, as camadas podem ser classificadas como de entrada, saída ou intermediária. A primeira camada, de entrada, recebe as informações diretamente do meio externo (neste caso, documentos). Esta camada atua, simplesmente, propagando as informações das entradas para a camada seguinte, não realizando nenhuma transformação. A última camada, de saída, transmite as saídas para o ambiente. Essas saídas são as respostas desejadas (categoria do documento que foi processado). As camadas intermediárias são responsáveis por interligar as camadas da RNA. São estas camadas que recebem, como entradas, as saídas de outras camadas e geram saídas, como entradas, para outras camadas (HAYKIN, 2001).

Existem dois processos de aprendizagem de máquina: o supervisionado e o não supervisionado. O aprendizado supervisionado é definido quando o processamento desejado é especificado utilizando um conjunto de sequências ordenadas formado por uma ou mais entradas e as saídas desejadas correspondentes. Durante o processamento, uma comparação entre o valor desejado e o valor de saída acontece, gerando um valor de erro, que é utilizado para ajustar os pesos da rede. O aprendizado não supervisionado acontece em modelos de RNA que podem se auto organizar e produzir saídas satisfatórias, utilizando somente os dados de entrada, sem ter as saídas respectivas. A aprendizagem acontece pela identificação de similaridades nos dados de entrada.

As redes MLP são muito utilizadas em processos de classificação. Neste tipo de RNA, o sinal de saída de cada neurônio é o resultado da aplicação da função de ativação sobre a soma ponderada dos sinais de entrada. Ao contrário das redes de uma só camada (perceptron simples), que são utilizadas para classificar padrões que podem ser separados

linearmente, as redes MLP são utilizadas para classificar padrões que não sejam linearmente separáveis. O fluxo dos dados acontece da camada de entrada para a camada de saída (*feedforward*). Normalmente, possuem uma ou mais camadas ocultas, que se conectam com as entradas e as saídas.

Para calcular os pesos adequados à rede, é necessária uma regra de treinamento. Comumente, é utilizado o algoritmo de retropropagação de erros (algoritmo *backpropagation*) nos casos de aprendizagem supervisionada. Estes algoritmos têm obtido bons resultados nas soluções de diversos problemas (McCLELLAND; CLEEREMANS, 2009). A aprendizagem por retropropagação de erro consiste em dois passos que acontecem entre as diversas camadas de rede: a propagação e a retropropagação (HAYKIN, 2001). A propagação, ou passo para frente, é realizada quando um padrão de atividade (entrada) é aplicado aos neurônios da rede e o seu efeito se propaga, gerando um conjunto de saídas. A retropropagação, ou passo para trás, acontece quando os pesos sinápticos são reajustados de acordo com a regra de correção de erro. Este sinal de erro é propagado retroativamente pela rede, na direção contrária das conexões sinápticas. Estes ajustes são feitos para aproximar a resposta real da rede da resposta desejada. Muitos autores (NAIK *et al.*, 2015; MISHU; RAFIUDDIN, 2016; VINODHINE; CHANDRASEKARAN, 2016; PRASANNA; RAO, 2017) avaliaram o desempenho de algoritmos de classificação de textos que utilizavam RNA e ressaltaram o seu desempenho comparado aos métodos tradicionais.

## 2.4 Métricas de avaliação dos algoritmos de classificação

As métricas apresentadas nesta seção foram utilizadas para avaliar, quantitativamente, o algoritmo de classificação utilizado nos experimentos. A acurácia é formalmente definida como a porcentagem dos exemplos de teste que são corretamente classificados (SHALEV-SHWARTZ; BEN-DAVID, 2014). Entretanto, segundo o autor, não é possível considerar um classificador como eficiente, ou não, somente com a medida de acurácia. Para complementar a avaliação do algoritmo utilizado neste trabalho, foram escolhidos os conceitos de sensibilidade e precisão (BAEZA-YATES; RIBEIRO-NETO, 1999). Sensibilidade é o percentual de instâncias classificadas corretamente como positivas dentre todas as instâncias da base que são de fato positivas:

$$\text{Sensibilidade} = \frac{VP}{VP+FN}$$

Precisão é o percentual de instâncias classificadas corretamente como positivas dentre todas as instâncias que foram classificadas como positivas:

$$\text{Precisão} = \frac{VP}{VP+FP}$$

onde:

- VP significa Verdadeiro Positivo e representa as instâncias que são positivas e foram corretamente classificadas como positivas;
- FP significa Falso Positivo e representa as instâncias que são negativas e foram incorretamente classificadas como positivas;
- FN significa Falso Negativo e representa as instâncias que são positivas e foram incorretamente classificadas como negativas.

A quarta métrica utilizada é a *F-Measure*, que é a média harmônica entre a precisão e a sensibilidade e é um valor que varia entre 0,0 e 1,0, sendo 1,0 precisão e sensibilidade perfeitas e 0,0 a ausência total de precisão e sensibilidade (SHALEV-SHWARTZ; BEN-DAVID, 2014).

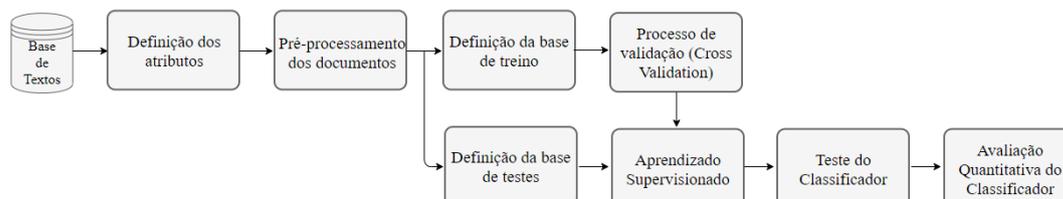
$$F\text{-Measure} = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$$

As quatro métricas são utilizadas, nesse trabalho, nas análises relativas aos dois experimentos propostos. A comparação das métricas definidas nessa seção mostra, objetivamente, a influência das etapas de pré-processamento e de treinamento do classificador no processo de classificação realizado. Considerando-se o que foi discutido para as métricas descritas, nesta seção, observa-se que, quanto mais próximo de 1,0 for o valor das métricas, melhor será a qualidade do classificador.

### 3 Metodologia

Esta seção contém uma descrição detalhada do corpo textual e das técnicas e ferramentas utilizadas nos processos de pré-processamento e de treinamento do classificador. A Figura 2 mostra o fluxo de processamento dos dados, que inicia-se na definição dos atributos da base textual que serão utilizados, perpassa o pré-processamento dos documentos, o processo de treinamento do classificador, a classificação e a posterior avaliação quantitativa do classificador.

Figura 2 – Fluxo de processamento dos textos



Fonte: Elaborada pelos Autores.

#### 3.1 Base textual

O *Center for Machine Learning and Intelligent Systems* (Centro de Aprendizado de Máquina e Sistemas Inteligentes), da *University of California, Irvine*, disponibiliza bases textuais que são utilizadas em experimentos. As pesquisas se concentram nos campos de aprendizado de máquina e sistemas inteligentes que abordam o problema fundamental de desenvolver técnicas e algoritmos que consigam lidar com a grande quantidade de dados digitais disponíveis no século XXI e utilizar estes dados de maneira criativa para resolver problemas do mundo real.

Foi utilizado para a análise um conjunto de avaliações de usuários submetidas aos *sites* Amazon, IMDB e Yelp. Os *sites* escolhidos fornecem um conjunto distinto (produtos, serviços e filmes) de elementos a serem avaliados. A base textual utilizada neste trabalho contém 988 avaliações originárias do *site* Amazon, 988 provindas do *site* Yelp e 956 do *site* IMDB. As avaliações foram apontadas como positivas ou negativas pelos próprios usuários que as submeteram. A Amazon oferece diversos produtos de varejo que são avaliados pelos consumidores, O IMDB é uma plataforma em que usuários podem submeter avaliações e opiniões sobre os filmes que assistiram e o Yelp permite que os usuários avaliem serviços como restaurantes, casas noturnas, lojas e outros. Porém, os elementos avaliados são diferentes o suficiente (produtos, serviços e filmes) para garantir uma diferença no vocabulário de expressão do usuário, justificando, assim, essa escolha.

O *corpus* empírico foi submetido ao analisador sintático com o objetivo de realizar a compreensão sintática das frases. As marcações foram levantadas pelo *parser*. Aplicando-se folhas de estilo aos arquivos gerados, foram obtidas categorias gramaticais e combinações gramaticais, tornando possível a criação dos vetores de apoio (descritos na seção de ferramentas). Para iniciar o processo de classificação, foi necessário estabelecer um conjunto de dados utilizado para o treinamento do classificador e outro para teste. Para averiguar a integridade dos resultados, garantindo uma representação realista dos conjuntos de treino e de teste, foi utilizado o *3-fold cross validation*, que propõe a divisão de um terço do conjunto para treinamento e dois terços para teste. As 2.932 avaliações disponíveis na base textual foram divididas de acordo com sua origem em 3 classes, uma para cada *site*. Cada subconjunto foi dividido por três novamente, sendo um terço utilizado para treinamento e dois terços para testes.

## 3.2 Ferramentas

Para a classificação dos textos, foi utilizada a ferramenta *RapidMiner* (MIERSWA *et al.*, 2006) previamente conhecida como YALE (*Yet Another Learning Environment*). A ferramenta conta com uma coleção de algoritmos e métodos relacionados ao aprendizado de máquina e é desenvolvida na linguagem Java. Estão disponíveis diversas implementações de algoritmos classificadores, incluindo instâncias

calibráveis das RNA MLP, que foram utilizadas como classificadores, neste trabalho, e ferramentas de pré-processamento de texto.

A ferramenta *WVTool*, desenvolvida por Michael Wurst na linguagem de programação Java, foi utilizada para fundamentar o pré-processamento dos textos. A ferramenta realiza uma correspondência entre os termos. Os seus valores numéricos foram obtidos por meio dos cálculos de relevância descritos anteriormente. Para realizar o pré-processamento, foram utilizadas as referências e as calibrações descritas na bibliografia do autor, que é um dos criadores do *RapidMiner* (RAPIDMINER, 2018).

Os  $n$  termos mais relevantes foram utilizados para construir um vetor local de cada subconjunto. Estes vetores foram combinados, formando um vetor geral. Este vetor geral serve como índice para os vetores de cada instância e suas posições representam a importância daquela instância no documento. Com isso, é possível construir um modelo de termos significativos que foram estabelecidos durante a fase de treino e, a partir desse modelo, avaliar a relevância do termo no processo de classificação do conjunto de testes, utilizando o coeficiente de confiança de cada termo.

Para o pré-processamento do segundo experimento foi utilizado o algoritmo desenvolvido por Vidal *et al.* (2012). O algoritmo leva em consideração um vocabulário extraído da Wikipedia para realizar o processamento textual.

### 3.3 Experimentos

O primeiro experimento teve o objetivo de verificar se a diferença entre os contextos de vocabulário era um fator significativo no processo de classificação. Para isso, o *corpus* empírico, que conta com 2.932 avaliações, foi dividido de acordo com a sua fonte. A Tabela 1 apresenta o número de avaliações positivas e negativas realizadas pelos usuários para cada um dos *sites*. Cada conjunto de avaliações foi pré-processado, gerando 3 subconjuntos diferentes. Em seguida, o processo de treinamento do classificador foi realizado de modo que cada subconjunto de treino foi aplicado ao seu próprio conjunto e aos outros dois conjuntos originários dos outros dois *sites*. Como exemplo, pode ser citado o processo treinado com um dos subconjuntos da Amazon, que classificou as avaliações provindas dos *sites* IMDB, Yelp e Amazon. O mesmo foi feito com os outros subconjuntos, gerando 9 resultados diferentes.

Tabela 1 – Número de avaliações para cada *site*

	<b>Amazon</b>	<b>Yelp</b>	<b>IMDB</b>
<b>Positivas</b>	527	527	517
<b>Negativas</b>	461	461	439
<b>Total</b>	988	988	956

Fonte: Dados da pesquisa.

O segundo experimento repetiu o primeiro experimento, com alteração do método de pré-processamento para utilizar o extrator da Wikipedia. Os dados foram pré-processados, como descrito na seção anterior, padronizados em caixa baixa e submetidos a validação (*3-fold cross validation*) para a criação do conjunto de treino. Em seguida, a base de testes foi classificada e, depois, foi gerado um novo modelo com os vetores de palavras para analisar os termos mais relevantes. Cada termo do vetor de palavras teve sua relevância calculada para comparação com os outros conjuntos.

Para os experimentos, foram utilizadas as redes MLP com algoritmo de treinamento *backpropagation*, disponíveis na ferramenta *RapidMiner*. Os parâmetros estabelecidos foram, inicialmente, calibrados em 0,9 para o valor de *momentum*, 0,1 para taxa de aprendizado e para a condição de parada, 300 épocas. A RNA possuía 10 neurônios na camada intermediária e 3 neurônios na camada de saída, que correspondiam aos subconjuntos. Estes valores foram utilizados nos estudos de Schmidhuber (2015). Para compreender se a alteração do contexto das avaliações afeta o treinamento do classificador, os resultados foram expostos em tabelas para serem comparados, levando em consideração as métricas acurácia, precisão, sensibilidade e *F-Measure*.

## 4 Apresentação e análise dos resultados

O pré-processamento da base de treinamento identificou entre 1500 e 1700 termos relevantes nos testes realizados com conjuntos de contextos diferentes e entre 800 e 950 termos relevantes em conjuntos do mesmo contexto. Nos casos de contextos distintos, muitas das palavras eram encontradas em somente uma ou duas instâncias e possuíam um *score* de relevância entre 0,1 e 0,2 e portanto oscilando entre 10 e 20% apenas da medida de relevância máxima, que é 1,0. Uma hipótese seria a de que estes vocabulários poderiam ser específicos de um determinado contexto e, por isso, só seriam encontrados naquele contexto.

### 4.1 Experimento 1: avaliação da etapa de treinamento do classificador

Cada conjunto de dados foi isolado e o classificador foi treinado e testado com os outros conjuntos. Foram realizados nove experimentos. Para exemplificar a metodologia utilizada, foram selecionados apenas três experimentos. Nesses experimentos, o algoritmo classificou as avaliações dos usuários da base da Amazon, tendo sido treinado com o próprio subconjunto da Amazon e, posteriormente, classificou as avaliações dos usuários dos *sites* Yelp e IMDB. As métricas de avaliação do processo de classificação citado são apresentadas nas Tabelas 2, 3 e 4. Nestas tabelas, a primeira coluna apresenta as métricas precisão, sensibilidade e *F-Measure* para a classificação dos sentimentos positivos expressos pelos

usuários. A segunda coluna apresenta as mesmas métricas para a classificação dos sentimentos negativos. A terceira coluna apresenta as métricas para a soma de todas as classificações realizadas.

Tabela 2 – Treinamento Amazon e teste IMDB (Acurácia: 0,75309)

	Classe	Classe	Total
<b>Precisão</b>	0,78056	0,72133	0,78054
<b>Sensibilidade</b>	0,77085	0,78684	0,76120
<b>F-Measure</b>	0,77567	0,75266	0,77075

Fonte: Dados da pesquisa.

Tabela 3 – Treinamento Amazon e teste Yelp (Acurácia: 0,66700)

	Classe	Classe	Total
<b>Precisão</b>	0,68356	0,71133	0,65354
<b>Sensibilidade</b>	0,67087	0,75994	0,71183
<b>F-Measure</b>	0,67716	0,73483	0,68144

Fonte: Dados da pesquisa.

Tabela 4 – Treinamento Amazon e teste Amazon (Acurácia: 0,98077)

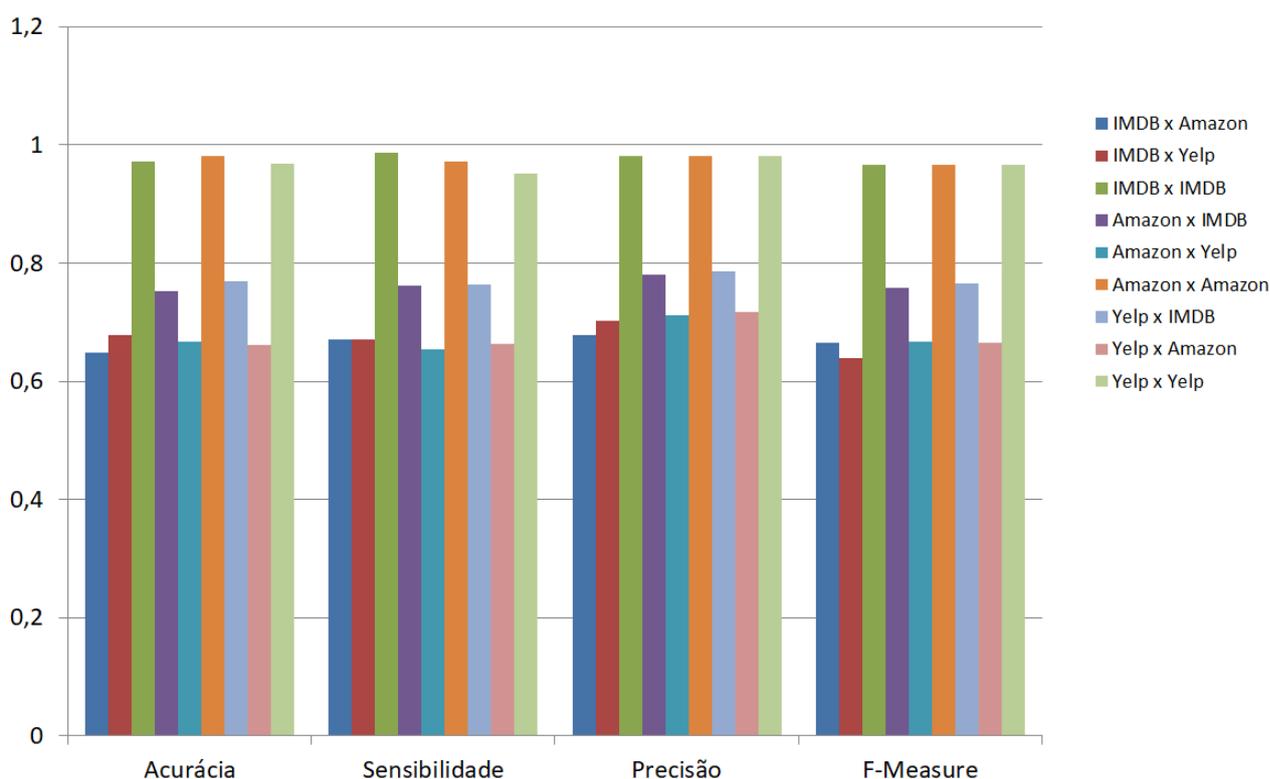
	Classe	Classe	Total
<b>Precisão</b>	0,98056	0,96133	0,98054
<b>Sensibilidade</b>	0,96685	0,98984	0,97120
<b>F-Measure</b>	0,97366	0,97538	0,97585

Fonte: Dados da pesquisa.

As Tabelas 2 e 3 apresentam as métricas de avaliação dos experimentos de classificação das avaliações dos *sites* IMDB e Yelp, respectivamente, utilizando-se para treinamento o subconjunto da Amazon. Nestes experimentos, as métricas variaram de 65% a 78%. Os valores de precisão, sensibilidade e acurácia apresentados na Tabela 4, cujo classificador utilizou os subconjuntos de treino da Amazon para classificar as avaliações do *site* da Amazon, variaram de 95% a 99%. Este percentual se repetiu nos três contextos, o de produtos, o de filmes e o de serviços. Portanto, os experimentos cujo subconjunto de treino e o conjunto de teste faziam parte do mesmo contexto geraram resultados mais satisfatórios. O Gráfico 1 exibe os resultados encontrados para as métricas de avaliação dos nove experimentos. No eixo vertical, são apresentados os resultados referentes às métricas acurácia, sensibilidade, precisão e *F-Measure*, dispostas no eixo horizontal. Para cada métrica, nove experimentos são avaliados. As três primeiras barras de cada

métrica são resultados referentes ao treinamento que utiliza os dados do IMDB. As três barras centrais utilizam dados da Amazon para treinamento. As três últimas barras são relativas ao treinamento que utiliza dados do Yelp. As barras que apresentam o valor da métrica mais elevado são sempre aquelas que utilizaram dados do próprio *site* para treinamento, que são as terceiras barras de cada conjunto, o que evidencia a importância de se utilizar dados para treinamento do classificador provenientes do conjunto que se quer classificar. A informação resultante da análise do gráfico de barras reafirma o fato de que as expressões utilizadas pelos usuários para avaliação dos produtos são próprias do contexto que está sendo avaliado.

Figura 5 – Resultados do experimento 1



Fonte: Elaborada pelos Autores.

## 4.2 Experimento 2: pré-processamento com extração de termos da Wikipedia

Utilizando o método de extração de palavras com informações provenientes da Wikipedia, todos os testes do experimento 1 foram refeitos. As Tabelas 5, 6 e 7 apresentam, agora, as métricas de avaliação após o novo pré-processamento da base de dados. É importante ressaltar que os valores de acurácia, precisão, sensibilidade e *F-Measure* são em média 9,8% superiores aos valores apresentados nas Tabelas 2, 3 e 4.

Tabela 5 – Treinamento Amazon e teste IMDB (Acurácia: 0,84309)

	<b>Classe</b>	<b>Classe</b>	<b>Total</b>
<b>Precisão</b>	0,83087	0,82143	0,84068
<b>Sensibilidade</b>	0,81031	0,81984	0,81350
<b>F-Measure</b>	0,82046	0,82063	0,82687

Fonte: Dados da pesquisa.

Tabela 6 – Treinamento Amazon e teste Yelp (Acurácia: 0,81757)

	<b>Classe</b>	<b>Classe</b>	<b>Total</b>
<b>Precisão</b>	0,81049	0,83154	0,78658
<b>Sensibilidade</b>	0,87485	0,78954	0,85163
<b>F-Measure</b>	0,84144	0,81000	0,81781

Fonte: Dados da pesquisa.

Tabela 7 – Treinamento Amazon e teste Amazon (Acurácia: 0,98571)

	<b>Classe</b>	<b>Classe</b>	<b>Total</b>
<b>Precisão</b>	0,98756	0,96133	0,98154
<b>Sensibilidade</b>	0,96695	0,98984	0,97521
<b>F-Measure</b>	0,97715	0,97538	0,97836

Fonte: Dados da pesquisa.

Constata-se, assim, pelos resultados obtidos, que o uso de um método mais refinado de pré-processamento afeta significativamente o resultado do processo de treinamento do classificador para identificar sentimentos positivos ou negativos expressos pelos usuários. Este resultado, também, ressalta a relevância do vocabulário específico de cada contexto para o processo de aprendizagem do classificador.

### 4.3 Análise dos experimentos

O primeiro experimento confirmou que o modo de expressão do usuário se altera dependendo do contexto que está sendo avaliado, influenciando o desempenho do algoritmo de classificação, especificamente, em seu processo de aprendizado.

O segundo experimento, que utilizou o pré-processamento com extração de termos da Wikipedia, produziu resultados que comprovaram

que um tratamento mais refinado no vocabulário melhorou significativamente, em até 30%, os resultados expressos pelas métricas utilizadas para avaliar o processo de classificação. Durante a etapa de pré-processamento, observou-se que o novo método reduz o universo dos termos relevantes (*cerca de 1150 para a base textual e 700 para cada um dos conjuntos*). Devido à exclusão de termos pouco relevantes para a classificação do sentimento expresso pelo usuário, os resultados foram melhores. Ainda assim, analisando o modelo de treino e o vetor de palavras e seus pesos, constatou-se que alguns termos eram consideravelmente prejudiciais para o processo de aprendizado. Para verificar essas ocorrências, a base de dados foi examinada com mais profundidade.

Verificando minuciosamente o vetor de termos relevantes, foi possível observar que alguns fatores da comunicação humana interferem fortemente no peso que é atribuído ao termo naquele subconjunto e, conseqüentemente, sua contribuição para o processo de aprendizado do classificador. Termos de avaliações carregadas de ironia, figuras de linguagem ou expressões idiomáticas regionalizadas eram prejudiciais ao processo, pois não representavam explicitamente o sentimento do avaliador. Um exemplo é a palavra *stale* do inglês: vencido, velho, passado (tradução livre) que é utilizada em seu significado formal para alimentos. Este termo não foi encontrado nenhuma vez no conjunto do Yelp, que fala de serviços, e muitos são restaurantes que tratam com alimentos. O termo foi encontrado 3 vezes no conjunto do IMDB, que fala de filmes. Este tipo de abstração dificulta o processo de aprendizado do classificador.

Outro exemplo que deve ser destacado são os adjetivos usados fora do seu contexto formal como as palavras *bright* e *shine* que significam, respectivamente, brilhante e brilho (tradução livre). Estas palavras foram encontradas em várias avaliações de filmes e atores com um sentido abstrato que denota um sentimento positivo. Entretanto, o mesmo termo foi encontrado em avaliações de estabelecimentos, denotando algo negativo como um local muito claro ou muito ofuscante. Este termo não é ignorado pelo algoritmo devido à alta incidência, mas, as avaliações em que aparece foram quase todas (93%) classificadas incorretamente.

Outra preocupação relativa ao processo de classificação foram as sentenças que expressavam expectativa frustrada, em que o autor, deliberadamente, estabelece um contraste relativo ao início do discurso, por exemplo: *"This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up"* (tradução livre: Este filme deveria ser brilhante. Parece uma boa trama, os atores são de primeira linha e os coadjuvantes são bons também e o Stallone está tentando atuar bem. Mas, o filme não se sustenta) ou *"I hate the Spice Girls... Why I saw this movie is a really, really, really long story, but I did, and one would think I'd despise every minute of it. But... Okay, I'm really ashamed of it, but I enjoyed it. I*

*mean, I admit it's a really awful movie ...the ninth floor of hell...The plot is such a mess that it's terrible. But I loved it.*" (tradução livre: Eu odeio as Spice Girls... O motivo de eu ter assistido esse filme é uma história muito muito longa, mas eu o fiz, e devem pensar que eu desprezei cada minuto. Mas... ok, eu estou muito envergonhado e é realmente um péssimo filme... o nono andar do inferno... a trama é tão bagunçada que é terrível. Mas eu amei). Nos dois exemplos citados, o ser humano conseguiria, facilmente, detectar o verdadeiro sentimento da avaliação. Classificadores automáticos, entretanto, não atuam bem nessas sentenças, já que muitas palavras indicam o oposto do sentimento que desejam expressar.

## 5 Considerações finais e trabalhos futuros

A análise de sentimentos, elemento essencialmente subjetivo e próprio do ser humano, é uma tarefa complexa de ser realizada por máquinas e ainda constitui um desafio para as pesquisas relacionadas a aprendizado de máquina. Os resultados obtidos, neste trabalho, mostram que as técnicas de classificação disponíveis conseguem alcançar resultados satisfatórios. Mesmo tendo sido utilizada uma base de dados curada, balanceada e tratada por especialistas para a realização de experimentos, foram identificadas algumas ressalvas, cujas análises foram objeto desse trabalho. Não foi possível eliminar as dificuldades de aprendizado e de compreensão do modelo para as classificações de sentimentos que envolviam características especiais ou particulares da expressão humana em que o modo de se manifestar é o oposto do que se quer relatar. Quanto mais subjetivo, mais sarcástico ou mais irônico o teor da avaliação do usuário, mais difícil é, para a máquina, classificar o texto como sendo o de uma avaliação positiva ou negativa.

O pré-processamento dos textos disponíveis possibilitou observar que, ainda que as possibilidades de avaliação em contextos distintos sejam iguais (ou seja, positiva ou negativa), o elemento a ser avaliado possui termos considerados relevantes pelo pré-processamento tradicional que não necessariamente o são. Este fato ocasiona o surgimento de muitos termos que são relevantes para poucas instâncias e estes termos possuem um valor de relevância muito baixo. O algoritmo que referencia a Wikipedia foi capaz de melhorar a escolha dos termos, identificando termos compostos e títulos, modelando a linguagem não apenas com termos formados por uma palavra, mas, também, com esses termos compostos, chamados de n-gramas (VIDAL et al., 2012). Essa diferença na etapa de pré-processamento foi suficiente para gerar um ganho médio aproximado de 10% nos resultados, evidenciado pelas Tabelas 5, 6 e 7 e validar a importância da linguagem do contexto no processo de classificação. Esta melhoria aplicada a bases de dados maiores do que as utilizadas neste trabalho pode gerar um ganho significativo para a classificação automática de textos.

Uma proposta de continuidade dos estudos para abordar as questões identificadas nesse trabalho seria realizar experimentos semelhantes em bases de dados não balanceadas e não curadas, extraídas diretamente de uma fonte de textos, como por exemplo, os próprios portais. Outra proposta seria encontrar alternativas para fazer com que esta etapa seja mais sensível às nuances da expressão de sentimentos dos seres humanos, tendo verificado que o pré-processamento textual impacta positiva e significativamente o processo de aprendizado. Além disso, existem algoritmos de desambiguação que, utilizando relações sintáticas e semânticas, poderiam ser usados para determinar o melhor sentido para os termos selecionados, definindo com maior assertividade se o sentimento expresso pelos usuários caracteriza um contexto positivo ou negativo. Espera-se, com a análise das etapas de pré-processamento e de treinamento do classificador utilizado, contribuir com os trabalhos de pesquisa relacionados à aprendizagem de máquina voltada para os processos de organização e de recuperação da informação.

## Referências

- ALLAHYARI, M. *et al.* A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 23, 2017, Halifax, [Proceedings...]. Halifax, SIGKDD, 2017. Disponível em: <https://arxiv.org/pdf/1707.02919.pdf>. Acesso em: 11 dez. 2018.
- ARANHA, C.; PASSOS, E. A Tecnologia de Mineração de Textos. *Revista Eletrônica de Sistemas de Informação*, Curitiba, v. 5, n. 2, p.1-8. 2006. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/171/66>. Acesso em: 30 out. 2017.
- BAEZA-YATES, R.; RIBEIRO-NETO; B. *Modern Information Retrieval*. New York: Addison- Wesley, 1999.
- BRAGA, A. de P. *et al.* *Redes Neurais Artificiais: teoria e aplicações*. São Paulo: LTC, 2007.
- HAYKIN, S. S. *Redes Neurais Princípios e Práticas*. 2. ed. New York: Bookman, 2001.
- LADEIRA, A. P. Processamento de linguagem natural: caracterização da produção científica dos pesquisadores brasileiros. *Perspectivas em Ciência da Informação*, v. 7, n. 4, p. 214-215, out./dez. 2012. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1685/1081>. Acesso em: 15 dez. 2017.
- LARA, J. E. *et al.* Do encanto à vingança: o processamento e o comportamento do consumidor sobre informações de atributos de produtos tecnológicos. *Perspectivas em Ciência da Informação*, v. 22, n. 4, p. 157-176, Belo Horizonte, out /dez. 2017. Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/3209/1977>. Acesso em: 15 dez. 2017.

MCCLELLAND, J. L.; CLEEREMANS; A. Connectionist Models. *In*: BYRNE, A; CLEERMANS, A.; WILKEN, P.(Eds.). *Oxford Companion to Consciousness*. New York: Oxford University Press, 2009.

MIERSWA, I.; WURST, M.; KLINKENBERG, R.; SCHOLZ, M.; EULER, T. YALE: Rapid prototyping for complex data mining tasks. *In*: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 12, 2006, Halifax. [*Proceedings...*] Halifax, SIGKDD, 2006. p. 935-940. Disponível em: <https://www.researchgate.net/publication/220017671>. Acesso em: 15 dez. 2018.

MISHU, Sadia Z.; RAFIUDDIN, S. M. Performance analysis of supervised machine learning algorithms for text classification. *In*: INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY, 19, 2016, Dhaka, [*Proceedings...*] Dhaka, North South University, 2016. p. 409-413, 2016. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7860233>. Acesso em: 13 dez 2018.

MUKHERJEE, Indrajit; SAHANA, Sudip; MAHANTI, P. K. An Improved Information Retrieval Approach to Short Text Classification. *International Journal of Information Engineering and Electronic Business*, Hong Kong, v. 9, n. 4, p. 31-37, 2017. Disponível em: <http://www.mecspress.org/ijieeb/ijieeb-v9-n4/IJIEEB-V9-N4-5.pdf>. Acesso em: 13 dez 2018.

NAIK, C.; KOTHARI, V.; RANA, Z. Document Classification using Neural Networks Based on Words. *International Journal of Advanced Research in Computer Science*, Udaipur, v. 6, n. 2, 2015. Disponível em: <https://search.proquest.com/docview/1682543110?pq-origsite=gscholar>. Acesso em: 13 dez. 2018.

NG, H. T.; GOH, W. B.; LOW, K. L. Feature selection, perceptron learning and a usability case study for text categorization. *In*: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 20, 1997, Philadelphia. [*Proceedings...*] Philadelphia: ACM SIGIR, 1997. p. 67-73. Disponível em: <https://app.dimensions.ai/details/publication/pub.1046686117>. Acesso em: 13 dez.2018.

PRASANNA, P. L.; RAO, D. R. Text classification using artificial neural networks. *International Journal of Engineering & Technology*, Dubai, v. 7, n. 1.1, p. 603-606, 2017. Disponível em: <https://www.sciencepubco.com/index.php/ijet/article/view/10785/3968>. Acesso em: 13 dez. 2018.

RAPIDMINER. Lightning Fast Data Science. Data Science Plataforma, 2018. Disponível em: < <https://rapidminer.com/>>. Acesso: 20 set. 2017.

REZENDE, S. O.; MARCACINI, R. M.; MOURA; M. F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. *Revista de Sistemas de Informação da FSMA*, n.7, p. 7-21, 2011. Disponível em: <

[http://www.fsma.edu.br/si/edicao7/FSMA\\_SI\\_2011\\_1\\_Principal\\_3.pdf](http://www.fsma.edu.br/si/edicao7/FSMA_SI_2011_1_Principal_3.pdf)>.

Acesso: 30 out. 2017.

RIES, B.E. Sensação e percepção. In: B. E. Ries & E. W. Rodrigues (ORGS), *Psicologia e Educação: fundamentos e reflexões*, Porto Alegre: EDIPUCRS, 2004.

SALTON, G.; WONG; A.; YANG; C. S. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, New York, v. 18, n. 11, p. 613-620, nov. 1975. Disponível em:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.446.5101&rep=rep1&type=pdf>. Acesso em: 13 dez. 2018.

SAMAL, Biswaranjan; PANDA, Mrutyunjaya. Performance Analysis of Supervised Machine Learning Techniques for Sentiment Analysis. In: *INTERNATIONAL CONFERENCE ON SENSING, SIGNAL PROCESSING AND SECURITY*, 3, 2017, Chennai, [Proccedings...] Tamilnadu, IEEE, 2017. p. Disponível em:

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8071579>.

Acesso: 13 dez. 2018.

SCHERER, K. *What are emotions? And how can they be measured?* 4. ed. *Social Science Information*, Pennsylvania, v. 44, n.4, p. 695-792, dec. 2005. Disponível em:

<https://www.researchgate.net/publication/229060094>. Acesso em: 13 dez. 2018.

SCHMIDHUBER, J. Deep Learning in neural networks: An overview. *Neural Networks Magazine*, Ithaca, n. 61, p. 85-117, 2015. Disponível em:

[https://ac.els-cdn.com/S0893608014002135/1-s2.0-S0893608014002135-main.pdf?tid=df160bd9-bee0-4fba-a5a2-babf35a3297f&acdnat=1544729470\\_0c480e3956f5a32a4b46fe5c62d382a8](https://ac.els-cdn.com/S0893608014002135/1-s2.0-S0893608014002135-main.pdf?tid=df160bd9-bee0-4fba-a5a2-babf35a3297f&acdnat=1544729470_0c480e3956f5a32a4b46fe5c62d382a8). Acesso em: 30 out. 2017.

SHAFIABADY, N. et al. Using unsupervised clustering approach to train the Support Vector Machine for text classification. *Neurocomputing*, v. 211, p. 4-10, 2016. Disponível em:

[https://ac.els-cdn.com/S0925231216305604/1-s2.0-S0925231216305604-main.pdf?tid=bb7842d6-2077-409d-b4fc-0c2bdb98e78f&acdnat=1544729621\\_f72d387b10db7f67d9071256f599357a](https://ac.els-cdn.com/S0925231216305604/1-s2.0-S0925231216305604-main.pdf?tid=bb7842d6-2077-409d-b4fc-0c2bdb98e78f&acdnat=1544729621_f72d387b10db7f67d9071256f599357a). Acesso em: 13 dez. 2018.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press, 2014.

VIDAL, M. et al. Selecting Keywords to Represent Web Pages Using Wikipedia Information. In: *BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB*, 18, 2012, [Proccedings...] São Paulo: Webmidia, 2012. p. 375-

382. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.711.588&rep=rep1&type=pdf>. Acesso em: 30 out. 2017.

VINODHINI, G.; CHANDRASEKARAN, R. M. A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University-Computer and Information Sciences*, Riade, v. 28, p. 2-12, 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1319157815001020>. Acesso em: 13 dez. 2018.