

Análise de Sentimento em Redes Sociais no Idioma Português com Base em Mensagens do Twitter

Kássio T. C. Junqueira, Anita Fernandes

Mestrado em Computação Aplicada – Universidade do Vale do Itajaí (UNIVALI)
Caixa Postal 360 – 88.302-202 – Itajaí – SC – Brazil

Kassio.junqueira@gmail.com, anita.fernandes@univali.br

***Abstract.** Sentiment Analysis applies to the analysis of opinions, sentiments, evaluations, attitudes and emotions about texts. These texts, when written in Portuguese, face difficulties of analysis because of the grammatical complexity and richness of words in this language. This paper presents a comparison among some machine learning algorithms and lexical approach using Twitter messages wrote in Portuguese related to 2016 Olympic Games. The results showed that the machine learning approach achieved better results in relation to the lexical approach, and that SVM algorithm obtained the best performance, reaching 89.5% of accuracy in the best scenario.*

***Resumo.** Análise de sentimentos refere-se à análise de opiniões, sentimentos, avaliações, atitudes e emoções a partir de textos. Estes textos quando escritos em português enfrentam dificuldades na análise devido à complexidade gramatical e riqueza de vocábulos presentes neste idioma. Este trabalho apresenta um comparativo entre os diversos algoritmos de aprendizado de máquina e de abordagem léxica utilizando mensagens do Twitter em português associadas às Olimpíadas de 2016. Os resultados mostraram que a abordagem de aprendizado de máquina alcançou melhor resultado em relação a abordagem léxica, sendo que o algoritmo SVM obteve o melhor desempenho atingindo 89,5% de acurácia no melhor cenário.*

1. Introdução

Os avanços tecnológicos permitem coletar, armazenar e processar grandes volumes de dados. Sites de notícias, *blogs* e redes sociais são fontes de dados que concentram grandes volumes de informação [Martins, Pereira e Benvenuto 2015]. Em razão da popularidade e enorme quantidade de dados disponíveis, várias aplicações surgem com o objetivo de extrair opiniões a partir dessas fontes. A análise de comentários expressos nessas fontes de dados requer muito esforço quando tratada de forma manual, principalmente devido ao volume de dados gerados [Alves *et al* 2014]. Uma forma de sumarizar estas opiniões de forma automática e extrair o que as pessoas pensam ou sentem em relação a determinado assunto é através da análise de sentimento.

A análise de sentimentos é o estudo que analisa opiniões, sentimentos, avaliações, atitudes e emoções em entidades como produtos, serviços, organizações, indivíduos, eventos, tópicos, entre outros. Análise de sentimentos também é comumente conhecida por vários outros termos, tais como: extração de opinião, mineração de

sentimento, análise de subjetividade, análise afetiva, análise de emoções e mineração de opinião [Liu 2012].

O elemento essencial para análise de sentimentos é uma opinião, que pode ser expressa em redes sociais, sites de notícias, sites de revisões de produtos ou comentários em páginas Web. Segundo Liu (2012) uma opinião é formada por dois elementos principais: um alvo e um sentimento expresso em relação ao alvo. Um alvo pode ser definido como uma entidade, ou seja, representa uma pessoa, marca de um determinado produto, ou qualquer sujeito que esteja relacionado à opinião. Já o sentimento é a opinião ou emoção expressa em relação ao alvo. Por exemplo: “os jogos do Rio 2016 estão surpreendentes”. Na frase anterior, a parte “os jogos do Rio 2016” representa o alvo, e a palavra “surpreendentes” representa o sentimento expresso em relação aos jogos do Rio.

Muitos desafios surgem na análise destas opiniões em grandes volumes de dados. Textos com erros ortográficos, sentenças sintaticamente malformadas, textos com conteúdo irônico ou sarcástico, abreviações de palavras e ambiguidade, são alguns dos desafios encontrados para análise de dados textuais [Chakraborty e Krishna 2014]. Além desses desafios, o idioma português apresenta algumas dificuldades em sua compreensão devido à riqueza de vocábulos e flexibilidade gramatical. A variação dos tempos e formas verbais, regras de concordância e flexões verbais, são desafios comumente encontrados no idioma português [Inoki, 1992].

A análise de sentimentos utiliza técnicas de vários campos da computação como o processamento da linguagem natural, recuperação da informação, mineração de dados, e estatística. Essa área de análise também é associada a extração de informação que lida com identificação de opiniões, emoções e avaliações expressas em textos. A análise de sentimentos utiliza abordagens ou métodos que se dividem em duas categorias: abordagens baseadas em aprendizado de máquina e abordagens léxicas [Medhat, Hassan e Korashy 2014].

As abordagens baseadas em aprendizado de máquina utilizam algoritmos com objetivo de identificar o sentimento expresso a partir de um conjunto de dados previamente treinado, tendo como referência textos onde o sentimento já é conhecido. SVM (*Support Vector Machine*), Naïve Bayes e Redes Neurais Artificiais, são exemplos de algoritmos aplicados nesta abordagem. As abordagens léxicas utilizam estatísticas e dicionários de palavras com anotações de sentimentos para extrair o sentimento expresso e determinada opinião. Os dicionários OpLexicon e SentiLex, são exemplos de dicionários comumente utilizados nesta abordagem.

Com base nisto, este artigo apresenta a comparação entre a abordagem baseada em aprendizado de máquina e a abordagem léxica, afim de verificar qual tem o melhor desempenho no contexto da análise de sentimentos de *tweets* escritos em português.

A seguir serão apresentados alguns conceitos necessários para a compreensão do trabalho, bem como o estudo realizado e seus resultados.

2. Técnicas para Análise de Sentimentos

A análise de sentimentos utiliza várias técnicas, dentre as principais é possível destacar Extração de Características, Detecção de Emoções, Transferência de Aprendizagem,

Construção de Recursos e a Classificação de Sentimento [Medhat, Hassan e Korashy 2014].

Neste trabalho, o foco é a Classificação de Sentimento, que tem por objetivo mediar a polaridade ou orientação da opinião em classes positivas ou negativas. Também é possível determinar se a opinião expressa é neutra, ou seja, quando a sentença ou documento é objetivo ou não expressa uma opinião [Medhat, Hassan e Korashy 2014]. As abordagens para classificação de sentimento podem ser divididas em abordagens léxicas e abordagens de aprendizado de máquina.

2.1. Classificação de Sentimento com Aprendizagem de Máquina

A classificação de sentimento através do aprendizado de máquina cria um modelo de classificação que utiliza mensagens associadas a um sentimento ou polaridade, para classificar mensagens não associadas a uma classe. Os dados utilizados para o modelo de classificação são chamados de dados de treinamento e geralmente são formados por mensagens de opinião onde cada mensagem está associada a uma classe que pode ser um sentimento ou polaridade. A partir desses dados, a classificação de sentimento busca associar mensagens que não estão associadas a uma classe, tendo por base as mensagens que possuem classes associadas [Medhat, Hassan e Korashy 2014]. Essa abordagem é dividida em métodos supervisionados e não supervisionados.

- **Métodos Supervisionados:** são dependentes de um conjunto prévio de treinamento. Tem por objetivo construir um modelo conciso de distribuição das classes para realizar predições [Kotsiantis 2011]. Esse método pode utilizar várias técnicas como árvores de decisão, modelos probabilísticos (Naïve Bayes, Redes Bayesianas, Máxima Entropia), lineares (Redes Neurais Artificiais, e SVM), e modelos baseados em regras [Wilson, Wiebe e Hoffmann 2009] [Medhat, Hassan e Korashy 2014] [Tan, Steinbach e Kumar 2009] [Zhang e Zhou 2010]
- **Métodos não Supervisionados:** não necessitam de dados previamente treinados e utilizam conjuntos de palavras, estatística ou dicionários para classificação. Uma das principais vantagens destes métodos é a independência do domínio, ou seja, o modelo de classificação não é baseado em um conjunto restrito de treinamento [Soni e Patel 2014].

Esse trabalho utiliza o método supervisionado através da comparação dos algoritmos de aprendizado de máquina Naïve Bayes, SVM, Máxima Entropia, Random Forest e Árvore de Decisão.

2.2. Classificação de Sentimento pela Abordagem Léxica

A abordagem léxica utiliza dicionários ou coleções de palavras de opinião com anotações que indicam o sentimento em positivo ou negativo. Essa abordagem pode ser criada de forma manual ou automática a partir de uma lista de palavras. Algumas pesquisas utilizam adjetivos como indicadores de orientação semântica do texto. Basicamente uma lista de adjetivos é compilada em um dicionário, então dado o texto, todos os adjetivos são extraídos e utilizados para pontuar os valores da orientação semântica [Turkey 2002]. A classificação de sentimento léxico se divide basicamente em duas abordagens: dicionário e *corpus*.

- A abordagem baseada em dicionários utiliza um conjunto de palavras com anotações de polaridade e classifica a sentença de acordo com a polaridade de cada palavra do dicionário. Os dicionários utilizados para a classificação podem conter ícones que expressam emoção, também conhecidos como *emoticons*, *hashtags*, expressões, palavras e símbolos que denotam algum sentimento [Avanço e Nunes 2014]. Alguns dicionários encontrados na literatura para o idioma português são: SentiLex¹, OpLexicon², Sentimeter-br³, e LIWC⁴.
- Abordagem baseada em *Corpus* é um método que depende de padrões sintáticos ou padrões que ocorrem no corpo do texto. *Corpus* é o corpo do texto de linguagem natural utilizado para acumular estatísticas sobre a linguagem natural. As informações extraídas incluem anotações de cada palavra que indicam suas partes do discurso como adjetivos, substantivos, advérbios, entre outros. Essa abordagem se divide em baseada em estatísticas e baseada na orientação semântica [Santos e Ladeira 2014] [Medhat, Hassan e Korashy 2014].

2.3. Métricas para Comparação de Abordagens de Análise de Sentimentos

Algumas métricas são comumente encontradas na literatura afim de analisar e comparar as diversas abordagens utilizadas na análise de sentimentos [Martins, Pereira e Benvenuto 2015]. Neste trabalho as métricas utilizadas foram precisão, acurácia, *recall*, abrangência e medida de concordância, conceituadas a seguir de acordo com Martins, Pereira e Benvenuto (2015).

A métrica Precisão mostra a quantidade de acertos na classificação de uma determinada classe (positivo ou negativo).

A acurácia mede o desempenho das abordagens considerando os registros corretamente classificados em relação a todos os registros. Não considera as classes positivas e negativas de forma isolada, mas sim o resultado de acordo com o total de registros.

A métrica *recall* apresenta a relação entre os registros classificados pertencentes a determinada classe em relação ao total de registros de cada classe que realmente fazem parte desta classe.

A abrangência apresenta a relação entre as palavras contidas nas mensagens de opinião e as palavras do dicionário. O objetivo é medir os termos comuns entre os dicionários e as mensagens classificadas.

A Medida de Concordância utilizada é o coeficiente Kappa cujo objetivo é definir um limiar de concordância entre os anotadores em tarefas de classificação. No processo de classificação das mensagens, o coeficiente Kappa (k) mediu o nível de

¹ A primeira versão foi disponibilizada em 2010 (SentiLex-PT01). Atualmente a versão 2 está disponível em: http://dmir.inesc-id.pt/project/SentiLex-PT_02

² Disponível em <http://ontolp.inf.pucrs.br/Recursos/downloads-OpLexicon.php>

³ Dicionário contém apenas palavras do domínio de cosméticos e não foi utilizado neste trabalho

⁴ Disponível em <http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>

concordância entre as pessoas que classificaram as mensagens e as mensagens classificadas.

3. Metodologia

Para a execução deste trabalho foram necessárias sete etapas: coleta de dados, armazenamento dos dados, seleção, treinamento e classificação manual, pré-processamento, mineração e análise dos resultados.

Na coleta dos dados foi desenvolvida uma aplicação para coleta e armazenamento diário de mensagens. Para desenvolver essa aplicação foi utilizada a API Twitter4J, que permitiu a integração com o serviço do Twitter. A aplicação buscou e armazenou *tweets* diariamente no período de 40 dias antes da realização dos jogos e 40 dias após o início das olimpíadas. Estes *tweets* referiam a mensagens que tinham o termo #rio2016. Foram excluídos *retweets*. Ao todo foram coletadas 988.512 mensagens.

As mensagens coletadas foram armazenadas em um bando de dados relacional e em seguida foi extraída uma amostra com sete mil mensagens aleatórias da base. O objetivo foi selecionar os dados para as etapas seguintes. Na etapa de classificação, os algoritmos supervisionados requerem dados previamente treinados. Para isso é necessário classificar mensagens de forma manual. Para não ter que classificar todas as mensagens da base manualmente, selecionou-se esta amostra inicial.

Na etapa de treinamento e classificação manual todas as mensagens selecionadas, conforme amostra inicial, foram lidas e classificadas manualmente por três pessoas falantes nativas do idioma português cada pessoa fez a leitura das mensagens e as classificou conforme sua polaridade.

A quantidade de mensagens selecionadas na amostra inicial foi baseada em um estudo preliminar, onde foi calculada a quantidade de mensagens que cada pessoa conseguia classificar diariamente. Neste estudo inicial cada pessoa conseguiu classificar em média 350 mensagens. A classificação das mensagens ocorreu durante o período de 4 semanas. Considerando o total de 20 dias úteis, cada pessoa conseguiu classificar 7 mil mensagens.

Após a leitura, cada mensagem era analisada e classificada em quatro classes: positiva, negativa, neutra ou *spam*. Às classes positiva e negativa foram atribuídas às mensagens que expressam uma opinião subjetiva em relação a qualquer acontecimento na Olimpíada. A classificação neutra foi atribuída às mensagens objetivas, ou seja, são mensagens que apresentaram conteúdo opinativo e geralmente narram um acontecimento. As mensagens que apresentaram conteúdo ambíguo ou apresentaram a polaridade positiva e negativa ao mesmo tempo, também foram classificadas como neutras. A classe *spam* foi atribuída às mensagens que apresentaram as seguintes características: mensagens relacionadas a protesto político, textos irônicos ou sarcásticos, textos de conteúdo publicitário, notícias, textos que continham conteúdo pornográfico ou palavras de baixo calão.

Após a classificação manual de todas as mensagens selecionadas, foi utilizado o índice Kappa para mensurar a medida de concordância entre as pessoas que classificaram as mensagens. A medida de concordância das mensagens foi de 74%, o

que é considerado como “boa medida” de concordância de acordo com a escala Kappa desenvolvida por Landis e Koch (1977).

Ao final dessa etapa de treinamento e classificação, foram excluídas as mensagens que apresentaram discordância na classificação, mensagens neutras e mensagens classificadas como *spam*.

Após a exclusão dessas mensagens, restaram 1.343 mensagens negativas. Para avaliar o desempenho dos algoritmos, foi selecionada a mesma quantidade de mensagens positivas e negativas. Dessa forma, a base de dados nesse estudo foi formada por 2.686 mensagens, sendo 1.343 mensagens positivas e 1.343 mensagens negativas. Vale ressaltar que a quantidade de mensagens selecionadas neste trabalho teve por objetivo de comparar o desempenho das abordagens e não classificar ou analisar todas as mensagens coletadas.

Após a coleta, armazenamento, seleção dos dados e classificação manual, as mensagens foram pré-processadas da seguinte forma:

- Conversão de todos os caracteres da base para minúsculo. Dessa forma o mesmo termo escrito de diferentes formas é avaliado como um único termo;
- Remoção das URLs nas mensagens. Estes termos não influenciam na opinião ou polaridade da mensagem;
- Remoção de letras duplicadas. Por exemplo: “muitoood bom o jogo” é transformado em “muito bom o jogo” para que ambas as frases sejam avaliadas como um único termo;
- Remoção de acentos, pontuação, caracteres especiais e *emoticons*;
- Remoção de *stopwords*; e
- Remoção de menções a usuários e *hashtags*.

Após a utilização dos recursos anteriores de pré-processamento, as mensagens foram convertidas em uma matriz de termos, cujo objetivo é preparar os dados para a etapa de treinamento e depois classificação. Essa matriz de termos consiste em estruturar as mensagens de maneira a torna-las processáveis pelos algoritmos de aprendizagem de máquina. A matriz utiliza o modelo espaço-vetorial, no qual cada mensagem ($d_1 \dots d_2$) é um vetor em um espaço multidimensional. Na Tabela 1 a letra \underline{T} representa cada termo ou palavra contida na mensagem, \underline{d} é cada dimensão da matriz e representa as mensagens, e a letra \underline{a} representa o valor do atributo que indica se um determinado termo está presente na mensagem ou não. O valor desse atributo geralmente é binário.

Tabela 1. Matriz de mensagens para classificação

	T_1	T_2	...	T_M
d_1	a_{11}	a_{12}	...	a_{1M}
d_2	a_{21}	a_{22}	...	a_{2M}
...
d_N	a_{N1}	a_{N2}	...	a_{nM}

4. Resultados

A avaliação das abordagens foi feita com objetivo de identificar qual a abordagem apresentou os melhores resultados para análise de sentimentos em português a partir das mensagens coletadas e classificadas. A seguir serão apresentados os resultados para cada abordagem.

4.1 Abordagem de Aprendizado de Máquina

Nos experimentos realizados, os algoritmos de aprendizado de máquina Naïve Bayes, SVM, Máxima Entropia, Random Forest e Árvore de Decisão, foram comparados. A proporção entre os dados de treinamento e testes variaram entre 50% e 90% com o objetivo de apresentar em qual dessas faixas o algoritmo obteve melhor desempenho.

Para identificar a correta proporção entre treino e teste, com objetivo de evidenciar o algoritmo com melhor desempenho, foram feitos experimentos variando o percentual de dados para treino entre 50% e 90% sendo este índice incrementado em 10% a cada novo experimento.

Considerando 50% dos dados para treino e 50% para teste, o algoritmo da Máxima Entropia teve o melhor desempenho, atingindo 84% de precisão na classificação das mensagens. Na sequência, o algoritmo SVM teve o segundo melhor resultado, alcançando 83% de precisão, seguido pelo algoritmo Random Forest com 82% e os algoritmos Naïve Bayes e Árvore de Decisão com 80,5% e 76,5% respectivamente.

Utilizando 60% dos dados para treino e 40% para testes, o algoritmo SVM teve o melhor desempenho, atingindo 85% de precisão na classificação de mensagens. Na sequência, o algoritmo de máxima entropia e Random Forest alcançaram 83% de precisão, seguidos pelo algoritmo Naïve Bayes e Árvore de Decisão com 82,5% e 77,5% respectivamente.

No cenário de 70% de dados para treino e 30% de teste, o algoritmo Random Forest teve o melhor desempenho atingindo 87% de precisão na classificação das mensagens. Na sequência, o algoritmo de máxima entropia alcançou 85,5% de precisão e SVM e Naïve Bayes alcançaram 85% de precisão. Quando o cenário considerou 80% dos dados para treino e 20% para testes, o algoritmo SVM teve um bom desempenho atingindo 88% de precisão na classificação das mensagens, empatado com Random Forest e Máxima Entropia.

Para o cenário utilizando 90% dos dados para treino e 10% para teste, o algoritmo SVM teve o melhor desempenho com 89,5% de precisão, seguido do algoritmo Random Forest com 89% e Máxima Entropia com 88% de precisão.

Considerando os cenários como um todo, pode-se verificar que o algoritmo SVM alcançou o melhor desempenho global entre os algoritmos analisados.

4.2 Abordagem Léxica

Nos experimentos utilizando a abordagem léxica as mensagens classificadas a partir dos dicionários foram validadas a partir das mensagens classificadas manualmente. A

métrica de abrangência foi utilizada para mensurar a cobertura do dicionário em relação as mensagens classificadas.

Aplicando os dicionários, percebeu-se que os três dicionários utilizados na pesquisa apresentaram menos de 50% de cobertura das mensagens: SentiLex obteve 47% de cobertura, seguido do OpLexicon, com 38% e LIWC com 32%.

O SentiLex além de ter apresentado a melhor cobertura entre os três, também teve a melhor acurácia (76%). Esse dicionário contém o maior número de palavras polarizadas em relação aos demais dicionários (20.663 palavras positivas e 53.656 palavras negativas).

O OpLexicon apresentou 72,5% de acurácia classificando com maior precisão as mensagens negativas em relação as mensagens positivas. Esse dicionário possui 8.675 palavras positivas e 14.469 palavras negativas.

O dicionário LIWC apresentou o menor desempenho quando comparado com os outros dicionários analisados. Sua acurácia foi de 70%. Ele tem 12.878 palavras positivas e 15.115 palavras negativas.

A quantidade de palavras contidas no dicionário não tem relação com os resultados alcançados. Isso porque quanto mais palavras relacionadas ao domínio existir no dicionário, maior será a sua abrangência e precisão dos resultados.

5. Conclusões

Este trabalho apresentou um comparativo entre os diversos algoritmos de aprendizado de máquina e de abordagem léxica utilizando mensagens do Twitter em português associadas às Olimpíadas de 2016. As mensagens coletadas foram submetidas a algoritmos de aprendizagem de máquina (SVM, Naïve Bayes, Árvore de Decisão, Máxima Entropia e Random Forest), bem como foram analisadas sob a ótica da abordagem léxica, utilizando os dicionários SentiLex, OpLexicon e LIWC.

Os algoritmos de aprendizado de máquina foram aplicados em diferentes cenários. Na comparação dos algoritmos de aprendizado de máquina, o SVM apresentou maior assertividade na classificação das mensagens, atingindo 89,5% de acurácia quando a proporção dos dados foi de 90% para treino e 10% para teste das mensagens.

Os dicionários utilizados na abordagem léxica tiveram uma abrangência baixa, principalmente se for levado em consideração o número de mensagens avaliadas (2.686).

Após avaliar os resultados obtidos, verificou-se que seria pertinente aplicar e comparar outras abordagens como as Redes Neurais Artificiais, Algoritmos Genéticos e SMO (Sequential Minimal Optimization) e também aplicar e comparar as abordagens estatísticas e semânticas, a fim de verificar se há uma melhoria no percentual de classificação.

Conclui-se com os resultados alcançados que a aprendizagem de máquina apresenta um melhor desempenho em relação a abordagem léxica, conforme os indicativos da literatura e que o algoritmo SVM se comporta melhor que os demais para o tipo de problema abordado.

Referências

- Alves, F. A. L.; Baptista, C. D. S.; Firmino, A. A.; Oliveira, M. G. D; Paiva, A. C. D. A Comparison of svm versus naïve-bayes techniques for sentiment analysis in tweets: a case study with the 2013 FIFA confederations cup. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 20., 2014, João Pessoa. *Proceedings...* New York: ACM, 2014. p. 123-130.
- Avanço, L. V.; Nunes, M. D. G. V. Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese. In: BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEMS (BRACIS), 2014, São Paulo. *Proceedings...* [S.I.]: IEEE, 2014. p. 277-281.
- Chakraborty, G.; Krishna, M. Analysis of unstructured data: applications of text analytics and sentiment mining. In: SAS GLOBAL FORUM, 2014, Washington. *Proceedings...* Cary: SAS Institute Inc, 2014. p. 1288-1302.
- Inoki, S. R. *Uma Gramática de um Fragmento do Português Baseado na Lógica Illocutória*. 1992. Dissertação (Mestrado) – Curso de Sistemas e Computação, Instituto Militar de Engenharia, Rio de Janeiro, 1992.
- Kotsiantis, S. B. Feature selection for machine learning classification problems: a recent overview. *Artificial intelligence review*, p. 1-20, May. 2011.
- Landis, J. R.; Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159-174, March. 1977.
- Liu, B. Sentiment analysis and opinion mining. *Synthesis Lectures On Human Language Technologies*, v. 5, n. 1, p. 1-167, May 2012.
- Martins, R. F.; Pereira, A.; Benevenuto, F. An approach to sentiment analysis of Web applications in Portuguese. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 21., 2015, Manaus. *Proceedings...* New York: ACM, 2015. p. 105-112.
- Medhat, W.; Hassan, A; Korashy, H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Engineering Journal*. v. 5, n. 4, p.1093-1113, December 2014.
- Santos F. L.; Ladeira M. The role of text pre-processing in opinion mining on a social media language dataset. In: BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEMS (BRACIS), 2014, São Paulo. *Proceedings...* [S.I]: IEEE, 2014. p. 50-54.
- Soni, V.; Patel, M. R. Unsupervised opinion mining from text reviews using sentiwordnet. *International journal of computer trends and technology (IJCTT)*, v.11, n. 5, p. 234-238, May 2014.
- Tan, P.; Steinbach, M.; Kumar, V. *Introdução ao data mining: mineração de dados*. Rio de Janeiro: Ciência Moderna, 2009.
- Turney, P. D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 40., 2002, Philadelphia. *Proceedings...* Stroudsburg: Association for Computational Linguistics, 2002. p. 417-424.

Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Computational linguistics*, v. 35, n. 3, p. 399-433, September 2009. ZHANG, Y.; JIN, R.; ZHOU, Z. Understanding bag-of-words model: a statistical framework. *International journal of machine learning and cybernetics*, v.1, n. 1, p. 43-52, December 2010.