

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA  
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

**RAFAEL PENCZKOSKI  
ROBSON JEAN PENTEADO**

**COMPARAÇÃO DE FERRAMENTAS DE PROCESSAMENTO DE  
LINGUAGEM NATURAL PARA ANÁLISE DE SENTIMENTO EM  
PORTUGUÊS: UM ESTUDO DE CASO EM AVALIAÇÕES ONLINE DE  
HOTÉIS**

**TRABALHO DE CONCLUSÃO DE CURSO**

**PONTA GROSSA**

**2019**

**RAFAEL PENCZKOSKI**  
**ROBSON JEAN PENTEADO**

**COMPARAÇÃO DE FERRAMENTAS DE PROCESSAMENTO DE  
LINGUAGEM NATURAL PARA ANÁLISE DE SENTIMENTO EM  
PORTUGUÊS: UM ESTUDO DE CASO EM AVALIAÇÕES ONLINE DE  
HOTÉIS**

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas, do Departamento Acadêmico de Informática, da Universidade Tecnológica Federal do Paraná.

Orientadora: Prof.<sup>a</sup> Dra. Helyane Bronoski Borges

**PONTA GROSSA**

**2019**



Ministério da Educação  
**Universidade Tecnológica Federal do Paraná**  
Câmpus Ponta Grossa

Diretoria de Graduação e Educação Profissional  
Departamento Acadêmico de Informática  
Tecnologia em Análise e Desenvolvimento de Sistema



---

## **TERMO DE APROVAÇÃO**

**COMPARAÇÃO DE FERRAMENTAS DE PROCESSAMENTO DE LINGUAGEM  
NATURAL PARA ANÁLISE DE SENTIMENTO EM PORTUGUÊS: UM ESTUDO DE  
CASO EM AVALIAÇÕES ONLINE DE HOTÉIS**

por

**RAFAEL PENCZKOSKI  
ROBSON JEAN PENTEADO**

Este Trabalho de Conclusão de Curso (TCC) foi apresentado em 2 de dezembro de 2019 como requisito parcial para a obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas. Os candidatos foram arguidos pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

---

**Prof.<sup>a</sup> Dra Helyane Bronoski Borges**  
Orientadora

---

**Prof.<sup>a</sup> Dra Simone de Almeida**  
Membro titular

---

**Prof. MSc Geraldo Ranthum**  
Membro titular

---

**Prof. MSc Geraldo Ranthum**  
Responsável pelo Trabalho de Conclusão  
de Curso

---

**Prof. Dr André Pinz Borges**  
Coordenador do curso

## RESUMO

PENCZKOSKI, Rafael; PENTEADO, Robson Jean. **Comparação de ferramentas de processamento de linguagem natural para análise de sentimento em português**: um estudo de caso em avaliações online de hotéis. 2019. 54 f. Trabalho de Conclusão de Curso de Tecnologia em Análise e Desenvolvimento de Sistemas - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2019.

Aplicar ferramentas de processamento de linguagem natural e ferramentas de Aprendizado de máquina para realizar uma análise de sentimento nas avaliações online de hotéis. A nota de uma avaliação não traz toda a expressividade da avaliação do usuário, com isso encontra-se a necessidade de aplicar uma análise nos textos das avaliações online para obter melhores parâmetros para uma análise da Reputação Online deste Hotel. Buscar ferramentas de Processamento de Linguagem Natural e ferramentas de Aprendizado de máquina para extrair os parâmetros de uma análise de sentimento nas avaliações. Espera-se que com a aplicação dessas ferramentas seja possível obter uma análise de sentimento nas avaliações de hotéis, assim sendo possível ter uma análise mais concisa da reputação online da empresa.

**Palavras-chave:** Análise de Sentimento. Processamento de Linguagem Natural. Avaliações Online.

## ABSTRACT

PENCZKOSKI, Rafael; PENTEADO, Robson Jean. **Comparison of Natural Language Processing tools for sentiment analysis in Portuguese:** a case study in online hotel reviews. 2019. 54 p. Work of Conclusion Course Graduation (Technology in Analysis and Development Systems) Federal Technology University - Paraná. Ponta Grossa, 2019.

Apply natural language processing tools and Machine Learning tools to conduct a sentiment analysis in online hotel reviews. The score of a review does not convey all the expressiveness of the user's review, so there is a need to apply an analysis in the texts of the online reviews to obtain better parameters for the hotel's online reputation statistics. Search for Natural Language Processing tools and Machine learning tools to extract the parameters from a sentiment analysis in the reviews. It is expected with the application of these tools it will be possible to obtain a sentiment analysis in hotels reviews, thus making it possible to have a more concise analysis of the company's online reputation.

**Keywords:** Sentiment Analysis. Natural Language Processing. Online Reviews.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Etapas gerais do processo de análise de sentimento .....	17
Figura 2 – Exemplo de avaliação positiva .....	35
Figura 3 – Exemplo de avaliação neutra .....	35
Figura 4 – Segundo exemplo de avaliação neutra .....	35
Figura 5 – Exemplo de avaliação negativa.....	35
Figura 6 – Exemplo de ruído encontrado na base de dados .....	37
Figura 7 – Exemplo de avaliação sem significado semântico .....	37
Figura 8 – Exemplo de avaliações em outro idioma .....	37
Figura 9 – Nuvem de palavras da base após o pré-processamento .....	40
Figura 10 – Resultados gerais das ferramentas.....	47
Gráfico 1 – Comparação em avaliações positivas.....	44
Gráfico 2 – Comparação das avaliações negativas .....	45
Gráfico 3 – Comparação das avaliações neutras .....	46
Gráfico 4 – Comparação geral na base de dados .....	47
Quadro 1 – Elaboração da matriz de confusão .....	22
Quadro 2 – Saídas dos léxicos utilizados.....	38

## LISTA DE TABELAS

Tabela 1 – Sumarização da base de dados .....	36
Tabela 2 – Matriz de confusão para o léxico Sentilex com NLTK .....	40
Tabela 3 – Matriz de confusão para o léxico Unilex com NLTK .....	41
Tabela 4 – Matriz de confusão para o léxico WordNet Affect BR com NLTK .....	41
Tabela 5 – Matriz de confusão para o léxico OpLexicon com NLTK .....	41
Tabela 6 – Matriz de confusão para o léxico Sentilex com <i>spaCy</i> .....	42
Tabela 7 – Matriz de confusão para o léxico Unilex com <i>spaCy</i> .....	42
Tabela 8 – Matriz de confusão para o léxico WordNet Affect BR com <i>spaCy</i> .....	42
Tabela 9 – Matriz de confusão para o léxico OpLexicon com <i>spaCy</i> .....	43

## LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

AM	Aprendizado de Máquina
API	<i>Application Programming Interface</i>
AS	Análise de Sentimentos
LIWC	<i>Linguistic Inquiry and Word Count</i>
MD	Mineração de Dados
NER	<i>Named Entity Recognition</i>
NLTK	<i>Natural Language Toolkit</i>
PLN	Processamento de Linguagem Natural



## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>10</b>
1.1 OBJETIVOS .....	11
1.1.1 Objetivo Geral .....	11
1.1.2 Objetivos Específicos .....	12
1.2 JUSTIFICATIVA .....	12
1.3 ORGANIZAÇÃO DO TRABALHO .....	13
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>14</b>
2.1 ANÁLISE DE SENTIMENTO .....	14
2.2 DESAFIOS .....	15
2.3 ETAPAS DO PROCESSO DE ANÁLISE DE SENTIMENTO .....	17
2.3.1 Coleta de Dados .....	17
2.3.2 Construção da Base de Dados .....	18
2.3.3 Pré-processamento .....	19
2.3.4 Classificação .....	20
2.3.5 Sumarização e Validação do Classificador .....	21
2.4 CLASSIFICAÇÃO DE ABORDAGENS DE ANÁLISE DE SENTIMENTOS .....	23
2.4.1 Abordagem Baseada em Aprendizado de Máquina .....	24
2.4.2 Abordagem Baseada em Léxico .....	24
2.4.3 Abordagens Estatística e Semântica .....	25
2.5 MÉTODOS E FERRAMENTAS DE ANÁLISE DE SENTIMENTOS .....	25
2.5.1 Ferramentas .....	26
2.5.2 Léxicos .....	29
2.6 CONSIDERAÇÕES DO CAPÍTULO .....	31
<b>3 METODOLOGIA</b> .....	<b>33</b>
3.1 ESCOLHA DAS FERRAMENTAS .....	33
3.2 COLETA DE DADOS E CONSTRUÇÃO DA BASE .....	34
3.3 PRÉ PROCESSAMENTO .....	36
3.4 SUMARIZAÇÃO .....	37
3.5 CONSIDERAÇÕES DO CAPÍTULO .....	38
<b>4 RESULTADOS</b> .....	<b>39</b>
4.1 VALIDAÇÃO DO PRÉ-PROCESSAMENTO .....	39
4.2 NLTK .....	40
4.3 SPACY .....	41
4.4 COMPARAÇÃO DOS RESULTADOS .....	43
4.5 CONSIDERAÇÕES DO CAPÍTULO .....	48
<b>5 CONCLUSÃO</b> .....	<b>49</b>
5.1 TRABALHOS FUTUROS .....	50
<b>REFERÊNCIAS</b> .....	<b>51</b>

## 1 INTRODUÇÃO

O Processamento de Linguagem Natural (PLN) é uma subárea de estudo da Inteligência Artificial (IA) que tem como principal objetivo estudar a capacidade e as limitações dos computadores compreenderem a linguagem dos seres humanos. Desta forma, é possível reconhecer o contexto, sintaxe, semântica, morfologia e analisar sentimentos. A Análise de Sentimentos (AS) tem o objetivo de obter a polaridade negativa ou positiva de um texto ou conjunto de textos por meio de PLN.

É importante para as pessoas que, durante o processo de tomada de decisão de uma compra, saibam o que outras pessoas pensam sobre um produto ou serviço (PANG; LEE, 2008). Atualmente, opiniões a respeito de produtos, marcas, serviços e lugares estão distribuídos por toda a internet. Essa massa de dados é de interesse diretamente as empresas que queiram gerenciar sua reputação e de seus produtos no mercado, identificando pontos de melhoria ou novas oportunidades de negócios (SALTER, 2011).

O interesse das empresas leva a um grande investimento em pesquisas relacionadas à análise de sentimentos. Por meio dessas pesquisas pode-se obter como resultado a automatização do processo de identificação do que é falado na internet sobre elas, compreensão de diálogos em mídias sociais e obtenção de conteúdo relevante que permita tomar ações específicas (BECKER, 2017).

Há inúmeros trabalhos no campo de extração de informações utilizando PLN. Como por exemplo “*SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods*” (RIBEIRO et al, 2016), “*Survey on mining subjective data on the web*” (TSYTSARAU; PALPANAS, 2012), “*Approaches, Tools and Applications for Sentiment Analysis Implementation*” (D’ANDREA et al., 2015), entre outros. Entretanto as implementações de ferramentas que dão suporte a língua portuguesa ainda apresentam resultados inferiores se comparados a outros idiomas (PIRES et al., 2017). Visando a melhoria desses resultados, as ferramentas seguem geralmente duas abordagens:

- Tradução de documentos para o inglês e em seguida a execução dos procedimentos de análise de sentimentos (DENECKE, 2008);
- Utilizar um léxico contendo palavras subjetivas em cada linguagem a ser abordada na análise (GINSCA et al., 2011).

Segundo Becker (2017), existem, entretanto, dificuldades na utilização de abordagens multilíngues, as quais têm como propósito analisar mais de um idioma dentro de um conjunto de textos, tais como:

- A baixa disponibilidade de ferramentas *open-source*;
- A tradução de textos por si só já é uma tarefa complexa, sendo que uma tradução errada no contexto afetará negativamente o resultado;
- Os sistemas de tradução podem gerar uma melhor combinação de resultado, limitando as diversidades e particularidades de determinada língua;
- O treinamento das abordagens propostas por redes neurais profundas pode levar um tempo considerável, apesar de apresentarem bons resultados.

Este trabalho visa elencar ferramentas que realizam o processamento de linguagem natural em conjuntos de dados em português, e realizar a comparação em uma base de dados de avaliações de hotéis *online*.

A escolha do estudo de caso se deu pelo motivo das avaliações *online* afetarem a decisão de compra do consumidor em hotéis (MURPHY; CHEN, 2014), além de que 68% dos consumidores dizem que avaliações positivas fazem eles confiarem mais em um negócio local (MURPHY, 2018).

A coleta das avaliações online será feita por *Web Scraping*, que se trata de uma técnica de extração de dados para coletar dados de sites na internet. Utilizando de robôs de extração, para salvar os dados em uma base de dados para posterior análise (VARGIU; URRU, 2013).

## 1.1 OBJETIVOS

### 1.1.1 Objetivo Geral

Realizar uma análise comparativa das ferramentas e técnicas de análise de sentimento, com suporte à língua portuguesa, utilizando como base de dados avaliações *online* de hotéis.

### 1.1.2 Objetivos Específicos

- Obter os dados para a comparação das ferramentas utilizando técnicas de *Web Scraping* de dados.
- Aplicar técnicas de pré-processamento nos dados obtidos para melhorar a qualidade da base de dados, e obter resultados mais precisos.
- Realizar um estudo de caso em avaliações *online* de hotéis para a comparação de resultados.

## 1.2 JUSTIFICATIVA

As avaliações *online* são um dos objetos de estudo do *Marketing Boca a Boca*, o qual estuda a comunicação entre indivíduos em relação a uma marca, produto ou serviço. Segundo Goldenberg et al. (2001), o *marketing boca a boca* é um fator que influencia fortemente na decisão de compra dos consumidores e é possível notar que avaliações *online* são um fator importante para negócios locais, pois afetam a tomada de decisão no processo de compra. O que pode ser observado na pesquisa de MURPHY (2018):

- 86% das pessoas leem avaliações *online* de negócios locais;
- 91% das pessoas entre 18 e 34 anos acreditam em avaliações *online* do mesmo modo que acreditam em recomendações pessoais;
- 80% das pessoas entre 18 e 34 anos já escreveram avaliações *online*.

O crescimento de aplicativos de internet e websites de turismo trouxe uma nova oportunidade para consumidores praticarem o *Marketing Boca a Boca* por meio de avaliações online (PAN et al, 2007).

As avaliações online são uma das métricas utilizadas para o ranqueamento de hotéis nas plataformas de buscas como *Google*, *Tripadvisor* e *Booking.com*, as quais serão usadas neste trabalho por serem as plataformas com maior popularidade dentre os consumidores.

### 1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho está estruturado em cinco capítulos. O Capítulo 2 apresenta a fundamentação teórica necessária para o desenvolvimento e entendimento do estudo a respeito de análise de sentimentos e um panorama geral sobre a disponibilidade de métodos e ferramentas na literatura.

O Capítulo 3 expõe os métodos utilizados para o desenvolvimento do trabalho, descrevendo as etapas executadas com objetivo de levantar dados para possibilitar a comparação entre os métodos selecionados para estudo.

O Capítulo 4 apresenta os resultados obtidos para cada experimento realizado e demonstra os mesmos por meio de gráficos e tabelas para facilitar a análise.

No Capítulo 5 as principais conclusões e contribuições deste trabalho são apresentadas, além de sugestões de trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

Este Capítulo apresenta a fundamentação teórica necessária para o desenvolvimento e entendimento deste trabalho. A Seção 2.1 aborda conceitos sobre a Análise de Sentimentos. A Seção 2.1 apresenta uma breve introdução sobre Análise de sentimentos. A Seção 2.2 revela os desafios encontrados na maioria dos estudos. Na Seção 2.3 serão descritas as etapas envolvidas no processo da análise de sentimentos. Na Seção 2.4 será mostrado um panorama geral de como as técnicas existentes na literatura são classificadas por diferentes autores. A Seção 2.5 apresenta algumas de ferramentas e léxicos mais conhecidos disponíveis para a realização de experimentos. Por fim, na Seção 2.6 são discutidas as considerações deste capítulo.

### 2.1 ANÁLISE DE SENTIMENTO

Segundo Liu (2010), a Análise de Sentimento (AS) também chamada de mineração de opinião, é uma área de pesquisa que reúne pesquisas de Mineração de Dados (MD), Aprendizado de Máquina (AM), Recuperação de Informação, Linguística e Processamento de Linguagem Natural (PLN). Essa área tem por objetivo analisar a opinião das pessoas, identificando seus sentimentos, avaliações, atitudes e percepções relacionadas a um determinado alvo no momento em que escreveram determinado texto.

Uma opinião é basicamente composta por dois elementos principais, sendo eles um alvo e um sentimento sobre esse alvo. Um alvo pode representar uma entidade, indivíduo, uma característica desse indivíduo e até mesmo uma empresa. Já um sentimento é representado por elementos subjetivos, como emoções, avaliações e experiências expressas pelo autor a respeito de mencionado alvo. Apesar do significado amplo do termo opinião, de maneira geral, a Análise de Sentimentos foca principalmente em opiniões que podem ser classificadas conforme sua polarização.

Determinar a polaridade de um sentimento significa determinar se ele representa algo positivo ou negativo, ou ainda neutro caso não seja possível defini-lo. Textos e sentenças podem conter informações fáceis de serem encontradas e

extraídas e outras nem tanto, pois demandam relacionamento complexos de palavras para serem compreendidas em sua totalidade. Na literatura encontram-se diferentes tipos de granularidade que tratam o assunto de AS. Conforme Liu (2012), define-se o nível a ser utilizado em cada aplicação de acordo com a quantidade de informação que se deseja extrair da mesma. Desta forma podem-se dividir os níveis de análise de granularidade em três principais grupos: documento, sentença e aspecto.

A análise a nível de documento realiza a classificação da polaridade de um documento por inteiro, verificando se a opinião expressa nele é positiva, negativa ou neutra (PANG; LEE, 2002). Esta granularidade é adequada quando o documento trata de um único alvo ou entidade (BECKER; TUMITAN, 2013), não sendo aplicado em situação diferente desta.

No que se refere a análise a nível de sentença, essa classifica a polaridade do sentimento expresso em cada sentença do documento. Esta abordagem pode ser utilizada quando deseja-se analisar o sentimento a respeito de alvos distintos contidos em um único documento ou realizar uma análise detalhada dos diferentes sentimentos a respeito de um alvo específico. Uma sentença também pode ser dividida em diversas cláusulas (THET; KHOO, 2010).

A análise a nível de aspecto também pode ser chamada de *feature level*, é onde é realizada a análise com a granularidade mais fina, com foco na opinião expressa, e não levando em consideração os construtos para expressá-la (documento, sentença). Este é um nível de análise que permite identificar o sentimento de um alvo sob diferentes perspectivas, considerando seus diferentes atributos.

## 2.2 DESAFIOS

Recentemente a Análise de Sentimento tornou-se um tópico importante em áreas de pesquisa como Aprendizagem de Máquina e Processamento de Linguagem Natural, motivados principalmente pelos desafios e dificuldades envolvidos nesse processo. Em tarefas de classificação textual, grande parte do tempo é utilizado na etapa de pré-processamento dos dados. Por exemplo, grandes quantidades de textos informais resultam em bases heterogêneas e prejudicam a capacidade de

generalização dos dados aumentando o espaço dimensional e exigindo que estratégias de pré-processamento sejam aplicadas.

Um problema relacionado ao processamento de textos é a alta dimensionalidade dos dados, ou seja, de implica em trabalhar com um grande volume de dados. Por menor que seja a coleção de dados utilizada para a mineração de textos, esta pode facilmente conter milhares de termos, muitas vezes repetidos ou desnecessários, tornando o processo de extração de conhecimento lento e prejudicando a qualidade dos resultados. Para contornar esse problema, deve-se ter uma estratégia de pré-processamento bem definida, que vise a redução da dimensionalidade dos dados, facilitando a montagem de uma representação estruturada para a classificação do sentimento.

Grande maioria das publicações textuais em redes sociais e websites de avaliações é dada por meio de textos curtos, limitando a quantidade de caracteres das mesmas. Isso pode causar dificuldades na hora do processamento dessas publicações e/ou avaliações. Um exemplo decorrente disso é a variação ortográfica de palavras e o uso de gírias, também devido ao fato de estarem em um ambiente informal. Ainda nesse cenário ocorre a ambiguidade, pois algumas palavras precisam de um contexto um pouco mais extenso para que possam ser corretamente classificadas. A ambiguidade consiste em uma palavra que possui mais de um sentido, podendo até expressar sentimentos opostos (LIU, 2012).

Outra característica que pode dificultar a compreensão da base analisada é o Reconhecimento de Entidades Nomeadas (do inglês NER - *Named Entity Recognition*), uma subárea de estudo do Processamento de Linguagem Natural que visa identificar e classificar entidades em um texto não estruturado (PIRES et. al, 2017). Dado o fato que substantivos, muitas vezes, são formados por palavras presentes no vocabulário cotidiano do português, os métodos de classificação podem falhar ao reconhecer algumas entidades. Caso o texto analisado possua sentimentos diferentes para alvos distintos, não será possível distinguir qual sentimento pertence a cada alvo. A avaliação *online* de um hotel a seguir exemplifica essa situação: “O café da manhã é bom, mas o banheiro é ruim”.

Muitas vezes as avaliações também apresentam sarcasmo e/ou ironia o que dificulta a detecção do sentimento pois causa a inversão do sentimento expressado. Segundo Maynard (2014), sarcasmo e ironia são utilizados quando pessoas desejam

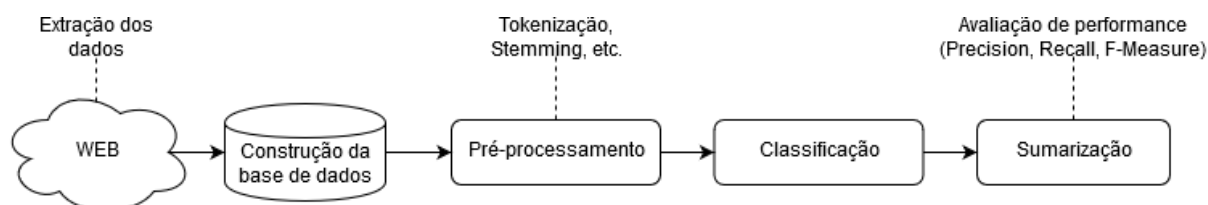


dizer o contrário de seus verdadeiros sentimentos, com objetivo de serem engraçados ou enfatizar determinado ponto.

## 2.3 ETAPAS DO PROCESSO DE ANÁLISE DE SENTIMENTO

O processo de análise de sentimento possui várias etapas, que podem variar de acordo com o método, ferramenta ou mesmo o objeto de estudo. Por exemplo, a etapa de pré-processamento em uma base de dados de texto formal pode exigir menos esforço que uma base informal. Com base na classificação de Kanakalakshmi e Chezian (2015) a Figura 1 apresenta as etapas do processo de AS.

**Figura 1 – Etapas gerais do processo de análise de sentimento**



**Fonte: Adaptado de Kanakalakshmi e Chezian (2015)**

### 2.3.1 Coleta de Dados

O primeiro passo para a análise de sentimentos consiste em coletar dados. Esses podem ser provenientes de *blogs*, fóruns, e redes sociais. Estes dados estão desorganizados, expressos em diferentes formas e fazendo uso de diferentes vocabulários, gírias e contexto de escrita, o que torna a análise manual praticamente impossível (D'ANDREA et al., 2015).

A obtenção de dados relevantes e que contenham opiniões objetivas a respeito de determinado alvo apresenta certa dificuldade em relação à determinada fonte de informações utilizada. Por exemplo, em jornais, *blogs* e *posts* não se conhece a priori as entidades envolvidas, podendo referenciar mais de uma entidade no mesmo segmento de texto (BECKER, TUMITAN, 2013). Se a coleta de informações não for direcionada a alvos previamente definidos, pode-se utilizar técnicas de reconhecimento de entidades nomeadas para a recuperação de informações (AGGARWAL; ZHAI, 2012) (SARAWAGI, 2008).

A complexidade desta tarefa de identificação do alvo da opinião pode variar baseado na mídia utilizada e de seu grau de estruturação. A aplicação mais frequente em mineração de opiniões é a de revisão de produtos e serviços onde o alvo pode ser identificado com mais facilidade. Assume-se que toda opinião expressa no documento se refere diretamente a uma única entidade, o alvo da revisão. Caso a análise ocorra num nível de granularidade menor, o desafio passa a ser a identificação dos aspectos dessa entidade (BECKER, TUMITAN, 2013). No exemplo de um estudo realizado em avaliações de hotéis, os aspectos poderiam ser acomodações, banheiro, café da manhã, etc.

Este processo também pode envolver o discernimento entre texto subjetivo (opinião) ou objetivo (fato), com o objetivo de melhorar os resultados da etapa seguinte, sendo que isto é comum quando o nível de análise de granularidade é menor. O critério utilizado para determinar se um conteúdo é subjetivo consiste na identificação de classes de palavras candidatas a expressar sentimentos, como por exemplo adjetivos (SILVA, 2016).

Uma das ferramentas frequentemente utilizadas para a tarefa de extração de dados são os *web crawlers*. Estes são definidos como *softwares* ou rotinas automatizadas que fazem buscas pela *web* em diversos tipos de domínios a fim de criar índices de conteúdo (AGUIAR; RODRIGUEZ, 2014).

### 2.3.2 Construção da Base de Dados

Para realizar a Análise de Sentimento é necessário rotular os dados obtidos por meio da Coleta de Dados em relação ao sentimento expresso no texto. Essas tarefas de rotular as avaliações demandam tempo e cuidado, além de normalmente ser uma tarefa executada por humanos.

Como a Análise de Sentimentos envolve técnicas de Aprendizado de Máquina, sendo que o Aprendizado Supervisionado necessita de um conjunto de dados para treinar um classificador, é necessário que as avaliações estejam rotuladas antes da classificação.

Sabendo que cada aplicação demanda uma quantidade de dados para o treinamento, não existe um número definido, mas é importante garantir que seja uma quantidade significativa de dados rotulados para cada classe do classificador.

### 2.3.3 Pré-processamento

O pré-processamento dos textos é um dos processos mais importantes da Análise de Sentimentos. Durante o pré-processamento são aplicadas diversas técnicas para remover conteúdos irrelevantes e alterar o formato de algumas palavras. As técnicas a serem aplicadas dependem diretamente dos dados a serem processados. Entre essas técnicas de pré-processamento pode-se citar (FELIX, 2016):

- Realizar a correção ortográfica das palavras, auxiliando no desempenho de outras ferramentas de processamento, tais como *Parsers* e *Taggers*.
- Converter as letras maiúsculas em minúsculas, com a finalidade de padronizar os dados.
- Remover os caracteres não alfabéticos e a pontuação, pois não agregam valor a classificação.
- Remover a acentuação das palavras, a fim de padronizar os dados.
- Aplicar a técnica de *stemming*. *Stemming* é uma técnica para separar o radical e a terminação de uma palavra, como por exemplo correr e corrida, no final o radical utilizado será corr.
- Remover as *Stop Words*, utilizando um dicionário de *stop words*, remove-se essas palavras do conteúdo, pois não são relevantes ou interessantes para o treinamento.
- Lematização, consiste em substituir as palavras pela sua forma canônica, sendo o singular masculino em substantivos e o infinitivo em verbos. Assim como o *Stemmer*, a lematização auxilia na redução da dimensionalidade dos atributos.
- Tokenização, consiste em quebrar o texto em um *array* de palavras. Por exemplo, "Café é muito bom" para ["Café", "é", "muito", "bom"].

Para a execução dessas técnicas são utilizadas bibliotecas que realizam Aprendizado de Máquina como por exemplo a ferramenta NLTK, do inglês *Natural Language Toolkit* (). A seção 2.5.1 apresenta mais detalhes sobre a mesma.

### 2.3.4 Classificação

Esta etapa consiste em classificar um objeto de texto em uma ou mais categorias pré-definidas, que podem variar de acordo com a aplicação. O algoritmo utilizado para a tarefa de classificação pode representar os objetos de texto nas mais variadas categorias, desde que sejam consideradas relevantes para o enriquecimento da representação textual destes, o que permite uma análise de texto ainda mais eficaz e profunda. As categorias podem ainda ser utilizadas para organizar os dados, facilitando sua consulta posterior (ZHAI; MASSUNG, 2016).

A análise de classificação de sentimentos é frequentemente um problema de classificação binário, ou seja, geralmente o classificador é responsável por rotular o sentimento de uma avaliação como “positivo” ou “negativo”. Contudo, classes adicionais podem ser consideradas para que a análise seja mais robusta, ou para aumentar o nível de detalhe dos resultados (BECKER; TUMITAN, 2013). Quando o texto analisado não representa um sentimento ou não é possível identificá-lo com precisão, pode-se fazer uso da classe “neutro”. Por exemplo, no caso da reprodução de uma notícia, onde um fato geralmente é expresso sem que uma opinião seja emitida a respeito do mesmo. A polaridade pode ser expressa ainda por meio de diferentes graus de intensidade, por exemplo de -5 para muito negativo, até 5 para muito positivo (TSYTSARAU; PALPANAS, 2012). Na etapa de classificação é atribuída a polaridade ou orientação da opinião e, por isso, pode ser considerada a etapa mais importante da análise de sentimentos (BECKER, 2017).

Para a realização dessa distinção em classes, são utilizados algoritmos presentes na literatura. Quando se trata de uma classificação realizada por meio de Aprendizagem de Máquina, algoritmos de aprendizagem supervisionada têm sido utilizados (FELIX, 2016). Um algoritmo de aprendizado é considerado supervisionado quando um conjunto de dados é usado para realizar o treino do classificador (BAEZA-YATES; RIBEIRO-NETO, 2013).

Entretanto, existem outras formas de executar a classificação textual. Por exemplo, com a utilização de léxicos de palavras, onde não é necessário realizar o treinamento de um classificador. Estas e outras abordagens são organizadas em diferentes grupos pelos autores. Na Seção 2.4 deste trabalho será discutida a classificação mais comumente encontrada na literatura.

### 2.3.5 Sumarização e Validação do Classificador

Com o modelo de classificação construído, faz-se importante analisar a sua capacidade preditiva utilizando diferentes amostras de dados e métricas de validação. Segundo Liu (2012), a opinião de uma única pessoa a respeito de um determinado tópico ou entidade não é suficiente para representar a opinião geral ou prevaiente dentro de um grupo de indivíduos, sendo necessário analisar uma grande quantidade de opiniões. É fundamental a criação de métricas e sumários para quantificar a diversidade de opiniões encontradas a respeito de um mesmo alvo.

A criação destes sumários é útil principalmente para que a enorme quantidade de informações expressas a respeito de diferentes alvos possa ser facilmente compreendida por usuários finais. A forma mais simples de representar os resultados é por meio de gráficos, facilitando a visualização e o entendimento, embora também possam ser apresentados em forma de texto em tabelas.

A aplicação dos sentimentos sumarizados é vasta, pode ser utilizada para realizar a previsão do resultado de eleições (TUMASJAN et al., 2010), analisar o comportamento da bolsa de valores (BOLLEN et al., 2011), medir a arrecadação de bilheteria de filmes (ASUR; HUBERMAN, 2010), embasar a definição de preços baseado no mercado (ARCHAK et al., 2007), entre outras. Entretanto, a classificação simplesmente entre sentimento positivo e negativo não reflete de maneira precisa todo o contexto analisado, e por isso é importante a criação de métricas para representar o sentimento em relação ao alvo. Na prática, várias técnicas utilizam a média entre sentimentos positivos e negativos e em alguns casos não levam em conta o aspecto da polaridade e sim a quantidade de menções a respeito das entidades analisadas (TUMASJAN et al., 2010) (BECKER; TUMITAN, 2013).

Considerando o método e os dados a serem avaliados, pode-se utilizar a estratégia de *10-fold cross-validation* para realizar a amostragem, onde a base de dados é dividida em dez partes aleatórias e cada uma é testada individualmente, utilizando as nove partes restantes como conjunto de treinamento (GIGLIOTTI, 2012).

Após os testes realizados com os classificadores que utilizam a base de dados rotulada como conjunto de treinamento, obtém-se as seguintes medidas, considerando que a base de treino possui as seguintes classes: positivo, negativo e neutro (SCHMITT, 2013).

- Verdadeiros positivos (VP): instâncias positivas corretamente classificadas;

- Verdadeiros negativos (VN): instâncias negativas corretamente classificadas;
- Verdadeiros neutros (VE): instâncias neutras corretamente classificadas;
- Falsos positivos (FP): instâncias negativas ou neutras classificadas como positivas;
- Falsos negativos (FN): instâncias positivas ou neutras classificadas como negativas;
- Falsos neutros (FE): instâncias positivas ou negativas classificadas como neutras.

A partir da contagem das instâncias classificadas nessas classes, cria-se uma matriz de confusão, que é uma forma de visualização de resultados utilizada para avaliar modelos preditivos que utilizem algoritmos de classificação. As linhas da matriz representam as instâncias previstas de uma classe, e as colunas representam as instâncias reais de uma classe. Assim, a diagonal principal indica os valores que foram previstos corretamente e a diagonal secundária representa os valores previstos incorretamente (FONTOURA, 2011).

Com a visualização do Quadro 1 é possível estimar o desempenho das previsões de cada abordagem. Entretanto, para a maioria das aplicações de processamento de linguagem natural, é desejável uma análise mais detalhada de acertos e erros (WEISS et al, 2005). As medidas de *Precision*, *Recall*, *F-measure* e *Accuracy* são interessantes medidas de qualidade de decisões binárias no problema de classificação de textos (SCHMITT, 2013).

**Quadro 1 – Elaboração da matriz de confusão**

<b>Classes</b>	Classe Verdadeira: Positivo (P)	Classe Verdadeira: Negativo (N)	Classe Verdadeira: Neutro (E)
Classe Prevista: Positivo (P)	<b>VP</b>	<b>FP</b>	<b>FP</b>
Classe Prevista: Negativo (N)	<b>FN</b>	<b>VN</b>	<b>FN</b>
Classe prevista: Neutro (E)	<b>FE</b>	<b>FE</b>	<b>VE</b>

Fonte: Adaptado de Schmitt (2013).

*Precision* (precisão) é a proporção de instâncias positivas que foram classificadas corretamente. Em problemas de classificação binária, o resultado pode ser obtido conforme mostra a Equação 1.

$$Precisão = \frac{VP}{VP + FP} \quad (1)$$

Já o *Recall* representa a quantidade classificada corretamente como instâncias positivas, e pode ser obtido pela por meio da Equação 2.

$$Recall = \frac{VP}{VP + FN} \quad (2)$$

As medidas de *precision* e *recall* podem ser enganosas quando examinadas separadamente. Quando uma dessas medidas apresenta um valor alto, geralmente significa que a outra será o oposto. Têm-se utilizado o que é chamado de ponto de equilíbrio para encontrar um valor intermediário entre essas duas medidas (FRAGOS et al., 2005). A *F-measure* é a média harmônica entre *precision* e *recall* e é definida pela Equação 3

$$F\text{-measure} = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (3)$$

A Acurácia (*accuracy*) ou taxa de acerto representa a soma de amostras positivas, negativas e neutras classificadas corretamente dividida pela soma de amostras positivas, negativas e neutras, ou seja, a proporção de classificações corretas. Seu valor pode ser obtido por meio da Equação 4.

$$Acurácia = \frac{VP + VN + VE}{VP + VN + VE + FP + FN + FE} \quad (4)$$

## 2.4 CLASSIFICAÇÃO DE ABORDAGENS DE ANÁLISE DE SENTIMENTOS

A classificação das abordagens mais utilizada na literatura é a de Tsytarau e Palpanas (2012), que abrange aprendizado de máquina, léxico ou dicionário de palavras, estatística e semântica. D'Andrea et al. (2015) as divide em três categorias principais, sendo elas aprendizado de máquina, léxica e híbrida.

É importante conhecer os tipos de abordagens bem como suas vantagens e limitações pois conforme Gonçalves et al. (2013) escolher o método que melhor se

adequa aos propósitos da pesquisa pode melhorar os índices de precisão. Contudo, a combinação desses métodos não implica necessariamente em melhores resultados. Por fim, embora exista uma predominância da abordagem léxica e de aprendizado de máquina, uma técnica não se sobressai sobre a outra em termos de desempenho (BECKER; TUMITAN, 2013).

#### 2.4.1 Abordagem Baseada em Aprendizado de Máquina

Nesta abordagem ocorre o uso predominante de algoritmos de regressão, convergindo para uma solução sofisticada na etapa de classificação que pode ser dividida em duas etapas. A primeira é o aprendizado supervisionado de um modelo a partir de um corpus com dados para treino. E a segunda é a classificação dos dados não conhecidos a partir do modelo treinado (TSYTSARAU; PALPANAS, 2012).

O objetivo do classificador é encontrar informações implicitamente representadas (BECKER; TUMITAN, 2013). Atualmente a maioria das abordagens em Aprendizado de Máquina para Processamento de Linguagem Natural consiste em obter um conjunto de caracteres organizados em estruturas como palavras, sentenças, e parágrafos, e então realizar a aplicação um algoritmo de classificação estatística nessas estruturas (SODERLAND, 2001).

Esta abordagem segue alguns passos genéricos. Primeiramente obtém-se um conjunto de dados (*dataset*), o qual deve ser rotulado com os sentimentos (positivos, negativos ou neutros). Na sequência, cada documento é representado como um vetor de características. Por fim, um classificador é treinado para distinguir entre sentimentos (previamente rotulados) analisando essas características relevantes. O resultado desse treinamento será usado para prever os sentimentos de novos documentos (TSYTSARAU; PALPANAS, 2012).

#### 2.4.2 Abordagem Baseada em Léxico

Abordagem baseada em dicionário utiliza um dicionário, ou léxico, de palavras que possuem uma polaridade definida, seja ela positiva, negativa ou neutra. Para classificar a polaridade de uma palavra encontrada no texto, deve-se anotar as sentenças, definindo a classe gramatical a que cada uma das palavras pertence.



Após, deve-se identificar palavras que expressem algum sentimento e então consultar as mesmas no dicionário a fim de definir sua polaridade. Além disso, pode-se utilizar um léxico contendo a polaridade de cada palavra, medidas em uma escala, onde a definição do sentimento da sentença dá-se por meio da média entre a polaridade das palavras presentes.

#### 2.4.3 Abordagens Estatística e Semântica

Abordagens estatísticas e semânticas segue a premissa que palavras que expressam alguma opinião frequentemente são encontradas juntas em uma sentença. Sendo assim, se uma palavra ocorre frequentemente ao lado de uma palavra positiva, então provavelmente ela também será positiva (BECKER; TUMITAN, 2013). Caso uma palavra esteja presente próxima de uma palavra positiva ou negativa, pode-se assumir que a polaridade entre elas é compartilhada. Entretanto, se uma palavra aparece em igual frequência entre as classes, pode-se classificá-la como neutra. A abordagem Semântica assemelha-se com a Estatística, diferenciando-se somente pelo uso de uma métrica para cálculo da distância entre as palavras (TSYTSAU; PALPANAS, 2012). Neste caso, as palavras que estão semanticamente próximas são classificadas com a mesma polaridade.

### 2.5 MÉTODOS E FERRAMENTAS DE ANÁLISE DE SENTIMENTOS

Diferentes trabalhos da literatura propõem variados métodos de classificação, porém poucos são os estudos que tentam analisar a influência desses métodos na tarefa de Análise de Sentimentos (RIBEIRO et al., 2016). Nesta Seção serão abordados alguns métodos utilizados na análise de sentimentos e que são mais citados na literatura. Estes métodos utilizam diferentes abordagens de classificação, como aprendizado supervisionado e não supervisionado, métodos combinados, utilização de escalas psicométricas, entre outros.

### 2.5.1 Ferramentas

**NLTK:** O *Natural Language Toolkit* (LOSER; BIRD, 2002), ou mais comumente o NLTK, é um conjunto de bibliotecas e programas para processamento simbólico e estatístico da linguagem natural (PLN), escrito na linguagem de programação *Python*. Foi desenvolvido no Departamento de Ciência da Computação e Informação da Universidade da Pensilvânia. O NLTK inclui demonstrações gráficas e dados de amostra. O NLTK tem como objetivo apoiar a pesquisa e o ensino na PLN ou em áreas estreitamente relacionadas, incluindo linguística, ciência cognitiva, inteligência artificial, recuperação de informações e aprendizado de máquina. O NLTK suporta funcionalidades de classificação, tokenização, *stemming*, análise e raciocínio semântico.

**LIWC:** *Linguistic Inquiry and Word Count* (LIWC) é uma ferramenta baseada em análise de texto, onde é feita a contagem de palavras relacionando seu significado de forma psicológica (TAUSCZIK; PENNEBAKER, 2010). Ela possui um dicionário léxico de aproximadamente 4500 palavras e raízes de palavras que fazem parte de oitenta categorias de sentimentos. Todas as palavras de raiva por exemplo, estão incluídas em pelo menos duas categorias que são emoções negativas e emoção geral. As categorias foram definidas e populadas com base em buscas em diversos dicionários como Thesaurus, questionários e listas feitas por pesquisadores. Para a correta definição, três juízes independentes julgaram cada uma das palavras e definiram em qual categoria ela deveria estar. O percentual de acordo entre os juízes ficou entre 93 e 100%. A construção do dicionário começou entre 1992 e 1994 e ainda apresenta alterações e incrementos.

**SentiStrength:** é uma ferramenta para estimar a polaridade de textos pequenos em linguagem informal, para isto utiliza um dicionário léxico rotulado por humanos e é automaticamente aprimorado por aprendizado de máquina (THELWALL, 2013). Seu idioma de origem é o inglês, mas possui também implementações que permitem a classificação em outros idiomas. Utiliza também técnicas de classificação baseada em aprendizado de máquina que permitem seu aprimoramento automaticamente. Também possui um dicionário de termos baseado em uma versão otimizada para redes sociais do dicionário LIWC. Dessa forma é caracterizado como um método híbrido.

Dada uma mensagem, a referida ferramenta classifica sua polaridade dentro de uma escala de [-5, 5], onde -5 representam as pontuações mais negativas e 5 as mais positivas.

**Happiness Index:** O método *Happiness Index* (DODDS; DANFORTH, 2009), é uma escala de sentimentos baseados no dicionário léxico *Affective Norms for English Words* (BRADLEY; LANG, 1999), que consiste em uma coleção de 1034 palavras associadas a dimensões afetivas de valência, excitação e dominância.

O referido método foi criado com base no já consolidado léxico ANEW e, tem como objetivo calcular pontuações com valores entre 1 e 9 para um dado texto, indicando a quantidade de felicidade que existe neste. Seus desenvolvedores calcularam a frequência em que cada palavra do ANEW aparece no texto e então computaram o peso médio encontrado, levando em consideração apenas o sentimento de valência (ou seja, com maior peso médio). A validação do método *Happiness Index* foi feita por meio de sua aplicação em letras e títulos de músicas e mensagens de *blogs*. Como resultado, encontrou-se que níveis de felicidade em letras de músicas tiveram um decréscimo entre 1961 e 2007, mas aumentaram nas amostras extraídas de *blogs*.

**PANAS-t:** é um método léxico adaptado a partir de uma versão expandida de uma escala bastante conhecida na psicologia chamada de *Positive and Negative Affect Schedule* (WATSON; CLARK; TELLEGEN, 1985). Tem como objetivo detectar as flutuações de humor dos usuários no *Twitter* por meio das postagens.

Consiste em uma escala psicométrica baseada em um largo dicionário léxico de palavras associadas a 11 categorias de humor: jovialidade, autoconfiança, serenidade, surpresa, medo, tristeza, culpa, hostilidade, timidez, fadiga e atenção. Foi desenvolvido para detectar qualquer acréscimo ou decréscimo dos níveis de sentimentos ao longo de um período (GONÇALVES; DORES; BENEVENUTO, 2012).

**OpenNLP:** A biblioteca Apache OpenNLP ([opennlp.apache.org](http://opennlp.apache.org), 2004) é um *kit* de ferramentas baseado em aprendizado de máquina para o processamento de texto em linguagem natural. Ele suporta as tarefas mais comuns da PNL, como detecção de idioma, tokenização, segmentação de sentença, marcação de parte do discurso, extração de entidade nomeada, análise e resolução de co-referência. Essas tarefas geralmente são necessárias para criar serviços de processamento de texto mais avançados.

**Opinion Finder:** é uma aplicação de Análise de Sentimento híbrida, que realiza análise de subjetividade, identificando automaticamente quando opiniões, sentimentos, especulações e outros estados pessoais estão presentes no texto. Ele visa identificar frases subjetivas para marcar vários aspectos da subjetividade nessas sentenças, incluindo a origem (titular) da opinião, e palavras que estão incluídas em frases que expressam sentimentos positivos ou negativos.

Este método funciona em dois modos: lote e interativo. No modo lote o sistema leva uma lista de documentos a serem processadas. O modo interativo fornece uma interface que permite ao usuário consultar fontes de notícias online para os documentos a serem processados. Já para o processamento geral do documento, ele primeiramente executa um analisador chamado Sundance (RILOFF; PHILLIPS, 2005), que fornece etiquetas de classe semântica, identifica entidades nomeadas, e combina padrões de extração que correspondem a linguagem subjetiva. Em seguida, o OpenNLP ([opennlp.apache.org](http://opennlp.apache.org), 2004) é utilizado para indexar e dividir a sentença em partes do discurso e etiquetar os dados. Para identificar palavras e frases utiliza um grande léxico baseado em linguagem subjetiva.

**VADER:** Proposto por (HUTTO; GILBERT, 2014) VADER (*Valence Aware Dictionary for Sentiment Reasoning*) é um método para análise de sentimentos desenvolvido para o contexto de mídias sociais.

Para sua construção os autores criaram uma lista de palavras com base em dicionários já bem estabelecidos como LIWC (TAUSCZIK; PENNEBAKER, 2010) e ANEW - *Evaluation of a word list for sentiment analysis in microblogs* (NIELSEN, 2011). Em seguida, foram adicionadas numerosas construções léxicas presentes em *microblogs* tais como *emoticons*, acrônimos e gírias que expressam sentimentos, resultando em 9000 novos candidatos a serem incluídos no dicionário. Em seguida, verificou-se quais destas possuíam realmente aplicabilidade por meio de pontuação atribuída por Turkers variando de -4 (extremamente negativa) a 4 (extremamente positiva).

Por fim, restaram 7000 construções léxicas no dicionário sendo que para a inclusão a média entre as avaliações obtidas com o AMT deveria ser diferente de 0 (neutro) e com desvio padrão entre as pontuações abaixo de um limiar estabelecido pelos autores. A palavra “*okay*”, por exemplo, possui uma pontuação de 0.9, enquanto “*great*” têm o valor 3.1 e “*horrible*” é -2.5 (BENEVENUTO; RIBEIRO; ARAÚJO, 2015).

**Pattern.en:** É um pacote da linguagem de programação *python* para trabalhar com processamento de linguagem natural. Um de seus módulos é responsável por inferir o sentimento no texto. Criado para ser rápido ele é baseado em polaridades associadas ao WordNet. Este módulo se chama “*text*” e contém um etiquetador rápido *part-of-speech* para o idioma inglês (que identifica substantivos, adjetivos, verbos, etc. em uma frase), ferramentas para conjugação de verbos e substantivo, singularização e pluralização em inglês, uma interface WordNet; um léxico com 8.500 verbos ingleses comuns e suas formas conjugadas, além de analisar sentimentos. O Pattern.en empacota um léxico de adjetivos (por exemplo, bom, mau, impressionante, irritante, etc.) que ocorrem com frequência em análises de produtos, anotado com as pontuações para uma polaridade de sentimento e subjetividade (SMEDT; NIJS; DAELEMANS, 2014).

A função *sentimento* retorna uma tupla (polaridade, subjetividade) para a sentença dada, com base nos adjetivos que ela contém, onde polaridade é um valor entre -1,0 e +1,0 e subjetividade entre 0.0 e 1.0. A sentença pode ser uma *string*, texto, frase, pedaço de uma frase ou uma palavra. A função positiva retorna “Verdadeiro” se a polaridade de determinada sentença está acima do limite. A função negativa retorna “Falso” se a polaridade de determinada sentença está abaixo do limite. O limite pode ser diminuído ou aumentado, mas no geral 0,1 dá os melhores resultados para análises de produtos. A precisão é de cerca de 75% para críticas de filmes.

**spaCy:** o spaCy (spacy.io, 2015) é uma biblioteca de software de código aberto para processamento avançado de linguagem natural, escrita nas linguagens de programação *Python* e *Cython*. A biblioteca é publicada sob a licença MIT e atualmente oferece modelos estatísticos de redes neurais para inglês, alemão, espanhol, português, francês, italiano, holandês e reconhecimento de entidades nomeadas multilíngue, além de tokenização para vários outros idiomas.

### 2.5.2 Léxicos

**Emoticons:** Os léxicos baseados em *emoticons* são desenvolvidos em sua grande maioria a partir de mensagens coletas de fóruns e mídias sociais (exemplo: Twitter, troca de mensagens via Skype, Yahoo! e Facebook). Um exemplo de léxico utilizado para classificação de polaridade em mensagens retiradas do Twitter foi

desenvolvido por (BENEVENUTO et al., 2013), ele inclui as variações mais comuns para emoticons que expressam polaridade positiva, negativa e neutra.

**AFINN:** É um dicionário léxico que fornece uma lista de palavras em inglês associadas com uma valência afetiva ou pontuação, esta lista inclui palavras com sentimentos, gírias da internet e palavras obscenas. As pontuações (*scores*) das palavras variam em uma escala de -5 indicando um sentimento muito negativo, a 5 indicando um sentimento muito positivo. Atualmente existem duas versões: AFINN-111 versão mais recente com 2477 palavras e frases, e AFINN-96 com 1468 palavras e frases originais em 1480 linhas (NIELSEN, 2011).

**Sentiment140:** Sentiment140-Lexicon ou Sentiment140 (MOHAMMAD; KIRIT-CHENKO; ZHU, 2013b), é um dicionário léxico de palavras associadas a sentimentos positivos e negativos. O método foi criado com uma base de dados que consiste em cerca de 1,6 milhões de *tweets* rotulados como positivo ou negativo.

Os *tweets* estão rotulados como unigramas, bigramas e pares de n-gramas. Cada recurso tem uma pontuação que reflete o quão positiva ou negativa é a mensagem. Se a palavra for vista em contextos mais positivos do que negativos, o resultado é positivo ou o contrário. A magnitude da pontuação é maior quando a distribuição é extremamente positiva, e a magnitude é mais próximo de zero quando a palavra aparece igualmente em ambos os contextos positivos e negativos. As palavras negativas são registradas com semelhantes valores negativos em vez de valores positivos.

**EmoLex:** O EmoLex (MOHAMMAD; TURNEY, 2013), é um léxico criado a partir do *Amazon Mechanical Turk Service*, no qual pessoas foram pagas para classificar os termos. Cada entrada está associada a 8 sentimentos básicos em inglês: *joy, sadness, anger, etc.* A base do EmoLex foi construída utilizando termos do Macquarie Thesaurus e palavras do General Inquirer do Wordnet. As palavras que o compõem incluem alguns substantivos, verbos, adjetivos e advérbios mais frequentes da língua Inglesa. O método tem, não apenas a presença de unigramas, mas também de vários bigramas, que são comumente utilizados em textos.

Em seu funcionamento o EmoLex associa as palavras a sua polaridade e a sua emoção. Por exemplo, a palavra “aberração” é negativa e pertence à categoria de emoção desgosto. Assim é possível não identificar apenas a polaridade da palavra, mas também a sua emoção.

De acordo com os desenvolvedores do método, apesar de algumas diferenças culturais, o léxico tem demonstrado que a maioria das normas afetivos são estáveis em todos os idiomas. Assim, versões do léxico são fornecidas em mais de vinte idiomas.

**Opinion Lexicon:** também conhecido como *Sentiment Lexicon* (HU; LIU, 2004) é um banco de dados léxico disponibilizado publicamente. Têm aproximadamente 6800 palavras, sendo 2006 com orientação semântica positiva e 4783 com negativa. Foi inicialmente construído a partir do Wordnet, uma base de dados de palavras em inglês agrupadas em conjuntos de sinônimos.

Por ser um léxico gerado automaticamente a partir de conteúdos extraídos de mídias sociais, ele contém palavras com erros ortográficos, o que conta como ponto positivo para a tarefa de análise de sentimentos, já que as mensagens postadas em redes sociais como o Twitter tendem a incluir grafias erradas e gírias da Internet (LIU, 2010). Por exemplo, no *Opinion Lexicon* pode-se encontrar ambas as palavras “impressionante” e “impessionante” na lista de palavras positivas e na lista negativa tanto “horrrível” como “orrível”. Essa característica o torna diferente dos demais léxicos, pois está mais atualizado para lidar com expressões em redes sociais e *reviews* de produtos.

**NRC Hashtag Sentiment Lexicon:** consiste em um dicionário de palavras com associações para sentimentos positivos e negativos, vinculadas a oito categorias de emoção: alegria, tristeza, raiva, medo, confiança, desgosto, antecipação e surpresa (MOHAMMAD; KIRITCHENKO; ZHU, 2012). O léxico é distribuído em três arquivos: unigramas-pmilexicon.txt (54129 termos), bigramas-pmilexicon.txt (316531 termos) e pares-pmilexicon.txt (480010 termos). Para criação deste léxico, milhares de *tweets* foram rotulados automaticamente com base na ocorrência de *hashtags* vinculadas a palavras positivas e negativas. A partir da rotulação automática verificou-se com contagens relativamente simples, quais palavras ocorriam com maior frequência em *tweets* positivos ou negativos.

## 2.6 CONSIDERAÇÕES DO CAPÍTULO

As avaliações *online* tornaram-se espaços onde as pessoas comentam sobre experiências anteriores e revelam se tiveram suas expectativas atendidas do serviço

que contrataram e acabam influenciando as opiniões de outras pessoas no momento de realizarem a escolha por determinado produto ou estabelecimento. Portanto, a Análise de Sentimentos é uma ferramenta muito importante na descoberta de opiniões de um grande número de pessoas em relação ao alvo de seus comentários, mesmo possuindo ainda alguns desafios, principalmente na obtenção de resultados satisfatórios em análises feitas na língua portuguesa.

Por se tratar de uma área que pode proporcionar uma grande riqueza de dados relevantes às empresas interessadas nessa análise, tem havido muito investimento e estudos relacionados. Percebeu-se que há uma ausência de estudos com foco em português, visto que experimentos como o *SentiBench* possuem uma comparação bastante completa, porém boa parte das ferramentas analisadas está disponível apenas no inglês (RIBEIRO et al, 2016).



### 3 METODOLOGIA

Neste Capítulo serão apresentados os métodos utilizados para o desenvolvimento desse estudo, cujo objetivo é comparar a efetividade de métodos léxicos disponíveis para o português na tarefa de classificação de sentimentos de uma base de dados montada a partir de avaliações *online* de hotéis. A seção 3.1 apresenta os parâmetros utilizados para a escolha das ferramentas. A seção 3.2 descreve como as informações para a construção da base foram obtidas e unificadas e os critérios utilizados para rotulação das avaliações. A seção 3.3 aborda as etapas de pré-processamento utilizadas com o objetivo de tratar os dados e o processo manual que foi feito para descartar informações irrelevantes para o estudo. Na seção 3.4 serão discutidas as métricas utilizadas para comparação do desempenho das ferramentas utilizadas. Por fim, a Seção 3.5 apresenta as considerações do capítulo.

#### 3.1 ESCOLHA DAS FERRAMENTAS

O principal objetivo deste experimento foi comparar algumas das principais ferramentas capazes de executar a classificação de sentimentos, com foco particular no idioma português. Entretanto, nem todas as ferramentas citadas no capítulo anterior possuem suporte ao idioma, então, dentre as disponíveis, optou-se pelas que possuem maior relevância em trabalhos realizados anteriormente alinhados aos seguintes critérios:

- A ferramenta deve ser totalmente gratuita;
- A ferramenta deve aceitar o fornecimento de um léxico em português.

A abordagem léxica foi escolhida pois ainda é a que fornece mais opções prontas para o uso, e também existe a possibilidade de combinação de diferentes dicionários de palavras. As abordagens baseadas em AM na maioria das vezes exigiam a implementação de algum algoritmo específico para o domínio e este não é o foco desse estudo. Com esses critérios em mente, foram escolhidas duas diferentes ferramentas: spaCy (spacy.io, 2015) e NLTK (LOSER; BIRD, 2002). Ambas usam *python* como linguagem principal e permitem o fornecimento de léxicos em português.

Em conjunto com essas ferramentas foram escolhidos 4 léxicos, com o objetivo de comparar cada um deles em cada ferramenta: Sentilex (CARVALHO;

SILVA, 2015), Unilex (SOUZA; PEREIRA, 2017), WordNet Affect BR (PASQUALOTTI; VIEIRA, 2008) e OpLexicon (HU; LIU, 2004).

### 3.2 COLETA DE DADOS E CONSTRUÇÃO DA BASE

Inicialmente, é necessário escolher as fontes de dados a serem utilizadas, no caso deste trabalho, foram escolhidos os websites de busca *Google Meu Negócio*, *TripAdvisor* e *Booking.com*. O Google Meu Negócio oferece uma API com acesso apenas para pessoas com permissão, com esta API é possível capturar as avaliações *online* de empresas dentro do serviço. Já para o caso do *TripAdvisor* e *Booking.com*, é necessário criar robôs de para a captura de dados, por meio de *Web Scraping*.

Para o *Web Scraping* foi implementado um algoritmo na linguagem *JavaScript*, utilizando o *framework* NodeJS (nodejs.org, 2009) e a biblioteca *Puppeteer*. Essa biblioteca fornece uma API (*Application Programming Interface*) de alto nível para controlar um navegador *Chrome* ou *Chromium*. Por meio dessa biblioteca foi possível acessar o *TripAdvisor.com* e o *Booking.com* e extrair os textos das avaliações *online* presente nestes websites. Esse algoritmo foi executado por aproximadamente 30 dias capturando cerca de 5000 avaliações, e foi hospedado dentro no servidor *Google Cloud Functions*.

Apesar do método escolhido para este estudo ser capaz de inferir uma classificação para as avaliações sem necessidade de treinamento prévio de um modelo, pois se utiliza da polaridade definida para cada palavra dentro do léxico fornecido, foi necessário construir uma base rotulada manualmente para que fosse possível averiguar a assertividade dos métodos posteriormente. Desta forma, foi coletada a maior quantidade possível de avaliações e rotuladas manualmente como “positivo”, “neutro” e “negativo”, representadas respectivamente por 1, 0 e -1.

Devido ao volume da quantidade de dados a serem classificadas, optou-se por inferir apenas a polaridade do sentimento, não destacando sua intensidade, como por exemplo “muito positivo” ou “muito negativo”.

As avaliações foram consideradas positivas quando faziam elogios e não apresentavam nenhum tipo de observação ou ponto de atenção, como demonstra a Figura 2.

**Figura 2 – Exemplo de avaliação positiva**

Ótimo hotel, com instalações novas, higiene impecável, café da manhã delicioso, boa localização, estacionamento incluso e funcionários muito simpáticos e atenciosos!!

Fonte: Autoria Própria.

Quando a avaliação possuía elogios, mas também críticas, esta foi considerada neutra, pelo fato de conter palavras com ambas as polaridades, como demonstra a Figura 3.

**Figura 3 – Exemplo de avaliação neutra**

Boa localização Bom atendimento funcionários corteses. A internet deixa um pouco a desejar

Fonte: Autoria Própria.

Nos casos em que não constava informação suficiente para a atribuição da polaridade, a mesma também foi considerada neutra, conforme exemplo da Figura 4.

**Figura 4 – Segundo exemplo de avaliação neutra**

Participei de um evento na sala de reuniões

Fonte: Autoria Própria.

Por fim, a Figura 5 exemplifica uma avaliação que continha apenas aspectos negativos do estabelecimento.

**Figura 5 – Exemplo de avaliação negativa**

café horrível, internet horrível, atendimento horrível.

Fonte: Autoria Própria.

Ao final do processo de rotulação o resultado foi uma base consistente com uma quantidade significativa de avaliações, como demonstra a Tabela 1 (página 36).

Tabela 1 – Sumarização da base de dados

Sentimento	Quantidade de avaliações
Positivo	3560
Negativo	202
Neutro	408
<b>Total</b>	<b>4170</b>

Fonte: Autoria Própria.

### 3.3 PRÉ PROCESSAMENTO

Embora tenham sido utilizadas duas ferramentas de PLN, as etapas de pré-processamento realizada com ambas foram muito similares, diferenciando-se apenas pela implementação de cada uma delas ao executar tarefas como remoção de *stop words*, onde cada uma possui um *corpus* específico. Estas etapas estão listadas adiante.

- Remoção de números;
- Substituição dos caracteres “\n” por “.”;
- Remoção de pontuação e caracteres não alfabéticos, pois não apresentam conteúdo semântico;
- Remoção do termo “*Translated by Google*”, dado que era um termo frequente na base de dados, pois o website de onde foram retirados realiza a tradução automática das avaliações para inglês.
- Tokenização, para converter as avaliações em vetores de palavras;
- Lematização, onde a palavra é transferida para sua forma canônica. Os léxicos utilizados já se encontram lematizados;
- *Stemming*, separando o radical do restante da palavra. Esta etapa foi aplicada tanto na base de dados quanto aos léxicos fornecidos, melhorando o reconhecimento dos termos contidos nos mesmos;
- Remoção de *Stop Words*, com o objetivo de retirar palavras muito comuns no idioma pois estes não possuem muito valor semântico (FILHO, 2014).

Além do pré-processamento aplicado utilizando as ferramentas também foi necessário realizar uma filtragem manual de algumas avaliações que não continham relação com o cenário estudado, como por exemplo a Figura 6, que demonstra um ruído nos dados, pois não pode ser considerada por nenhuma das classes analisadas.

**Figura 6 – Exemplo de ruído encontrado na base de dados**

Fonte: Autoria Própria.

Embora as avaliações neutras encontradas na base possam ajudar o classificador a aprender a distinguir itens dessa classe, nem todas apresentam conteúdo relevante, como no exemplo da Figura 7, onde o conjunto do texto perderá total significado quando o emoticon for removido durante o pré-processamento. A remoção de avaliações como essa também foi feita manualmente.

**Figura 7 – Exemplo de avaliação sem significado semântico**

Fonte: Autoria Própria.

Apesar da extração de dados ser direcionada para hotéis no Brasil, ocorreram coletas de avaliações realizadas por estrangeiros e que precisaram ser removidas, pois não acrescentam no aprendizado do classificador e também podem interferir no cálculo de acertos, pois os experimentos utilizaram léxicos inteiramente em português, conforme o exemplo da Figura 8.

**Figura 8 – Exemplo de avaliações em outro idioma**

Fonte: Autoria Própria.

### 3.4 SUMARIZAÇÃO

Devido ao fato de que cada léxico utilizado apresenta um formato de saída diferente, é preciso realizar a padronização do arquivo de saída de cada experimento realizado para que seja possível realizar a comparação com a polaridade real das avaliações. O Quadro 2 demonstra o tipo de saída para os léxicos utilizados.

Para a análise dos resultados foi necessário padronizar os resultados de cada algoritmo, ou seja, utilizou-se uma função para transformar todas as saídas dos algoritmos em -1 (Negativas), 0 (Neutras) e 1 (Positivas).

**Quadro 2 – Saídas dos léxicos utilizados**

Léxico	Saída
Sentilex (CARVALHO; SILVA, 2015)	[-8, +8]
Unilex (SOUZA; PEREIRA, 2017)	-1, 0, 1
WordNet Affect BR (PASQUALOTTI; VIEIRA, 2008)	-1, 0, 1
OpLexicon (HU; LIU, 2004)	-1, 0, 1

Fonte: Autoria Própria.

Para a obtenção de métricas de comparação utilizou-se a ferramenta *Scikit-learn* (PEDREGOSA et al, 2011). O *Scikit-learn* é um módulo da linguagem *python* que integra uma ampla gama de algoritmos de aprendizado de máquina de ponta para problemas supervisionados e não supervisionados de média escala.

Com o uso desta biblioteca, as matrizes de confusão foram calculadas utilizando a função a matriz de confusão do módulo "metrics", e o relatório final com os dados de precisão, recall, e f1-score, além da acurácia utilizando a função "classification\_report" do módulo "metrics".

### 3.5 CONSIDERAÇÕES DO CAPÍTULO

Durante o desenvolvimento deste estudo foram aplicadas as técnicas de *Web Scraping* em alguns websites de avaliações com o objetivo de construir uma base de informações relacionada à hotéis. Cada avaliação foi classificada manualmente em classes "positiva", "negativa" ou "neutra".

Foram utilizadas duas ferramentas combinadas com quatro léxicos. Para cada combinação foi realizada a etapa de pré-processamento dos dados com o objetivo de preparar a base para a realização dos experimentos. Após isso os dados foram classificados entre as possíveis classes e suas predições foram comparadas com os dados iniciais avaliados manualmente, possibilitando a contagem de erros e acertos para comparar os métodos entre si.

## 4 RESULTADOS

Este Capítulo apresenta os resultados dos experimentos realizado por meio das ferramentas combinadas com diferentes léxicos. A Seção 4.1 discute a qualidade do pré-processamento realizado. A Seção 4.2 e 4.3 sumarizam a assertividade de cada léxico utilizado em conjunto com as ferramentas NLTK e spaCy, respectivamente por meio de matrizes de confusão. A Seção 4.4 apresenta os resultados obtidos pelas ferramentas e com base em estudos realizados anteriormente referentes a análises semelhantes no idioma inglês. Por fim, a Seção 4.5 apresenta as considerações do Capítulo.

### 4.1 VALIDAÇÃO DO PRÉ-PROCESSAMENTO

Diversas etapas de pré-processamento descritas na Seção 3.3, foram utilizadas para tratar a base de avaliações *online* com o objetivo de excluir termos que não são úteis para a classificação. Ao observar a nuvem de palavras na Figura 9 (página 40), é possível perceber que o pré-processamento foi bem-sucedido em sua maioria, pois a maior parte das palavras que aparecem na nuvem faz referência a termos específicos ao ambiente em geral de hotéis, como acomodações, higiene, refeições, preço, recepção, entre outros.

Pode-se perceber que palavras de natureza positiva como “bom”, “boa”, “excelente” e “ótima” foram bastante frequentes nas avaliações, correspondendo aos dados presentes na Tabela 1, que indica que uma grande maioria dos comentários deixados por usuários tende a ser positiva.

Também é possível perceber que todas as palavras presentes na nuvem estão escritas corretamente, isto é, não falta nenhuma letra e é possível entender seu significado, há a ausência de gírias e abreviações, o que significa que os dicionários utilizados, alinhados ao processo de lematização e *stemming* funcionaram de maneira correta.

Como termo principal pode-se notar a presença de “café da manhã”, o que pode indicar que este seria um fator considerado muito relevante, e também um motivador para usuários deixarem suas impressões da estadia no hotel em sites de busca e avaliação, como os que foram utilizados neste estudo.

Figura 9 – Nuvem de palavras da base após o pré-processamento



Fonte: Autoria Própria.

## 4.2 NLTK

A combinação da ferramenta NLTK e o léxico Sentilex conseguiu uma acurácia de 79,9%, com precisão de 91,1%, e *F-measure* de 80,6%. Os dados foram calculados utilizando a função “classification\_report” da biblioteca *scikit-learn* com base na matriz de confusão representada pela Tabela 2.

Tabela 2 – Matriz de confusão para o léxico Sentilex com NLTK

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	3121	64	241
	Negativo	119	115	73
	Neutro	320	23	94

Fonte: Autoria Própria.

O léxico Unilex apresentou acurácia de 60,8%, com precisão de 75,2% e *F-measure* 66,6% conforme cálculo realizado a partir da matriz de confusão presente na Tabela 3.



Tabela 3 – Matriz de confusão para o léxico Unilex com NLTK

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	2403	110	261
	Negativo	386	78	94
	Neutro	771	14	53

Fonte: Autoria Própria.

Apesar de atingir um valor alto para precisão de 71,0%, o léxico *WordNet Affect BR* obteve acurácia de apenas 27,2% e *F-measure* de 34,0%. A Tabela 4 apresenta a matriz de confusão obtida para o mesmo.

Tabela 4 – Matriz de confusão para o léxico WordNet Affect BR com NLTK

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	855	68	133
	Negativo	91	35	31
	Neutro	2614	99	244

Fonte: Autoria Própria.

E por último, representando a maior assertividade para a ferramenta NLTK, o léxico *OpLexicon* obteve acurácia de 83,8%, com precisão de 80,0% e *F-measure* igual a 81,5%, conforme a matriz de confusão da Tabela 5.

Tabela 5 – Matriz de confusão para o léxico OpLexicon com NLTK

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	3387	136	322
	Negativo	38	51	29
	Neutro	135	15	57

Fonte: Autoria Própria.

#### 4.3 SPACY

Ao utilizar a ferramenta de processamento de linguagem natural Spacy combinada com o primeiro dos léxicos analisado, o *Sentilex*, obteve-se acurácia de 69,3%, com precisão de 80,4% e *F-measure* de 73,8%. Para obtenção dessas

métricas e dos dados presentes na Tabela 6, também foi utilizada a função “classification\_report” da biblioteca *scikit-learn*.

**Tabela 6 – Matriz de confusão para o léxico Sentilex com *spaCy***

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	2675	58	204
	Negativo	169	118	106
	Neutro	716	26	98

Fonte: Autoria Própria.

Ao combinar a ferramenta Spacy com o léxico Unilex, foram obtidos os dados da matriz de confusão da Tabela 7, resultando em precisão de 73,1%, *F-measure* de 54,7% e acurácia de 45,4%.

**Tabela 7 – Matriz de confusão para o léxico Unilex com *spaCy***

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	1760	107	213
	Negativo	608	78	138
	Neutro	1192	17	57

Fonte: Autoria Própria.

Para a combinação do léxico *WordNet Affect BR* e a ferramenta *spaCy*, foi obtida uma precisão de 70,2%, com *F-measure* de 34,3% e, semelhante ao experimento com o NLTK, uma acurácia de apenas 27,4%, indicando que este léxico foi o que apresentou pior desempenho, independente da ferramenta utilizada. A Tabela 8 apresenta os dados da matriz de confusão elaborada para o mesmo.

**Tabela 8 – Matriz de confusão para o léxico WordNet Affect BR com *spaCy***

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	877	79	137
	Negativo	76	23	27
	Neutro	2607	100	244

Fonte: Autoria Própria.

Por fim, a última combinação do método Spacy com o léxico OpLexicon apresentou os dados da matriz de confusão da Tabela 9, com precisão de 77,7%, *F-measure* de 75,7% e acurácia de 74,0%.

**Tabela 9 – Matriz de confusão para o léxico OpLexicon com spaCy**

		Classe verdadeira		
		Positivo	Negativo	Neutro
Classe prevista	Positivo	2959	122	283
	Negativo	72	56	55
	Neutro	529	24	70

Fonte: Autoria Própria.

#### 4.4 COMPARAÇÃO DOS RESULTADOS

Nesta Seção são comparados os resultados obtidos com os experimentos realizados com as duas ferramentas em combinação com 4 diferentes léxicos escolhidos, objetivando mostrar quais métodos tiveram melhor e pior desempenho.

Para tanto, quatro resultados tiveram de ser avaliados: precisão, *recall*, *F-measure* e acurácia. Dentre estes o que apresenta maior relevância para o experimento é o *F-measure* pois é a medida frequentemente utilizada para comparação do desempenho alcançado pelos diferentes métodos de AS. Todos os dados dessa seção foram obtidos por meio das matrizes de confusão apresentadas no capítulo anterior.

Para simplificar a análise dos resultados obtidos, foram elaborados gráficos com o resultado final de cada ferramenta. Também foram elaborados gráficos para facilitar a visualização de qual ferramenta obteve melhor e pior desempenho em cada classe distinta da classificação (negativo, neutro, positivo).

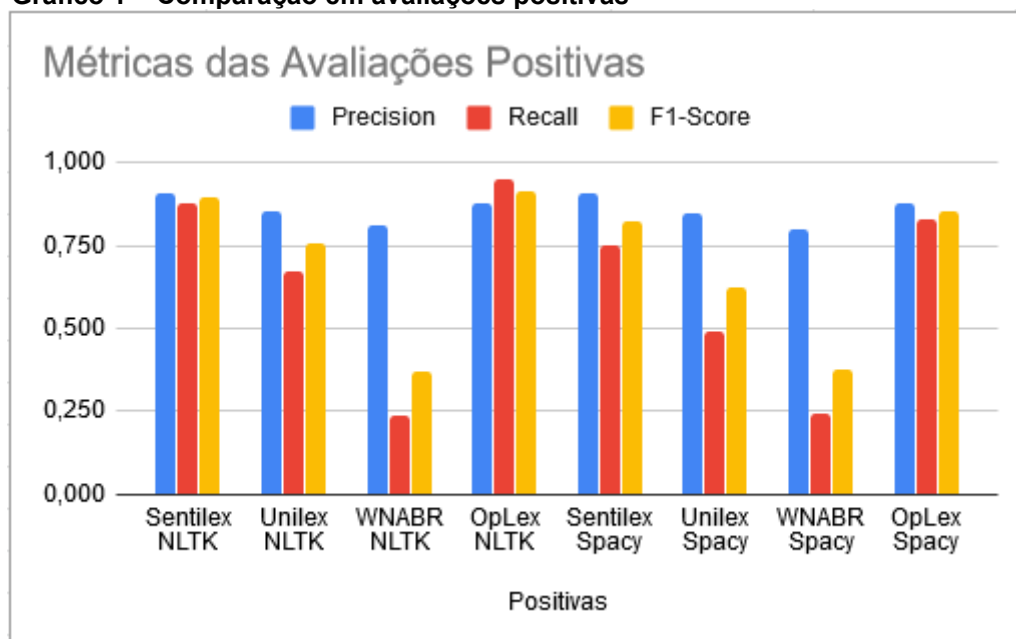
Lembrando que o *recall* é a capacidade do método classificar corretamente as avaliações de uma classe de polaridade específica, enquanto a precisão equivale a assertividade do método em classificar corretamente avaliações de determinada classe levando-se em conta todas as avaliações classificadas corretamente e incorretamente para tal classe. Já o *F-measure* é a média harmônica entre precisão e *recall*, utilizada para medir a capacidade geral de classificação do método para determinada classe.

Em relação à cada gráfico apresentado a seguir, o eixo Y representa uma escala equivalente a porcentagem alcançada por cada método em cada métrica de avaliação, ou seja, o desempenho de classificação do método. Já o eixo X representa cada método de análise de sentimento testado.

Considerando-se 3560 avaliações positivas, conforme apresentado no Gráfico 1, o método que obteve maior precisão foi o Sentilex tanto quando utilizado junto com o NLTK tanto com o spaCy, com uma taxa de acerto de 91,1%. Enquanto o método com maior *recall* foi o OpLexicon e NLTK com percentual de 95,1% de acerto. Em termos de desempenho geral, considerando o *F-measure* o método com o melhor desempenho também foi o OpLexicon utilizado junto com o NLTK.

Inversamente a estes resultados, o método com o pior desempenho em todos os indicadores foi o WordNet Affect BR, com precisão de 80,2%, *recall* de 24,6% e *F-measure* de 37,7%.

**Gráfico 1 – Comparação em avaliações positivas**

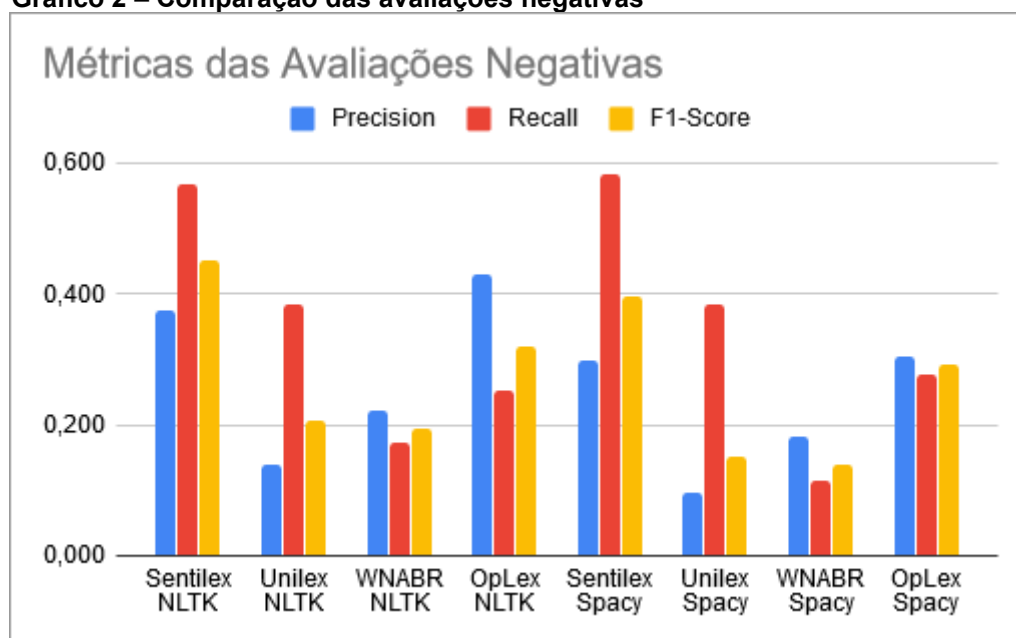


Fonte: Autoria Própria.

Analisando o cenário de 202 avaliações classificadas manualmente como negativas, segundo o Gráfico 2, o melhor desempenho para precisão foi do OpLexicon e NLTK, com percentual de 43,2%. O maior *recall* foi atingido pelo Sentilex e Spacy com percentual de 56,9%. No panorama geral, o maior *F-measure* foi registrado também pelo Sentilex e NLTK.

Em relação aos menores percentuais, a precisão mais baixa foi do método Sentilex e spaCy com apenas 9,5% de acerto. O menor *recall* foi registrado pelo WordNet Affect BR e spaCy com 11,4% de acerto. O menor desempenho no geral foi também do WordNet Affect BR e spaCy com 14,0% de *F-measure*.

**Gráfico 2 – Comparação das avaliações negativas**



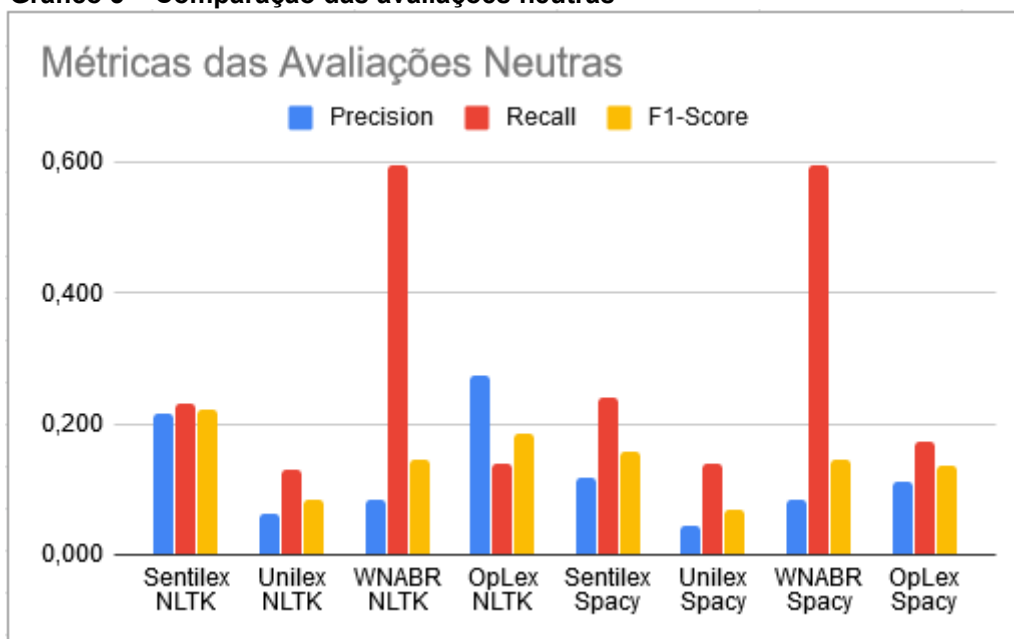
Fonte: Autoria Própria.

Para o total de 408 avaliações neutras, os melhores resultados obtidos foram os seguintes, de acordo com o Gráfico 3 (página 46).

- Precisão: OpLexicon e NLTK com taxa de acerto de 27,5%;
- *Recall*: WordNet Affect BR com ambas as ferramentas com taxa de acerto de 59,8%;
- *F-measure*: Sentilex e NLTK com assertividade de 22,2%.

Já os índices mais baixos foram obtidos pelo Unilex e spaCy com apenas 8,3% de precisão e 14,0% de *recall* e 6,8% de *F-measure*.

Gráfico 3 – Comparação das avaliações neutras

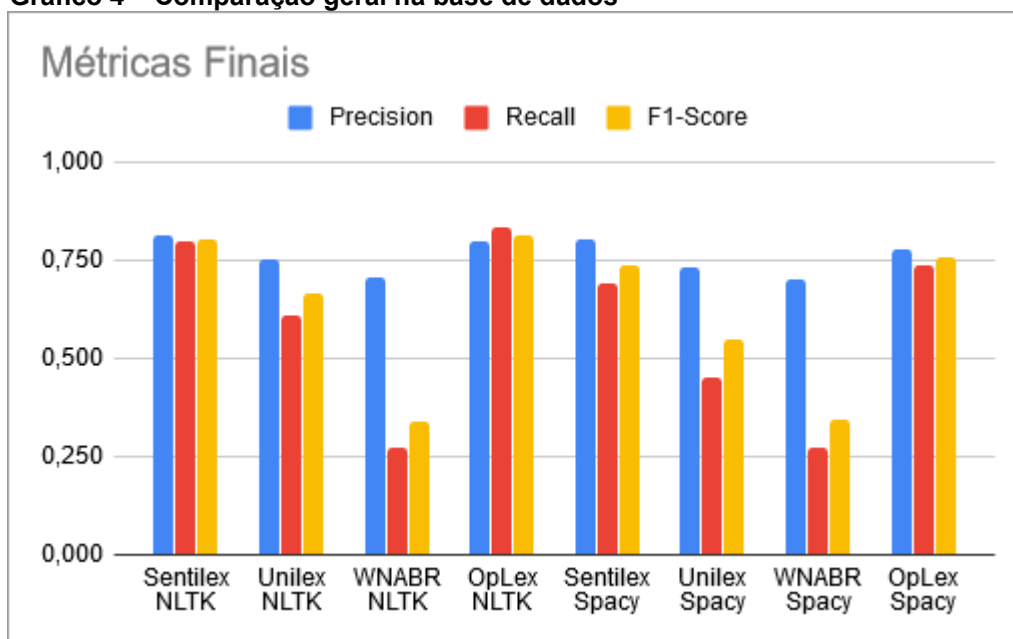


Fonte: Autoria Própria.

Por fim, o resultado geral e mais significativo, pois analisa a base como um todo levando 4170 avaliações em consideração. De modo geral, conforme ilustrado pelo Gráfico 4 (página 47), os índices obtidos foram os seguintes:

- **Precisão:** Melhor desempenho para o método Sentilex e NLTK com taxa de acerto de 81,7%. Pior desempenho para WordNet Affect BR e spaCy com taxa de acerto de 70,2%
- **Recall:** Melhor método foi o OpLexicon juntamente com o NLTK com percentual de 83,8%. O pior método foi o também o WordNet Affect BR e spaCy com acerto de apenas 27,4%.
- **F-measure:** Melhor desempenho atingido por OpLexicon e NLTK com percentual de 81,5% enquanto que o desempenho mais baixo dessa vez foi do WordNet Affect BR junto com o NLTK com percentual de 34,0%.

Gráfico 4 – Comparação geral na base de dados



Fonte: Autoria Própria.

A Figura 10 apresenta uma tabela com os valores detalhados da assertividade geral de cada método e destaca os que obtiveram melhores e piores desempenhos para cada uma das medidas analisadas.

Figura 10 – Resultados gerais das ferramentas

Ferramentas	Precisão	Revocação	F1 Score
Sentilex NLTK	81,7%	79,9%	80,6%
Unilex NLTK	75,2%	60,8%	66,6%
WNABR NLTK	71,0%	27,2%	34,0%
OpLex NLTK	80,0%	83,8%	81,5%
Sentilex spaCy	80,4%	69,3%	73,8%
Unilex spaCy	73,1%	45,4%	54,7%
WNABR spaCy	70,2%	27,4%	34,3%
OpLex paCy	77,7%	74,0%	75,7%

Fonte: Autoria própria.

Verifica-se com estes resultados que a análise em uma base com maior quantidade de dados possibilitou melhor desempenho do algoritmo, dados que a quantidade de termos aumenta proporcionalmente com o aumento dos dados possibilitando melhor correlação com o léxico e melhor assertividade como consequência.

#### 4.5 CONSIDERAÇÕES DO CAPÍTULO

Neste capítulo foi realizada uma análise na base depois de realizado o pré-processamento, com o objetivo de validar a qualidade do mesmo. O resultado obtido foi satisfatório, visto que a maioria dos termos que constaram após a execução dessa etapa apresentam significado semântico, o que ajuda o classificador na hora de aferir a polaridade da avaliação, pois encontra mais termos significativos e que correspondem a outros termos do léxico fornecido.

Após a avaliação dos oito métodos propostos neste estudo por meio da elaboração de matrizes de confusão, cálculo da precisão, cálculo do *recall*, e o cálculo do *F-measure* de cada uma delas, foi possível identificar a combinação da ferramenta NLTK junto com o léxico OpLexicon como a que obteve o melhor desempenho para o estudo de caso.

A partir dos resultados obtidos é possível concluir que a aplicação de uma ferramenta confiável para a tarefa de NLP em conjunto com um léxico grande o suficiente para o idioma utilizado, pode gerar bons resultados, inclusive com taxas de assertividade bastante próximas de tarefas tradicionais de processamento de linguagem natural e também de análise de sentimentos realizadas em inglês.



## 5 CONCLUSÃO

Neste trabalho buscou-se avaliar o desempenho quantitativo de duas ferramentas que possibilitam a análise de sentimentos em português combinadas com quatro diferentes léxicos, aplicando as mesmas em uma base de dados obtida a partir de avaliações *online* de hotéis em websites de reservas.

Para a realização deste estudo primeiramente foi necessário o levantamento de website que disponibilizam avaliações *online* de hotéis e em seguida a aplicação de métodos de *web scrapping* para a extração desses dados e unificação em uma base de dados com cerca de quatro mil e quinhentos registros. Ao final desta etapa o primeiro objetivo específico foi atingido visto que o resultado foi uma base com domínio voltado para o tema de estudo.

O segundo objetivo específico realizado foi a aplicação de técnicas gerais de tratamento de dados utilizadas em processos de análise de sentimentos, com foco no pré-processamento e classificação, etapas que podem ser consideradas as mais importantes. Para tanto, foi necessária uma pesquisa na literatura em geral para obter as melhores práticas e com pontos de vista diferentes de vários autores.

Ao realizar a aplicação de métricas de comparação nos resultados obtidos para as ferramentas selecionadas, o objetivo principal deste trabalho foi atingido, possibilitando a consulta dos dados para trabalhos futuros que visem a comparação de outras ferramentas e técnicas, ou até mesmo a utilização das mesmas ferramentas em outro domínio.

Levando em consideração a métrica de *F-measure*, foi possível constatar que dentre os 8 métodos analisados, o que teve melhor capacidade de classificação correta das avaliações foi o OpLexicon juntamente com o NLTK. Um ponto importante observado foi que as ferramentas analisadas possuem uma assertividade relativamente alta quando se trata de classificação de avaliações positivas, porém para uma melhoria da classificação de avaliações neutras e negativas é necessário obter uma base de dados mais diversificada.

Este estudo tem uma elevada importância tratando-se de organizações e empresas que utilizam os resultados gerados por métodos de análise de sentimentos para auxílio no processo de tomada de decisão, pois oferece um panorama das ferramentas já disponíveis para o português e o que ainda precisa ser desenvolvido.

Como limitações deste estudo, pode-se citar o fato da prevalência de avaliações positivas perante as negativas e neutras na base construída. E também maior complexidade para a implementação de algoritmos que utilizem o método de aprendizado de máquina, resultando no foco do estudo para abordagens léxicas.

## 5.1 TRABALHOS FUTUROS

Como trabalhos futuros pretende-se realizar a análise do desempenho de ferramentas de aprendizado de máquina, comparando com os resultados obtidos neste trabalho e na literatura em geral em relação a outros idiomas estudados.

A metodologia aplicada neste estudo também pode ser aplicada para futuras avaliações de escopos diferentes de avaliações de hotéis, como por exemplo, outras situações em que se deseja obter um panorama sobre a opinião de usuários em uma base extensa para avaliação manual.

## REFERÊNCIAS

AGUIAR, R. F.; RODRIGUEZ, W. F. G. **SOMTOOLS**: Sentiment and Opinion Mining tool for Twitter Data Stream. 2014. 112 f. Monografia (Graduação) – Universidade de Federal do Sul e Sudeste do Pará. Marabá, 2014.

ARCHAK, N.; GHOSE, A.; IPEIROTS, P. G. **Show me the money!**: deriving the pricing power of product features by mining consumer reviews: Simpósio Brasileiro de Banco de Dados, 2007.

AROSA, B. **Análise de Sentimento em textos curtos baseada em Processamento de Linguagem Natural e Aprendizado de Máquina**. 2018. 81 f. Tese (Doutorado) – Departamento de Eletrônica e Computação, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2018.

ASUR, S.; HUBERMAN, B. A. Predicting the future with social media: Web Intelligence and Intelligent Agent Technology (WI-IAT) IEEE/WIC/ACM **International Conference**, v. 1, p. 492–499, 2010.

BECKER, Willian Eduardo. **Uma abordagem de redes neurais convolucionais para análise de sentimento multi-lingual**. 2017. 85 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, Pontifícia Universidade Católica do Rio Grande do Sul. Porto Alegre, 2017.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock Market. **Journal of Computational Science**, p. 1–8, 2012.

CHEZIAN, R. M.; KANAKALAKSHMI, C. Performance Evaluation of Machine Learning Techniques for Text Classification. In: PROCEEDINGS OF THE UGC SPONSORED NATIONAL CONFERENCE ON ADVANCED NETWORKING AND APPLICATIONS, 2015.

DENECKE, Kerstin. Using SentiWordNet for multilingual sentiment analysis: 2008 In: IEEE INTERNATIONAL CONFERENCE ON DATA ENGINEERING WORKSHOP, p. 507-512, 2008.

D'ANDREA, A.; FERRI, F.; GRIFONI, P.; GUZZO, T. Approaches, Tools and Applications for Sentiment Analysis Implementation. **International Journal of Computer Applications**, Roma (Itália), v. 125, n. 3, p. 26–33, set. 2015.

FELDMAN, R; SANGER, J. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**. 1. ed. New York: Cambridge University Press, 2007.

FRAGOS, K.; MAISTROS, Y.; SKOURLAS, C. A weighted Maximum Entropy Language Model for Text Classification. In: PROCEEDINGS OF THE 2ND INTERNATIONAL WORKSHOP ON NATURAL LANGUAGE UNDERSTANDING AND COGNITIVE SCIENCE (NLUCS). p. 55–67, 2005.

FONTOURA, V. D. da. **Predição de falhas em projetos de software livre baseada em métricas de redes sociais**. 2011. 51 f. Trabalho de Conclusão de Curso (Graduação) – Universidade Tecnológica Federal do Paraná, Campo Mourão, 2011.

GIGLIOTTI, W. **Em Busca de Más Notícias**. 2012. 37 f. Monografia (Graduação). Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo, 2012.

GINSCA, A.; BOROS, E.; IFTENE, A.; TRANDABĂȚ, D.; TOADER, M.; CORICI, M.; PEREZ, C.; CRISTEA, D. Sentimatrix – Multilingual Sentiment Analysis Service. In: 2ND WORKSHOP ON COMPUTATIONAL APPROACHES TO SUBJECTIVITY AND SENTIMENT ANALYSIS, p. 189–195, **Association for Computational Linguistics**, 2011.

GOLDENBERG, J.; LIBAI, B.; MULLER, E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. **Marketing Letters**, v. 12, n. 3, p. 211–223, 2001.

GONÇALVES, P.; ARAÚJO, M.; BENEVENUTO, F. Comparing and Combining Sentiment Analysis Methods. In: PROCEEDINGS OF THE FIRST ACM CONFERENCE ON ONLINE SOCIAL NETWORKS, **ACM**, p. 27–38, 2013.

HAYKIN, S. **Redes Neurais: Princípios e Práticas**; trad. Paulo Martins Engel. 2 ed. Porto Alegre: Bookman, 2008.

LOPER, E.; BIRD, S. NLTK: the natural language toolkit. In: PROCEEDINGS OF THE ACL WORKSHOP ON EFFECTIVE TOOLS AND METHODOLOGIES FOR TEACHING NATURAL LANGUAGE PROCESSING AND COMPUTATIONAL LINGUISTICS, 2002, Filadélfia. **Association for Computacional Linguistics**, 2002.

MAYNARD, D.; GREENWOOD, M. A. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In: LANGUAGE

RESOURCES AND EVALUATION CONFERENCE (LREC), 2014. Reykjavik, Iceland, 2014. p. 26–31.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. **Ain Shams engineering journal**, v. 5, n. 4, p. 1093–1113, 2014.

MOZETIČ, Igor; GRČAR, Miha; SMAILOVIĆ, Jasmina. Multilingual twitter sentiment classification: The role of human annotators. **PloS one**, v. 11, n. 5, p. e0155036, 2016.

MURPHY, H.; CHEN, M. The multiple effects of review attributes on hotel choice decisions: A conjoint analysis study. **Ecole hôtelière de Lausanne**, 2014. Disponível em: <[http://agrilife.org/ertr/files/2014/02/enter2014\\_RN\\_17.pdf](http://agrilife.org/ertr/files/2014/02/enter2014_RN_17.pdf)>. Acesso em: 08 de maio de 2019.

MURPHY, Rosie. Local Consumer Review Survey 2018. **BrightLocal. Retrieved**. v. 19, n. 3, 2018. Disponível em: <<https://www.brightlocal.com/research/local-consumer-review-survey/>>. Acesso em: 08 de maio de 2019.

NIELSEN, F. Å. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In: PROCEEDINGS OF THE ESWC2011 WORKSHOP ON “MAKING SENSE OF MICROPOSTS”: BIG THINGS COME IN SMALL PACKAGES, 2011. **arXiv preprint arXiv:1103.2903**, 2011, p. 93–98.

OLIVEIRA, F. W. C. de. **Análise de sentimentos de comentários em português utilizando SENTIWORDNET**. 2013. Monografia (Especialização) – Departamento de Informática, Universidade Estadual de Maringá. Maringá, 2013.

PAN, B.; MACLAURIN, T.; CROTTS, J. C. Travel Blogs and the Implications for Destination Marketing. **Journal of Travel Research**, v. 46, p.35–45, 2007.

PANG, Bo; LEE, Lillian. **Opinion mining and sentiment analysis: Foundations and Trends in Information Retrieval**. 2 ed. Now Publishers Inc, 2008.

PIRES, A. S.; DEVEZAS, J. L.; NUNES, S. **Benchmarking Named Entity Recognition Tools for Portuguese**. Faculdade de Engenharia, Universidade do Porto. Porto, 2017.

SALTER, V. L. **Decision Support Systems For Business Intelligence**. 2 ed. WILEY, 2011.

SCHMITT, V. F. **Uma Análise Comparativa De Técnicas De Aprendizagem De Máquina Para Prever a Popularidade De Postagens No Facebook**. 2013. 57 f. Trabalho de Conclusão de Curso (Graduação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2013.

SILVA, P. S. **Avaliação do Desempenho de Métodos de Análise de Sentimentos na Presença das Figuras de Linguagem Sarcasmo e Ironia**. 2016. 115 f. Trabalho de Conclusão de Curso (Graduação) – Instituto de Geociências e Engenharia, Universidade Federal do Sul e Sudeste do Pará. Marabá, 2016.

TUMASJAN, A.; SPRENGER, T.; SANDNER, P.; WELPE, I. Predicting elections with twitter: What 140 characters reveal about political sentiment: In PROCEEDINGS OF THE FOURTH INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA. **AAAI Publications** p. 178–185, 2010.

VARGIU, E.; URRU, M. Exploiting web scraping in a collaborative filtering based approach to web advertising. **Artificial Intelligence Research**, v. 2, n. 1, 2013.

YE, Q.; ZHANG, Z.; LAW, R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. **Expert System with Applications**, v. 36, p. 6527–6535, 2009.