

**UNIVERSIDADE NOVE DE JULHO - UNINOVE**  
**Programa de Mestrado em Engenharia de Produção**

**COMPARANDO MODELOS DE CLASSIFICAÇÃO DOS COMENTÁRIOS DE  
CLIENTES: MINERAÇÃO DE OPINIÕES EM EMPRESA BRASILEIRA DE  
CLASSIFICADOS ONLINE DE EMPREGOS**

**MARCELO DRUDI MIRANDA**

**SÃO PAULO**  
**2014**

**MARCELO DRUDI MIRANDA**

**COMPARANDO MODELOS DE CLASSIFICAÇÃO DOS COMENTÁRIOS DE  
CLIENTES: MINERAÇÃO DE OPINIÕES EM EMPRESA BRASILEIRA DE  
CLASSIFICADOS ONLINE DE EMPREGOS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Nove de Julho, parte dos requisitos exigidos para a obtenção do grau de Mestre em Engenharia de Produção.

Prof. Renato José Sassi, Dr.–Orientador, UNINOVE

**SÃO PAULO**

**2014**

Miranda, Marcelo Drudi.

Comparando modelos de classificação dos comentários de clientes: mineração de opiniões em empresa brasileira de Classificados online de empregos. /Marcelo Drudi Miranda. 2014.

82 f.

Dissertação (mestrado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2014.

Orientador (a): Prof. Dr. Renato José Sassi.

1. Mineração de opiniões. 2. Análise de sentimentos. 3. Modelos de classificação. 4. Classificação de comentários em língua portuguesa.

I. Sassi, Renato José.

II. Título

CDU 658.5

São Paulo, 28 de agosto de 2014.

TERMO DE APROVAÇÃO

Aluno: Marcelo Drudi Miranda

Título da Dissertação: COMPARANDO MODELOS DE CLASSIFICAÇÃO DOS COMENTÁRIOS DE CLIENTES: MINERAÇÃO DE OPINIÕES EM EMPRESA BRASILEIRA DE CLASSIFICADOS ONLINE DE EMPREGOS.

Presidente: PROF. DR. RENATO JOSÉ SASSI



Membro: PROF. DR. MARCIRIO SILVEIRA CHAVES



Membro: PROF. DR. JOSÉ CARLOS CURVELO SANTANA



Dedico este trabalho aos meus filhos Lucas e Lorena, pois foram o motivo de eu ter seguido em frente, mesmo diante das dificuldades.

## AGRADECIMENTOS

Agradeço primeiramente à minha mãe, Darli Dalva Drudi Miranda, e ao meu pai, Alvaro Miranda Junior, por todo o esforço, todo incentivo e pela base moral que me proporcionaram, forjada através de seu próprio exemplo de vida.

Ao meu orientador, Professor Dr. Renato José Sassi, meus agradecimentos pela sua dedicação à pesquisa e ao ensino, pela sua excelente orientação, pelas cobranças e pelo trabalho árduo por ele realizado, que me fez crescer não só como pesquisador, mas também como ser humano.

Agradeço à minha esposa, Estela de Oliveira Baltieri, pela paciência e compreensão pelas inúmeras noites e finais de semana em que não pude ficar ao seu lado, quando eu estava mentalmente ou fisicamente ausente, me dedicando a este trabalho.

Agradeço aos docentes do programa de Mestrado em Engenharia de Produção e funcionários da Universidade Nove de Julho.

Aos membros das bancas de qualificação e defesa: Prof. Dr. José Carlos Curvelo Santana, Prof. Dr. Marcirio Silveira Chaves e Prof. Dr. Leonardo Junqueira, agradeço pelas contribuições importantes a esse trabalho.

Em especial, gostaria de agradecer ao Prof. Dr. Marcirio Silveira Chaves pelo incentivo, pela disposição em ajudar, e pelas excelentes contribuições, que ajudaram a nortear a pesquisa.

Agradeço aos colegas do grupo, que me receberam muito bem e sempre estiveram dispostos a ajudar. Em especial agradeço ao amigo Edquel Bueno Farias, por toda ajuda e coautoria.

Finalmente agradeço à Universidade Nove de Julho, pela bolsa de estudos no Programa de Mestrado em Engenharia de Produção.

## RESUMO

A Internet hoje é uma realidade presente na vida das pessoas, possibilitando o surgimento de muitas empresas que fornecem serviços *online*. Porém, para se manterem no mercado, é fundamental que essas empresas se preocupem com a qualidade dos serviços prestados. Neste contexto, torna-se importante conseguir avaliar a satisfação do cliente em relação a esses serviços. Uma das formas utilizadas para avaliar o sentimento dos clientes em relação aos serviços prestados é a Mineração de Opiniões, que se refere ao conjunto de técnicas utilizadas para extrair e avaliar sentimentos expressos em fontes de dados textuais. O objetivo deste trabalho foi comparar Modelos de Classificação dos comentários de clientes para a Mineração de Opiniões em empresa brasileira de classificados *online* de empregos. Foram implementados três Modelos de Classificação: baseado no *software* comercial Repustate; baseado em Dicionário e baseado em Naive Bayes. Os modelos foram aplicados em uma base de dados contendo comentários não estruturados de clientes em língua portuguesa, coletados no formulário de cancelamento do serviço. Um comentário não estruturado apresenta erro de grafia, de concordância e pode ser quase ininteligível. Classificar comentários não estruturados em língua portuguesa é uma tarefa complexa para um classificador por dois motivos: os erros contidos nos comentários e a escassez de modelos de classificação de comentários em língua portuguesa disponíveis para serem tomados como exemplo. Estes motivos tornam os modelos desenvolvidos nesse trabalho relevantes para as pesquisas em Análise de Sentimentos e Mineração de Opiniões. A avaliação do desempenho dos modelos de classificação foi feita usando o índice Kappa e a Matriz de Confusão. O desempenho na classificação dos três Modelos foi analisado por meio da comparação com a classificação realizada por jurados (seres humanos), que também tiveram a avaliação de sua concordância realizada com o índice Kappa e a Matriz de Confusão. A característica não estruturada dos comentários provocou divergência na sua classificação por parte dos jurados e também dos Modelos de Classificação. A concordância entre os classificadores e os jurados chegou no máximo a moderada, com melhor desempenho para o classificador baseado em Naive Bayes. Os modelos foram aplicados na base de dados e, apesar das características dos comentários, a Mineração de Opiniões foi realizada. Conclui-se, então, que o desempenho dos classificadores na mineração de opiniões em empresa brasileira de classificados *online* de empregos foi positivo e o objetivo do trabalho foi atingido. Vale destacar que a Mineração de Opiniões em comentários não estruturados em língua portuguesa é uma tarefa complexa, que exige pesquisa e que este cenário está aberto para novos estudos.

**Palavras-chave:** Mineração de Opiniões. Análise de Sentimentos. Modelos de Classificação. Classificação de Comentários em Língua Portuguesa. Recuperação de Informações. Processamento de Linguagem Natural.

## ABSTRACT

The Internet nowadays is a reality in people's lives, enabling the growth of many online services companies. However, to maintain their activities and stay in the market, it's important for these companies to worry about the quality of the provided services. In this context, it becomes important to be able to assess the client satisfaction regarding those services. One way to assess the clients' sentiment regarding the provided services is Opinion Mining, which refers to the set of techniques used to extract and assess the sentiment expressed in texts. The goal of this work was to compare Classification Models for the clients' comments for Opinion Mining in a Brazilian online job search company. Three models were implemented: one based on commercial software named Repustate; one Dictionary based and one based on Naive Bayes. The Models were applied to a database containing non-structured comments of clients in Portuguese, captured in a service cancellation form. A non-structured comment contains typos, concordance errors and can be almost unintelligible. Classifying non-structured comments in Portuguese is a complex task for a classifier for two reasons: the errors in comments and the scarcity of Classification Models for comments in Portuguese to be taken as examples. Those reasons make the Models developed in this work relevant for the research field of Sentiment Analysis and Opinion Mining. The performance evaluation of the Classification Models was performed using Kappa coefficient and the Confusion Matrix. The classification performance of the three models was analyzed by comparing it with the classification made by human judges, that in turn, had their agreement among them assessed using Kappa and Confusion Matrix. The non-structured characteristic of the comments caused divergence in the judges' classification and also among the Classification Models. The agreement among the classifiers and the judges was moderate at best, with the best performance achieved by the Naïve Bayes based classifier. The models were applied to the database and, despite the characteristics of the comments the Opinion Mining was performed. The conclusion is that the performance of the classifiers in the Opinion Mining in a Brazilian online search company was positive and the goal of this work has been reached. It 's worth to note that the Opinion Mining in non-structured comments in Portuguese is a complex task, that demands research and this scenario is open for new studies.

**Keywords:** Opinion Mining. Sentiment Analysis. Classification Models. Comment Classification in Portuguese. Information Retrieval. Natural Language Processing.

## **LISTA DE ILUSTRAÇÕES**

Figura 1 - Conceituação de um <i>e-service</i> .....	8
Figura 2 - Modelo de qualidade em <i>e-services</i> .....	10
Figura 3 - Arquitetura <i>Opinion Mining</i> .....	12
Figura 4 - Etapas da Mineração de Opiniões .....	16
Figura 5 - Fases principais da Metodologia Experimental .....	23
Figura 6 – Modelo de Classificação dos Comentários baseado no Repustate .....	26
Figura 7 - Modelo de Classificação dos Comentários baseado em Dicionário .....	29
Figura 8 - Modelo de Classificação dos Comentários baseado no Naive Bayes .....	31
Figura 9 - Distribuição dos comentários com classificador baseado no Repustate.....	55
Figura 10 - Distribuição percentual com classificador baseado no Repustate .....	55
Figura 11 - Distribuição dos comentários com classificador baseado em Dicionário .....	56
Figura 12 - Distribuição percentual com classificador baseado em Dicionário .....	56
Figura 13 - Distribuição dos comentários com classificador Naive Bayes .....	57
Figura 14 - Distribuição percentual com classificador Naive Bayes .....	57

## LISTA DE TABELAS

Tabela 1 - Exemplos de opiniões sumarizadas de um produto.....	12
Tabela 2 - Parâmetros relacionados às opiniões dos usuários .....	23
Tabela 3 - Intervalos de <i>score</i> para classificação de sentimentos com o Repustate .....	27
Tabela 4 - Interpretação dos valores de Kappa .....	33
Tabela 5-Exemplo de Matriz de Confusão .....	34
Tabela 6 - Classificação dos Comentários com o Classificador baseado no Repustate .....	36
Tabela 7 - Classificação dos comentários com o Classificador baseado em Dicionário (SentiLex) .....	37
Tabela 8 - Classificação dos Comentários com Naive Bayes .....	37
Tabela 9 - Classificação dos Comentários pelo Jurado 1 .....	38
Tabela 10 - Classificação dos Comentários pelo Jurado 2 .....	38
Tabela 11 - Classificação dos Comentários pelo Jurado 3 .....	39
Tabela 12 - Classificação dos Comentários pelo Jurado 4 .....	39
Tabela 13 - Classificação dos Comentários pelo Jurado 5 .....	39
Tabela 14 - Matriz de Confusão entre o Jurado 1 e o Jurado 2 .....	40
Tabela 15 - Matriz de Confusão entre o Jurado 1 e o Jurado 3 .....	40
Tabela 16 - Matriz de Confusão entre o Jurado 1 e o Jurado 4 .....	41
Tabela 17 - Matriz de Confusão entre o Jurado 1 e o Jurado 5 .....	41
Tabela 18 - Matriz de Confusão entre o Jurado 2 e o Jurado 3 .....	42
Tabela 19 - Matriz de Confusão entre o Jurado 2 e o Jurado 4 .....	42
Tabela 20 - Matriz de Confusão entre o Jurado 2 e o Jurado 5 .....	43
Tabela 21 - Matriz de Confusão entre o Jurado 3 e o Jurado 4 .....	43
Tabela 22 - Matriz de Confusão entre o Jurado 3 e o Jurado 5 .....	44
Tabela 23 - Matriz de Confusão entre o Jurado 4 e o Jurado 5 .....	44
Tabela 24 - Matriz de Confusão entre o Jurado 1 e o Repustate.....	46
Tabela 25 - Matriz de Confusão entre o Jurado 2 e o Repustate .....	46
Tabela 26 - Matriz de Confusão entre o Jurado 3 e o Repustate .....	47
Tabela 27 - Matriz de Confusão entre o Jurado 4 e o Repustate.....	47

Tabela 28 - Matriz de Confusão entre o Jurado 5 e o Repustate .....	48
Tabela 29 - Matriz de Confusão entre o Jurado 1 e o classificador baseado em Dicionário ..	48
Tabela 30 - Matriz de Confusão entre o Jurado 2 e o classificador baseado em Dicionário ..	49
Tabela 31 - Matriz de Confusão entre o Jurado 3 e o classificador baseado em Dicionário ..	49
Tabela 32 - Matriz de Confusão entre o Jurado 4 e o classificador baseado em Dicionário ..	50
Tabela 33 - Matriz de Confusão entre o Jurado 5 e o classificador baseado em Dicionário ..	50
Tabela 34 - Matriz de Confusão entre o Jurado 1 e o classificador baseado em Naive Bayes ..	51
Tabela 35 - Matriz de Confusão entre o Jurado 2 e o classificador baseado em Naive Bayes ..	52
Tabela 36 - Matriz de Confusão entre o Jurado 3 e o classificador baseado em Naive Bayes ..	52
Tabela 37 - Matriz de Confusão entre o Jurado 4 e o classificador baseado em Naive Bayes ..	53
Tabela 38 - Matriz de Confusão entre o Jurado 5 e o classificador baseado em Naive Bayes ..	53
Tabela 39 - Percentuais de coincidência dos classificadores com os jurados .....	54
Tabela 40 - Médias do Valor de Kappa para cada um dos Três Classificadores .....	54
Tabela 41 - Vantagens e desvantagens dos três Modelos de Classificação .....	64

## LISTA DE ABREVIATURAS E SIGLAS

API - *Application Programming Interface*

NLTK - *Natural Language Toolkit*

Web - *World Wide Web*

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 SATISFAÇÃO DO CLIENTE NO SETOR DE SERVIÇOS .....	1
1.2 MINERAÇÃO DE OPINIÕES .....	2
1.3 JUSTIFICATIVA E MOTIVAÇÃO .....	4
1.4 PROBLEMA DE PESQUISA.....	4
1.5 OBJETIVOS .....	4
1.5.1 Objetivo Geral .....	4
1.5.2 Objetivos Específicos .....	5
1.6 ORGANIZAÇÃO DO TRABALHO .....	5
<b>2 FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>6</b>
2.1 QUALIDADE EM SERVIÇOS E SATISFAÇÃO DO CLIENTE .....	6
2.2 MINERAÇÃO DE OPINIÕES E ANÁLISE DE SENTIMENTOS .....	9
2.2.1 Definição das Opiniões .....	12
2.2.2 Etapas da Mineração de Opiniões .....	13
2.2.3 Tipos de Opiniões.....	15
2.2.4 Abordagens para Classificação de Polaridade de Opinião .....	16
2.2.5 Classificação de Sentimento entre Linguagens .....	17
<b>3 MATERIAIS E MÉTODOS.....</b>	<b>19</b>
3.1 CARACTERIZAÇÃO METODOLÓGICA .....	19
3.2 CARACTERIZAÇÃO DA EMPRESA DE CLASSIFICADOS <i>ONLINE</i> DE EMPREGOS .....	20
3.3 METODOLOGIA EXPERIMENTAL .....	21
<b>4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS .....</b>	<b>36</b>
4.1 RESULTADOS DOS MODELOS DE CLASSIFICAÇÃO .....	36
4.2 RESULTADOS DA VALIDAÇÃO PELOS JURADOS.....	38
4.3 AVALIAÇÃO DA CONCORDÂNCIA ENTRE OS JURADOS .....	40
4.4 AVALIAÇÃO DA CONCORDÂNCIA ENTRE OS JURADOS E OS CLASSIFICADORES .....	45
4.4.1 Classificador Repustate com os Jurados .....	45
4.4.2 Classificador Baseado em Dicionário com os Jurados.....	48
4.4.3 Classificador Baseado em Naive Bayes com os Jurados .....	51
4.5 ANÁLISE DA DIVERGÊNCIA ENTRE JURADOS E CLASSIFICADORES .....	55
4.5.1 Classificador Repustate com os Jurados .....	55
4.5.2 Classificador Baseado em Dicionário com os Jurados.....	58
4.5.3 Classificador Baseado em Naive Bayes com os Jurados .....	61
4.6 APRESENTAÇÃO DA MINERAÇÃO DE OPINIÕES .....	64
<b>5 CONCLUSÃO.....</b>	<b>69</b>
5.1 CO-ORIENTAÇÃO EM PROJETO DE INICIAÇÃO CIENTÍFICA .....	72
5.2 PUBLICAÇÕES DO AUTOR .....	72

<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>74</b>
<b>APÊNDICE A: CÓDIGO PYTHON PARA TRADUÇÃO DOS COMENTÁRIOS .....</b>	<b>78</b>
<b>APÊNDICE B – CÓDIGO PYTHON PARA GERAÇÃO DOS SCORES DE COMENTÁRIO COM O REPUSTATE.....</b>	<b>79</b>
<b>APÊNDICE C – INSTRUÇÕES AOS JURADOS PARA CLASSIFICAÇÃO DOS COMENTÁRIOS .....</b>	<b>81</b>
<b>APÊNDICE D – EXEMPLOS DE COMENTÁRIOS DOS CLIENTES DA EMPRESA .....</b>	<b>82</b>

## 1 INTRODUÇÃO

A *Internet* vem despertando mudanças nas relações econômicas, sociais, políticas e culturais. Essas mudanças ainda estão em aberto e se transformam à medida que a própria *Internet* redefine seu escopo e alcance (PINHO, 2011).

No mercado de trabalho, isso não é diferente, e percebe-se que há cada vez mais competitividade neste meio. Muitas vezes as empresas têm dificuldade em preencher postos de trabalho com candidatos qualificados. Nesse ponto, a utilização da *Internet* como ferramenta de apoio na busca por candidatos qualificados traz resultados positivos.

Por outro lado, o recrutamento através da *Internet* pode trazer benefícios adicionais, pois estimativas indicam que o custo de se contratar utilizando anúncios *online* pode chegar a 5% do custo de se contratar através de meios tradicionais (CAPPPELLI, 2001).

Neste panorama, surgiram no Brasil empresas de classificados *online* de empregos. Empresas deste segmento oferecem o serviço de anúncio de vagas de emprego, que ficam disponíveis para busca e consulta via *Internet*. Algumas empresas, além do anúncio de vagas, oferecem também o anúncio de currículos profissionais, servindo tanto como um meio para os candidatos buscarem vagas de emprego quanto para as empresas buscarem profissionais para suas vagas.

Neste ramo de negócio a competição é grande, com várias empresas atuando no mercado de recolocação e recrutamento, tanto no mundo físico quanto no virtual. Por esse motivo, é importante a preocupação com a qualidade dos serviços prestados.

De acordo com Parasuramam (1985), a qualidade em serviços é uma comparação entre as expectativas do cliente e o desempenho do serviço, ou seja, a qualidade em serviços está estreitamente ligada à satisfação do cliente.

Porém, no caso dos serviços *online*, é difícil conhecer as expectativas dos clientes, pois, de acordo com Zeithaml et. al. (2002), os clientes desse tipo de serviço em muitos casos não têm expectativas bem definidas. Por esse motivo, conhecer o sentimento dos clientes após a prestação do serviço pode auxiliar a avaliação de sua satisfação em relação a esses serviços.

### 1.1 SATISFAÇÃO DO CLIENTE NO SETOR DE SERVIÇOS

A satisfação dos clientes é um fator crítico para o sucesso das empresas em um ambiente competitivo, e está relacionada ao atendimento das necessidades explícitas e

implícitas do consumidor por meio dos atributos do serviço (TONTINI e SANT'ANA, 2008). Porém, as operações de serviços em geral são complexas, baseadas nas relações humanas entre o cliente e os empregados que os atendem (BATAGLIA et. al., 2010).

O setor de serviços possui características que o difere de outros setores econômicos, tornando difícil avaliar sua qualidade. As características que diferem este setor dos demais são, a intangibilidade, heterogeneidade, e inseparabilidade (PARASURAMAM, 1985).

A intangibilidade refere-se ao fato dos serviços serem essencialmente performances, e não objetos. A heterogeneidade refere-se à variação do desempenho, e sua percepção ser dependente das expectativas dos clientes. Finalmente, os serviços são inseparáveis, pois não é possível separar sua produção do consumo.

Dentro do setor de serviços, encontra-se o setor de serviços eletrônicos, ou *e-services*. Desenvolvido a partir do *marketing* na *Internet* e da literatura tradicional sobre qualidade em serviços, o conceito de qualidade em *e-services* pode ser definido como a avaliação geral da excelência e qualidade das ofertas de *e-services* na *Internet* (SANTOS, 2003).

Segundo Zeithaml et. al. (2002), a avaliação da qualidade dos *e-services* é ainda mais complexa do que os serviços tradicionais, pois além das três características intrínsecas aos serviços, i.e., inseparabilidade, heterogeneidade e intangibilidade, a percepção de qualidade nos *e-services* baseia-se em cinco critérios:

- Disponibilidade das informações;
- Usabilidade;
- Privacidade;
- Estilo gráfico;
- Realização ou satisfação.

Portanto, ao mesmo tempo em que é importante para as empresas avaliar a satisfação dos clientes em relação aos serviços prestados, esta avaliação é também um grande desafio.

## 1.2 MINERAÇÃO DE OPINIÕES

As opiniões exercem papel central em quase todas as atividades humanas pois exercem influência no nosso comportamento. Em situações onde se faz necessário tomar decisões, é importante conhecer a opinião de outras pessoas. Empresas e organizações querem conhecer as opiniões dos consumidores ou do público em geral sobre seus produtos e serviços (LIU, 2012).

O surgimento da *Web 2.0* e das mídias sociais criou muitas oportunidades para entender a opinião do público geral e dos consumidores sobre eventos sociais, movimentos políticos, estratégia das empresas, campanhas de *marketing* e preferência por produtos. Muitas perguntas referentes à opinião dos consumidores sobre determinado assunto poderiam ser respondidas ao se analisar os milhares de comentários em blogs, mídias e redes sociais como Orkut, Facebook e Youtube ou sites de notícias (CHEN e ZIMBRA, 2010).

A partir de 2009, foi possível testemunhar que a presença de postagens em mídias sociais contendo opinião ajudou a remodelar empresas e a influenciar sentimentos e emoções do público em geral, impactando profundamente nossos sistemas sociais e políticos. Essas postagens chegaram a mobilizar as massas por mudanças políticas, como aquelas que aconteceram em alguns países Árabes em 2011. Tornou-se então uma necessidade coletar e estudar as opiniões na *Internet* (LIU, 2012).

Vale ressaltar que documentos contendo opiniões não existem somente na *Internet*, muitas empresas também mantêm dados internos, como, por exemplo o *feedback* dos clientes coletados por *e-mail* e em *call centers*, ou resultados de pesquisas conduzidas pelas organizações (LIU, 2012).

As informações sobre satisfação do cliente são normalmente suscitadas de forma estruturada, *surveys* ou grupos de estudos apresentam aos clientes perguntas construídas para coletar determinados pedaços de informação que são de interesse da empresa. O conjunto de resultados estruturados pode ser facilmente analisado estatisticamente e convenientemente agregado de acordo com as questões da *survey* (GAMON et. al., 2005).

Porém, essa abordagem, além de ser dispendiosa, consegue somente captar os aspectos que foram previstos durante a elaboração do questionário, deixando escapar as nuances que porventura existam quando os clientes podem expressar suas opiniões livremente (GAMON et. al., 2005).

Mineração de Opiniões, uma subdisciplina da Mineração de Dados e da Linguística Computacional, refere-se às técnicas computacionais para extrair, classificar, entender e avaliar as opiniões expressas em várias fontes de notícias *online*, comentários em mídias sociais e outros conteúdos criados por usuários (CHEN e ZIMBRA, 2010).

É o campo de estudo que analisa as opiniões, sentimentos, avaliações, atitudes e emoções das pessoas em relação a entidades como produtos, serviços, organizações,

indivíduos, problemas, eventos, tópicos, e seus atributos, representando um amplo campo de estudo (LIU, 2012).

### 1.3 JUSTIFICATIVA E MOTIVAÇÃO

De acordo com Parry e Tyson (2008) a tendência em longo prazo do mercado de recrutamento *online* é de crescimento. Neste cenário, este mercado tende a se tornar mais competitivo, com a entrada de novas empresas neste segmento. Desta forma, conseguir avaliar a satisfação de seus clientes se torna um fator importante para seu sucesso.

A empresa em estudo nesta dissertação possui milhares de clientes, e uma base de dados com milhões de comentários em língua portuguesa a serem avaliados, o que torna impraticável ler todos os comentários para interpretar se estes são positivos, negativos ou neutros. Ademais, os comentários contidos nesta base de dados são não estruturados, ou seja, são comentários que apresentam erros de grafia, erros de concordância e comentários quase ininteligíveis.

Além da dificuldade em avaliar esses comentários, a empresa trata reclamações dos usuários de forma pontual. Esta abordagem não oferece um panorama fiel da qualidade dos serviços prestados pela empresa, pois não consegue englobar todos os seus clientes e tampouco trata de uma amostra aleatória, o que justifica a aplicação da Mineração de Opiniões para avaliar a opinião dos clientes sobre os serviços prestados pela empresa.

Apesar de já existirem várias ferramentas consolidadas para Mineração de Opiniões em língua inglesa, existem poucas ferramentas em língua portuguesa, o que também motiva a realização desse trabalho.

### 1.4 PROBLEMA DE PESQUISA

Os Modelos de Classificação dos comentários de clientes comparados neste trabalho podem minerar opiniões em empresa brasileira de classificados online de empregos?

### 1.5 OBJETIVOS

#### 1.5.1 Objetivo Geral

O objetivo deste trabalho foi comparar Modelos de Classificação dos comentários de clientes para a Mineração de Opiniões em empresa brasileira de classificados *online* de empregos.

### 1.5.2 Objetivos Específicos

Os objetivo específicos deste trabalho são:

- Implementar o Modelo de Classificação dos Comentários com o Classificador baseado no **software comercial Repustate**;
- Implementar o Modelo de Classificação dos Comentários com o Classificador baseado em **Dicionário**;
- Implementar o Modelo de Classificação dos comentários com o Classificador baseado em **Naive Bayes**;
- Comparar o resultado dos Modelos de Classificação com a classificação realizada pelos jurados humanos.
- Avaliar o desempenho dos Modelos de Classificação usando o índice Kappa e a Matriz de Confusão.
- Analisar a divergência entre os Modelos de Classificação e os jurados.
- Minerar opiniões na base de dados de comentários.

## 1.6 ORGANIZAÇÃO DO TRABALHO

Além deste capítulo introdutório (Capítulo 1), este trabalho está estruturado da seguinte forma:

Capítulo 2 - Fundamentação Teórica. Neste capítulo são apresentados os conceitos abordados no desenvolvimento deste trabalho: Qualidade em serviços, com ênfase em *e-services* e Mineração de Opiniões.

Capítulo 3 - Materiais e Métodos. Neste capítulo é apresentada a metodologia utilizada e os modelos propostos.

Capítulo 4 - Resultados Experimentais. São apresentados neste capítulo os resultados dos experimentos realizados e a análise desses resultados.

Capítulo 5 - Conclusão. Neste capítulo é apresentada a conclusão deste trabalho.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os referenciais teóricos sobre os temas abordados neste trabalho. Inicialmente é apresentado o referencial teórico sobre qualidade em serviços e satisfação do cliente. A seguir é apresentado o referencial teórico sobre Mineração de Opiniões e Análise de Sentimentos. Finalmente o referencial teórico sobre redes neurais artificiais, com ênfase nas redes neurais artificiais do tipo SOM é apresentado.

### 2.1 QUALIDADE EM SERVIÇOS E SATISFAÇÃO DO CLIENTE

Os serviços possuem certas características que o diferem de outros setores econômicos no que se refere à percepção de qualidade. Essas características são a intangibilidade, heterogeneidade e a inseparabilidade (PARASURAMAM, 1985).

Os serviços são intangíveis, pois são desempenhos, e não objetos. Muitos serviços não podem ser medidos, contados, inventariados, testados ou verificados antes do ato da sua prestação, de forma a garantir sua qualidade (BERRY, 1980).

Os serviços são heterogêneos, pois seu desempenho é variável. Dependem do fornecedor e do cliente. E a experiência que a empresa pretende proporcionar pode ser diferente das expectativas do cliente (BOOMS e BITNER, 1981).

Os serviços são inseparáveis, pois a sua produção e consumo não podem ser separados. Por esse motivo, não se pode garantir a qualidade durante a linha de produção da fábrica e em seguida entregá-la intacta ao cliente (UPAH, 1980).

Existem também os serviços prestados através da *Internet*, também conhecidos como *e-services*. Esses serviços possuem características especiais que os diferenciam dos serviços tradicionais. Por exemplo, as expectativa do cliente em relação ao *e-service* tem menor importância do que no caso dos serviços tradicionais (ZEITHAML et. al.2002).

A avaliação da qualidade dos *e-services*, possui algumas peculiaridades. Conforme Parasuraman et. al. (2005), a percepção da qualidade desse tipo de serviço depende da familiaridade do cliente com tecnologia.

De acordo com Zeithaml et. al. (2002), a qualidade percebida em um *website* se baseia em cinco critérios:

- Disponibilidade de informação e conteúdo;
- Usabilidade, ou facilidade de uso;

- Privacidade e segurança;
- Estilo gráfico;
- Realização ou satisfação.

As expectativas dos clientes ao utilizar serviços *online* são diferentes das expectativas dos clientes de serviços tradicionais (*off-line*). Em grande parte dos casos, os clientes não têm expectativas muito bem definidas, e em muitos casos, seus padrões de consumo anteriores são inexistentes ou imprecisos (ZEITHAML et. al., 2002).

De acordo com o modelo proposto por van Riel et. al.(2001) para a oferta de *e-services*, mostrado na Figura 1, os *e-services* podem ser divididos em cinco componentes.

- o serviço principal;
- serviços facilitadores;
- serviços de suporte;
- serviços complementares;
- interface com o usuário, através da qual os clientes tem acesso ao serviço.

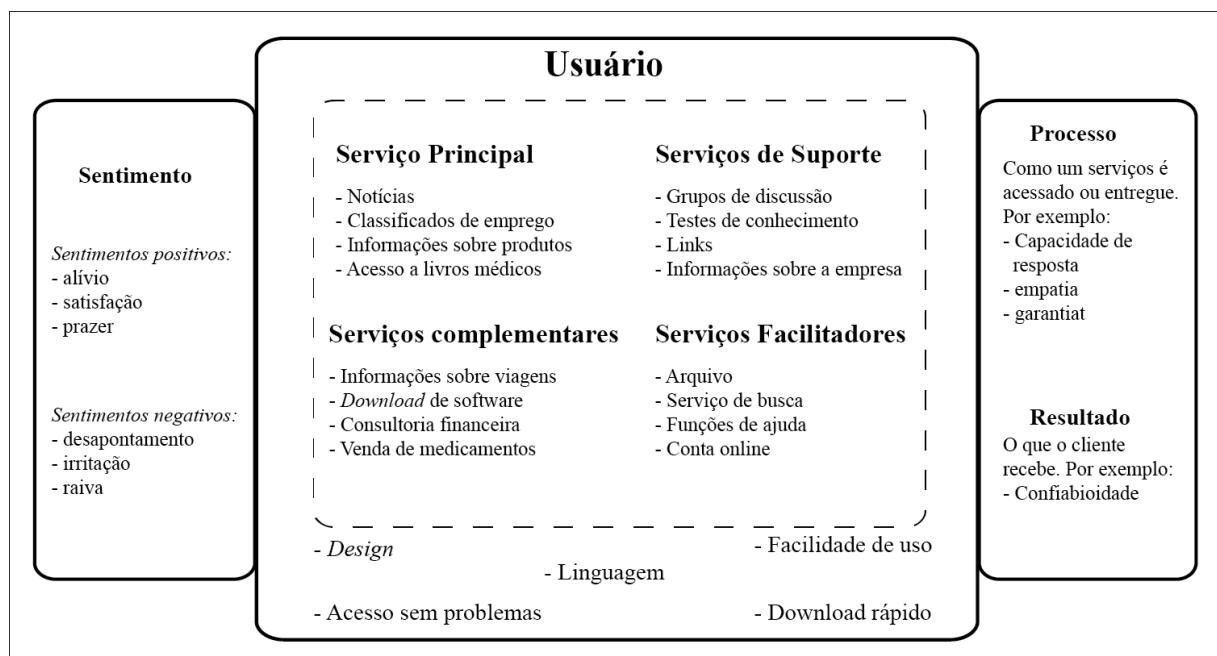


Figura 1 - Conceituação de um *e-service*. Fonte: Adaptado de van Riel et al. (2001)

Uma grande variedade de serviços facilitadores e de suporte, como máquinas de busca ou fóruns de comunicação, podem adicionar valor ao serviço principal. Lucro pode ser gerado a partir do serviço principal, ou pelos serviços complementares que alavancam o segmento de mercado. Os serviços complementares podem ser providos pela empresa que fornece o serviço principal, ou então por empresas terceirizadas (VAN RIEL et al., 2001).

O conceito de qualidade em *e-services* tem sua origem no *marketing* digital e na literatura tradicional sobre qualidade em serviços. Pode ser definida como a avaliação geral dos consumidores sobre a excelência e qualidade das ofertas de *e-services* no mercado digital (SANTOS, 2003).

Em relação à percepção da qualidade de cada um dos componentes dos *e-services*, junto com o julgamento de valor e da qualidade, os clientes podem também experimentar emoções distintas em relação a cada um dos componentes. Por exemplo, enquanto o serviço principal pode ser percebido como confiável, os serviços facilitadores podem ser percebidos como não confiáveis, o que pode causar frustração e raiva (VAN RIEL *et al.*, 2001).

Em contraste com a avaliação da qualidade nos serviços tradicionais, os consumidores são menos propensos a avaliar cada sub-processo envolvido em uma visita a um *website*. Ao invés disso, os clientes tendem a perceber o *e-service* como um processo único, que gera um único resultado (VAN RIEL *et al.*, 2001).

O modelo proposto por Santos (2003) para a qualidade em *e-services* possui duas dimensões principais. A primeira é a dimensão incubada. Esta dimensão refere-se às características da qualidade que podem ser tratadas antes do *website* ser lançado. Esta dimensão inclui:

- facilidade de uso (usabilidade);
- aparência;
- acoplamento;
- estrutura e leiaute;
- conteúdo.

A outra dimensão proposta no modelo é a dimensão ativa. É definida como o suporte, velocidade, e manutenção que o *website* que provê o e-service pode oferecer aos seus clientes. Consiste nos seguintes características:

- confiança;
- eficiência;
- suporte;
- comunicação;
- segurança;
- incentivo.

O modelo proposto por Santos (2003) é mostrado na Figura 2. Nesta figura, além de exemplos de algumas das características do modelo, é também indicada a ordem de importância das características das duas dimensões.

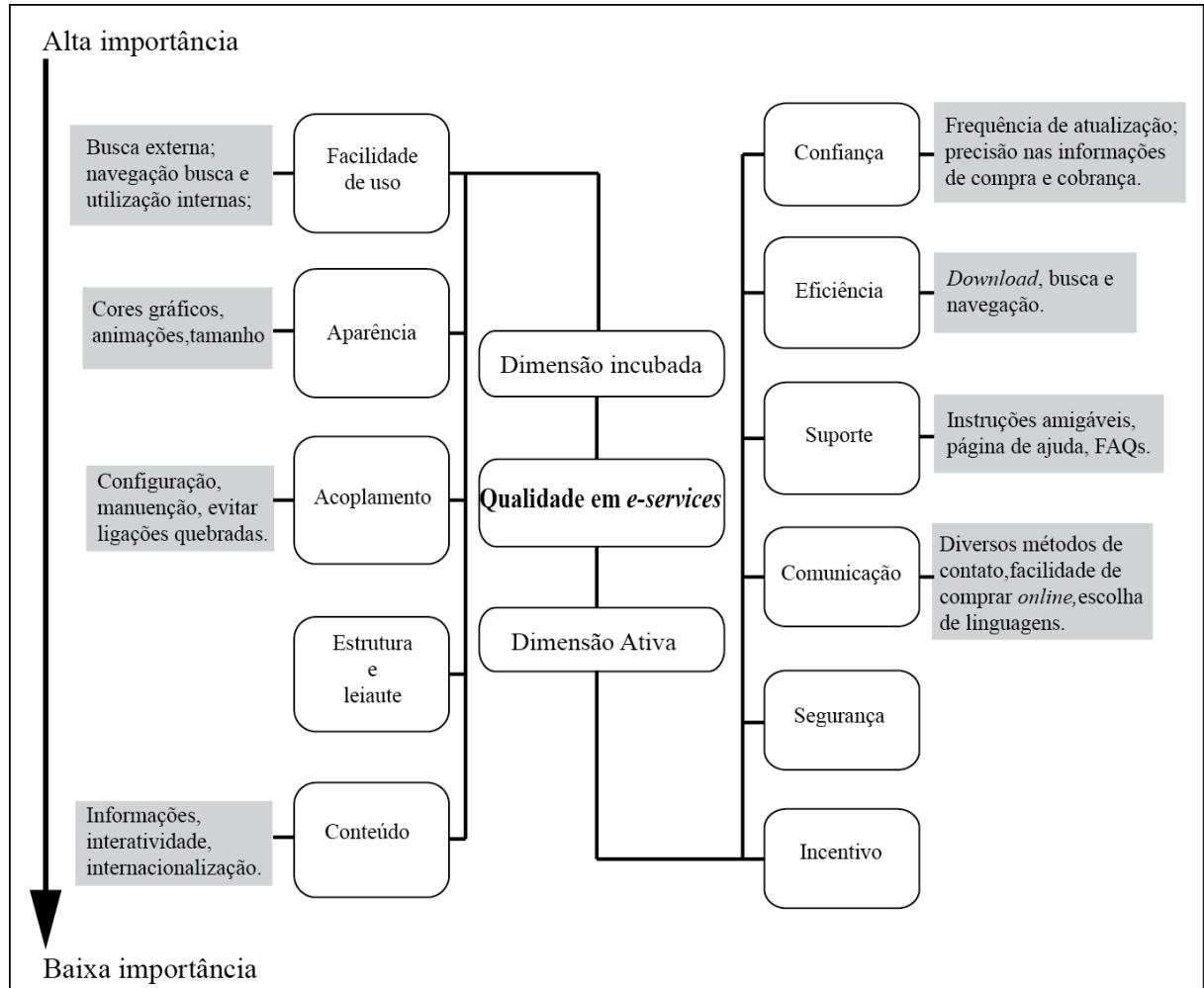


Figura 2 - Modelo de qualidade em *e-services*. Fonte: Adaptado de Santos (2003).

## 2.2 MINERAÇÃO DE OPINIÕES E ANÁLISE DE SENTIMENTOS

O surgimento da *Web 2.0* e das mídias sociais criou muitas oportunidades para entender a opinião do público geral e dos consumidores sobre eventos sociais, movimentos políticos, estratégia das empresas, campanhas de *marketing* e preferência por produtos (LIU, 2012). Muitas perguntas referentes à opinião dos consumidores sobre determinado assunto poderiam ser respondidas ao se analisar os milhares de comentários em blogs, mídias e redes sociais como Orkut, Facebook e Youtube ou sites de notícias.

Neste cenário de abundância de dados contendo opiniões, surgiu a necessidade de técnicas e ferramentas capazes de identificar essas opiniões, sem a necessidade de intervenção humana. Surgiu assim a Mineração de Opiniões.

A Mineração de Opiniões, uma subdisciplina dentro da Mineração de Dados e da Linguística Computacional, refere-se às técnicas computacionais para extrair, classificar, entender e avaliar as opiniões expressas em várias fontes de notícias *online*, comentários em mídias sociais e outros conteúdos criados por usuários (CHEN e ZIMBRA, 2010).

O termo Mineração de Opiniões é mais comum no meio acadêmico, enquanto Análise de Sentimentos é mais comum nas organizações. Porém, os dois termos se referem ao mesmo assunto (LIU, 2012).

Na linguagem da Mineração de Opiniões, o conteúdo criado por cada usuário é chamado de documento, por exemplo, uma postagem em um fórum, um comentário ou postagem em um blog, ou ainda uma revisão de um produto, se chama documento.

A Mineração de Opiniões não se preocupa em identificar o assunto de um documento, mas em identificar e classificar as opiniões expressas nele (CHEN e ZIMBRA, 2010). Os dados no texto de um documento podem ser divididos em duas grandes categorias, sendo elas fatos e opiniões. Fatos são declarações objetivas, enquanto opiniões são declarações subjetivas (VALARMATHI e PALANISAMY, 2011).

Para identificar as opiniões expressas em um documento, pode-se utilizar Mineração de Opiniões no nível de características, por exemplo, opiniões expressas sobre características de um produto, e dessa forma descobrir o sentimento associado a diferentes características (WANG; ZHOU, 2009).

Liu e Hu (2004) descrevem uma técnica para sumarizar as opiniões expressas em um conjunto de *reviews* (análise de produtos ou serviços feitos por consumidores) de um produto. Este processo consiste em dois passos principais, a extração das características e a identificação da opinião associada àquelas características, podendo ser a opinião positiva ou negativa. A Figura 3 mostra a arquitetura desenvolvida por Liu e Hu (2004).

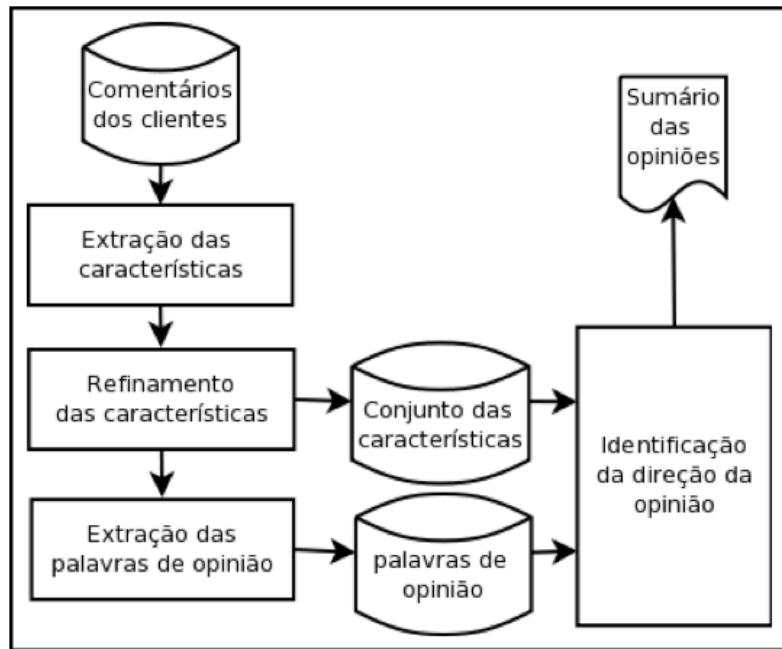


Figura 3 - Arquitetura *Opinion Mining*. Fonte: Adaptado de Hu e Liu (2004)

Na Figura 3, as entradas do sistema são os documentos contendoreviews de um produto. Inicialmente é realizada a identificação das características do produto que são tratadas no documento. Em seguida é feito o refinamento dessas características, por exemplo, eliminando duplicidades. Em seguida são identificadas as palavras que expressão opinião. Então, a partir desses dois conjuntos, é realizada a classificação dos *reviews* baseando-se nas características do produto. A saída do sistema é um sumário com as características e as opiniões, como no exemplo da Tabela 1 em que o produto é uma câmera digital.

Tabela 1 -Exemplos de opiniões summarizadas de um produto.

Qualidade da foto	Positivo:	253
	Negativo:	6
Tamanho	Positivo:	134
	Negativo:	10

Fonte: Adaptado de Hu e Liu (2004)

De acordo com Liu (2012), a Mineração de Opiniões pode ser realizada em três níveis:

1. **Nível de documentos:** O primeiro nível é o nível de documento, onde se preocupa em extrair a opinião geral expressa no documento. Neste tipo de análise, é importante que

o documento trate somente de uma entidade (o objeto, pessoa ou produto a que se refere à opinião).

2. **Nível de sentença:** O segundo nível é o nível de sentenças, onde cada sentença de um documento é classificada separadamente. Bastante utilizado quando um mesmo documento contém opinião sobre várias entidades.
3. **Nível de aspectos:** O terceiro nível é o nível de aspecto. Neste nível, existe a preocupação de identificar o que exatamente o autor gostou ou não gostou na entidade, ou seja, quais os aspectos, ou características, da entidade para as quais o autor manifestou opinião.

#### 2.2.1 Definição das Opiniões

De maneira geral as opiniões estão expressas em textos de forma não estruturada, e isso dificulta seu estudo. Para resolver este problema, faz-se necessária uma definição formal, que apresente uma opinião de forma estruturada, de modo que ela possa ser processada computacionalmente. A fim de resolver este problema, Liu (2012) definiu opinião como uma quíntupla (1):

$$O = (e_i, a_{ij}, s_{ijkl}, h_k, t_l) \quad (1)$$

Na qual:

- $e_i$  é o nome de uma entidade;
- $a_{ij}$  é um aspecto da entidade  $e_i$ . Caso a opinião seja sobre a entidade como um todo, o aspecto especial GERAL é utilizado;
- $s_{ijkl}$  é o sentimento associado ao aspecto  $a_{ij}$  da entidade  $e_i$  e pode ser positivo, negativo ou neutro, ou então ser expresso por diferentes níveis de intensidade;
- $h_k$  é o emissor da opinião;
- $t_l$  é o momento em que a opinião foi expressa pelo emissor  $h_k$ .

Nesta definição, é importante notar, e isso é reforçado pelos subscritos, que deve haver correspondência entre os itens da quíntupla. Vale também ressaltar que todos os componentes são essenciais. A falta, por exemplo, de  $t_l$  impede que se analise a opinião de acordo com o tempo. Isto pode ser problemático, já que a opinião sobre a característica de um produto emitida no passado pode não ser mais relevante para os dias atuais.

Como exemplo, pode-se considerar um texto fictício escrito por um usuário que comprou uma câmera digital, e as opiniões extraídas deste texto:

- João, 11 de maio de 2012: “A câmera é muito boa. É leve e muito fácil de usar. Consegue capturar fotos excelentes. O zoom funciona muito bem. Porém o foco automático é um pouco lento.
- (Câmera, geral, positivo, João, 11/05/2012)
- (Câmera, peso, positivo, João, 11/05/2012)
- (Câmera, facilidade de uso, positivo, João, 11/05/2012)
- (Câmera, qualidade das fotos, positivo, João, 11/05/2012)
- (Câmera, zoom, positivo, João, 11/05/2012)
- (Câmera, foco automático, João, negativo, 11/05/2012)

Como se pode observar, foi possível extrair seis quíntuplas do texto. Os aspectos foram o peso, a facilidade de uso, a qualidade das fotos, o zoom e o foco automático. O usuário também deixa explícita sua opinião sobre a entidade, de forma geral.

### 2.2.2 Etapas da Mineração de Opiniões

Dada a definição de opinião como uma quíntupla, é possível agora definir o objetivo e as etapas principais da análise de sentimentos (LIU, 2011). O objetivo da ou Mineração de Opiniões é, dado um documento  $d$  contendo opiniões, descobrir todas as quíntuplas ( $e_i, a_{ij}, s_{ijk}, h_k, t_l$ ) contidas em  $d$ .

As tarefas necessárias para descobrir as quíntuplas são relacionadas a cada um dos componentes da quíntupla. Primeiramente é necessário descobrir as entidades  $e_i$  presentes no documento. Para tanto, é necessário extrair as expressões que correspondem a entidades no documento. Após esse passo, como muitas vezes mais de uma expressão pode corresponder à mesma entidade, é necessário agrupar as entidades em categorias. Cada entidade deve ser reconhecida por um único nome dentro de uma aplicação. O mesmo ocorre com os aspectos, onde mais de uma expressão pode se referir ao mesmo aspecto (LIU, 2012).

O próximo componente na definição de opinião é o sentimento. A este componente relaciona-se a tarefa de classificar o sentimento sobre determinado aspecto como positivo, negativo ou neutro. Os outros componentes são o emissor da opinião e o tempo no qual a opinião foi emitida (LIU, 2012).

Desta forma, Liu (2012) define os seguintes modelos:

- **Entidade:** Uma entidade  $e_i$  é representada por ela mesma como um todo e um conjunto finito de aspectos  $A_i = \{a_{i1}, a_{i2}, a_{i3}, \dots, a_{in}\}$ . Sendo que  $e_i$  pode ser expresso com qualquer componente do conjunto finito de expressões que denotam a entidade  $\{ee_{i1}, ee_{i2}, ee_{i3}, \dots, ee_{is}\}$ . Cada aspecto  $a_{ij} \in A_i$  da entidade  $e_i$  pode ser expressa como qualquer elemento do conjunto de expressões que denotam aspectos  $\{ae_{i1}, ae_{i2}, ae_{i3}, \dots, ae_{ijm}\}$ .
- **Documento de opinião:** Um documento de opinião  $d$  contem opiniões sobre um conjunto de entidades  $\{e_1, e_2, \dots, e_r\}$  e um subconjunto originadas por um conjunto de emissores  $\{h_1, h_2, \dots, h_p\}$ , em algum momento particular.

Definidos esses dois modelos, é possível definir as tarefas necessárias para atingir o objetivo da Análise de Sentimentos conforme a definição de Liu (2012):

- **Tarefa 1:** Extração e categorização das entidades. É necessário extrair e agrupar as expressões que denotam entidades, de forma que cada grupo indique uma única entidade  $e_i$ .
- **Tarefa 2:** Extração e categorização dos aspectos: É necessário extrair todas as expressões de aspectos das entidades e agrupar essas expressões de forma que cada grupo represente um aspecto  $a_{ij}$  da entidade  $e_i$ .
- **Tarefa 3:** Extração e categorização do emissor da opinião. Da mesma maneira que as outras duas tarefas, é necessário extrair os emissores presentes no texto e categoriza-los.
- **Tarefa 4:** Extração e padronização do tempo. É necessário identificar os tempos em que as opiniões foram emitidas, e padronizar os formatos.
- **Tarefa 5:** Classificação do sentimento em relação aos aspectos. É necessário determinar se a opinião sobre um aspecto  $a_j$  é positiva, negativa ou neutra. Ou então atribuir uma classificação numérica ao aspecto.
- **Tarefa 6:** Geração das quíntuplas de opinião. Montar todas as quíntuplas de opinião ( $e_i, a_{ij}, s_{ijkl}, h_k, t_l$ ) expressas no documento  $d$ , baseando-se nos resultados acima.

Com as quíntuplas de opinião, é possível criar um sumário de opiniões, sendo possível assim analisar as opiniões de determinado grupo de pessoas tanto quantitativamente quanto qualitativamente. Um exemplo de sumário de opiniões é apresentado em Liu e Hu (2004), na Tabela 1. A Figura 4 resume todo o processo de Mineração de Opiniões:



Figura 4 - Etapas da Mineração de Opiniões. Fonte: Adaptado de Liu (2012).

### 2.2.3 Tipos de Opiniões

As opiniões podem ser classificadas entre opiniões regulares e opiniões comparativas. O tipo de opinião normalmente tratada na literatura é a opinião regular, quanto as opiniões do tipo comparativa são menos comuns (JINDAL e LIU, 2006). As opiniões podem também ser divididas entre explícitas e implícitas (LIU, 2012).

A opinião regular é o tipo mais simples e é geralmente tratado na literatura simplesmente como opinião. Pode ser dividida em dois subtipos (LIU, 2011):

- **Opiniões diretas:** Referem-se a opiniões expressas diretamente sobre uma entidade ou o aspecto de uma entidade. Como, por exemplo, em “A qualidade da foto é excelente.”
- **Opiniões indiretas:** São expressas indiretamente sobre uma entidade ou aspectos de uma entidade, baseando-se nos efeitos causados em alguma outra entidade, por exemplo, a sentença “Depois da injeção do medicamento, minha cabeça começou a doer” descreve

um efeito colateral indesejável do medicamento sobre a cabeça. Neste caso, a entidade é o medicamento, e o aspecto é o “efeito na cabeça”.

Outro tipo de opinião refere-se às opiniões comparativas. Uma opinião comparativa expressa uma relação de similaridades ou diferenças entre duas ou mais entidades, ou a preferência de emissor da opinião, baseando-se em atributos comuns às entidades (JINDAL e LIU, 2006). Por exemplo, as sentenças “O presunto da empresa X é mais saboroso que o presunto da empresa Y” ou “A mortadela da empresa A é a mais saborosa” expressam duas opiniões comparativas.

As opiniões podem também ser divididas entre explícitas e implícitas. Uma opinião explícita é uma afirmação subjetiva que fornece ou uma opinião regular, ou uma comparativa. Ja uma opinião implícita é uma afirmação objetiva que implica em uma opinião regular ou comparativa (LIU, 2012).

Um exemplo de opinião explícita seria: “O sabor do presunto da empresa X é fantástico!”. Enquanto um exemplo de uma opinião implícita seria: “Formou-se um vale no colchão que eu comprei na semana passada”.

#### 2.2.4 Abordagens para Classificação de Polaridade de Opinião

Polaridade é o termo usado para designar uma opinião como positiva, negativa ou neutra. A polaridade de uma opinião pode ser obtida através de várias abordagens. De acordo com Tsytserau e Palpanas (2012), existem quatro principais métodos, ou abordagens, para a classificação da polaridade de opinião. São eles:

- os métodos baseados na aprendizagem de máquina;
- os métodos baseados em dicionário;
- os métodos estatísticos;
- e os métodos semânticos.

A abordagem baseada em aprendizagem de máquina pode ser dividida em duas categorias, sendo elas aprendizado supervisionado e aprendizado não-supervisionado. No aprendizado supervisionado, existe um conjunto de treinamento em que se conhece as entradas do modelo e o resultado que o modelo deve fornecer. Já no aprendizado não-supervisionado, conhece-se somente as entradas do modelo.

O primeiro trabalho a utilizar aprendizagem de máquina para a classificação da polaridade de sentimento foi o trabalho de Pang et. al. (2002). Neste trabalho são comparados os métodos Naive Bayes, Máxima Entropia, e *Support Vector Machines*(SVM), na classificação de *reviews* de filmes.

A abordagem baseada em dicionário se baseia na utilização de um dicionário construído previamente, contendo palavras de opinião e sua polaridade associada. O dicionário mais utilizado atualmente é o SentiWordNet (ESULI e SEBASTIANI, 2006; BACCIANELLA, ESULI e SEBASTIANI, 2010).

Para a língua portuguesa, existe o trabalho desenvolvido por Chaves *et. al.* (2012), para a avaliação de *reviews online* do setor de acomodações. Este trabalho apresenta o algoritmo PIRPO, que realiza Análise de Sentimentos no nível de aspectos em língua portuguesa.

Na abordagem estatística, identifica-se a polaridade de uma palavra ao se estudar a frequência com que essa palavra ocorre em um *corpus* anotado (MIAO et. al., 2009). Se a palavra ocorre com mais frequência em textos positivos, ela é considerada positiva. Se ocorre com mais frequência em textos negativos, a palavra é negativa. Se ocorre com frequência igual tanto em textos positivos quanto negativos, é considerada neutra (TSYTSARAU e PALPANAS, 2012).

A abordagem semântica, assim como a abordagem estatística fornece diretamente a polaridade de sentimento, porém utiliza-se de princípios diferentes para calcular a similaridade entre palavras. O princípio por trás dessa abordagem é que palavras semanticamente próximas devem receber valores de sentimento similares (TSYTSARAU e PALPANAS, 2012).

## 2.2.5 Classificação de Sentimento entre Linguagens

A classificação de sentimentos entre linguagens é a utilização de técnicas, conjuntos de dados e heurísticas para classificação do sentimento de documentos escritos em uma determinada linguagem, para classificar documentos escritos em uma linguagem diferente. Geralmente as ferramentas são desenvolvidas para a língua inglesa.

Existem duas principais motivações para a classificação de sentimentos entre linguagens. Na primeira motivação, pesquisadores de diferentes países querem desenvolver sistemas para análise de sentimentos pra suas próprias línguas. Na segunda motivação,

empresas querem conhecer e comparar as opiniões de seus clientes sobre seus produtos ou serviços em diferentes países (LIU, 2012).

Muita pesquisa já foi realizada em relação à classificação de sentimentos para textos em língua inglesa, mas por outro lado, a pesquisa para classificação de documentos em outras línguas é escassa. A maioria dos métodos e ferramentas desenvolvidos contemplam a língua inglesa (LIU, 2012).

Wan (2008) utilizou recursos para Análise de Sentimentos em inglês para realizar a classificação de *reviews* em chinês. Seu método consistiu em realizar a tradução dos documentos escritos em chinês para o inglês, utilizando vários sistemas de tradução automática, gerando assim várias versões em inglês de cada documento originalmente em chinês. A classificação de cada documento é realizada, e em seguida os resultados de cada versão são combinados para retornar o resultado final.

Brooke (2009) também utilizou a tradução para a língua inglesa, porém, a linguagem alvo era o espanhol, e utilizou-se de somente um sistema de tradução automática. Após a tradução dos documentos do espanhol para o inglês, o autor usou um método baseado em dicionário e baseado em aprendizagem de máquina para classificação dos documentos em espanhol.

Wan (2009) propôs um método de co-treinamento que utilizou um *corpus* (conjunto de textos) anotado em inglês para a classificação de *reviews* em chinês, de forma supervisionada, sem a utilização de nenhum recurso em chinês. Os *reviews* em inglês rotulados são traduzidos para o chinês e os *reviews* em chinês não rotulados são traduzidos para o inglês. Em seguida, é utilizado um algoritmo de co-treinamento utilizando SVM para treinar dois classificadores. Os dois classificadores são então combinados em um só, e utilizado para classificar os *reviews* traduzidos para a língua inglesa.

No próximo capítulo os Materiais e Métodos utilizados neste trabalho são apresentados.

### 3 MATERIAIS E MÉTODOS

Neste capítulo são apresentados os Materiais e Métodos utilizados no trabalho e a descrição detalhada da Metodologia de realização dos experimentos.

#### 3.1 CARACTERIZAÇÃO METODOLÓGICA

A Metodologia de Pesquisa adotada neste trabalho foi definida como bibliográfica, exploratória e experimental. A pesquisa bibliográfica abrange a leitura, análise e interpretação de livros, periódicos, documentos, mapas, imagens, manuscritos, etc. Todo material recolhido deve ser submetido a uma triagem, a partir da qual é possível estabelecer um plano de leitura. Trata-se de uma leitura atenta e sistemática que se faz acompanhar de anotações e fichamentos que, eventualmente, poderão servir à fundamentação teórica do estudo (GIL, 2002).

A realização da pesquisa bibliográfica foi embasada em consultas a fontes bibliográficas e de referencial teórico, tais como: artigos, livros, teses, dissertações e *sites* com conteúdos sobre o setor de serviços, o setor de serviços eletrônicos (*e-services*), qualidade no setor de serviços, qualidade nos *e-services*, Mineração de Opiniões e Análise de Sentimentos.

Foram consultadas as seguintes bases de dados: SCIELO, IEEE Xplore, SCOPUS e de congressos da área da Engenharia de Produção, como o Encontro Nacional de Engenharia de Produção (ENEGET), o Simpósio de Engenharia de Produção (SIMPEP) e o Encontro Mineiro de Engenharia de Produção (EMEPRO).

Segundo Yin (2006) a pesquisa exploratória permite uma maior familiaridade entre o pesquisador e o tema pesquisado, visto que este ainda é pouco conhecido, pouco explorado. Nesse sentido, caso o problema proposto não apresente aspectos que permitam a visualização dos procedimentos a serem adotados, será necessário que o pesquisador inicie um processo de sondagem, com vistas a aprimorar ideias, descobrir intuições e, posteriormente, construir hipóteses.

Uma pesquisa exploratória visa proporcionar maior familiaridade com o problema com vistas a torná-lo explícito ou a construir hipóteses. Pode-se dizer que esta pesquisa objetiva o aprimoramento de ideias ou a descoberta de intuições. Seu planejamento é, portanto, bastante flexível para que possibilite a consideração dos mais variados aspectos relativos ao fato estudado.

A pesquisa experimental determina um objeto de estudo, selecionam-se as variáveis que seriam capazes de influenciá-lo, definem-se as formas de controle e de observação dos efeitos que a variável produz no objeto (GIL, 2002).

### 3.2 CARACTERIZAÇÃO DA EMPRESA DE CLASSIFICADOS *ONLINE* DE EMPREGOS

A empresa de classificados *online* de empregos alvo do estudo atua no mercado brasileiro desde 1996. Ela oferece tanto o anúncio de vagas para empresas quanto o cadastro de currículos profissionais para os clientes que procuram emprego. O modelo de negócios adotado por esta empresa é o de oferecer o anúncio de vagas gratuito para as empresas, e cobrar o serviço de busca de vagas e envio de currículos, dos profissionais que procuram emprego. Vale ressaltar que esse modelo difere de outras empresas do mesmo segmento no mundo.

Através dos sistemas da empresa, além de buscar por vagas de emprego, os candidatos às vagas podem enviar seus currículos para as vagas anunciadas. As empresas anunciantes, por sua vez, podem visualizar os currículos enviados através da plataforma oferecida pela empresa de classificados *online*, podendo gerenciar o *portfólio* de currículos da vaga, criar e gerenciar processos seletivos, marcar entrevistas, entre outras ações relacionadas ao recrutamento, ou seja, as empresas anunciantes têm acesso a um sistema de recrutamento *online*.

Nessa empresa, como geralmente o objetivo do cliente é conseguir uma posição no mercado de trabalho, após atingir esta meta é comum que o cliente cancele o serviço. Isso não quer dizer necessariamente que o cliente não ficou satisfeito com o serviço, mas pode significar justamente inverso. Por esse motivo, não é tarefa trivial avaliar a satisfação dos clientes desta empresa.

A empresa possui uma grande base de dados contendo comentários dos clientes sobre os serviços prestados. Estes comentários foram coletados através de um formulário apresentado ao cliente no ato do cancelamento do serviço. Devido aos motivos já expostos, estes comentários podem ser positivos, indicando que o cliente está satisfeito, ou negativos, indicando que o cliente não está satisfeito. Existe ainda a possibilidade do cliente não emitir opinião em relação ao serviço, ou emitir opinião neutra. Estes comentários possuem a característica de serem não estruturados, ou seja, apresentam erros de grafia, de concordância

ou podem ser quase ininteligíveis. Exemplos desses comentários podem ser encontrados no Apêndice D.

Desta forma, a retenção dos clientes não pode ser utilizada como parâmetro para medir a satisfação do cliente. Isso acontece, pois clientes que cancelam o serviço podem estar fazendo isso justamente por ter conseguido um emprego através dos serviços da empresa. Por outro lado, o cliente pode permanecer durante muito tempo utilizando os serviços da empresa, pois alguns gostam de manterem o serviço para verificar e sondar o mercado de trabalho. Em ambos os casos, sendo o tempo de permanência curto ou longo, o cliente pode tanto estar satisfeito quanto insatisfeito.

### 3.3 METODOLOGIA EXPERIMENTAL

A Metodologia Experimental foi dividida em 4 Fases principais:

- Primeira Fase: Seleção dos Comentários;
- Segunda Fase: Modelos de Classificação dos Comentários;
- Terceira Fase: Avaliação do Desempenho dos Modelos de Classificação dos Comentários;
- Quarta Fase: Apresentação dos Resultados da Mineração de Opiniões.

A Figura 5 detalha os passos constituintes de cada uma das Fases.

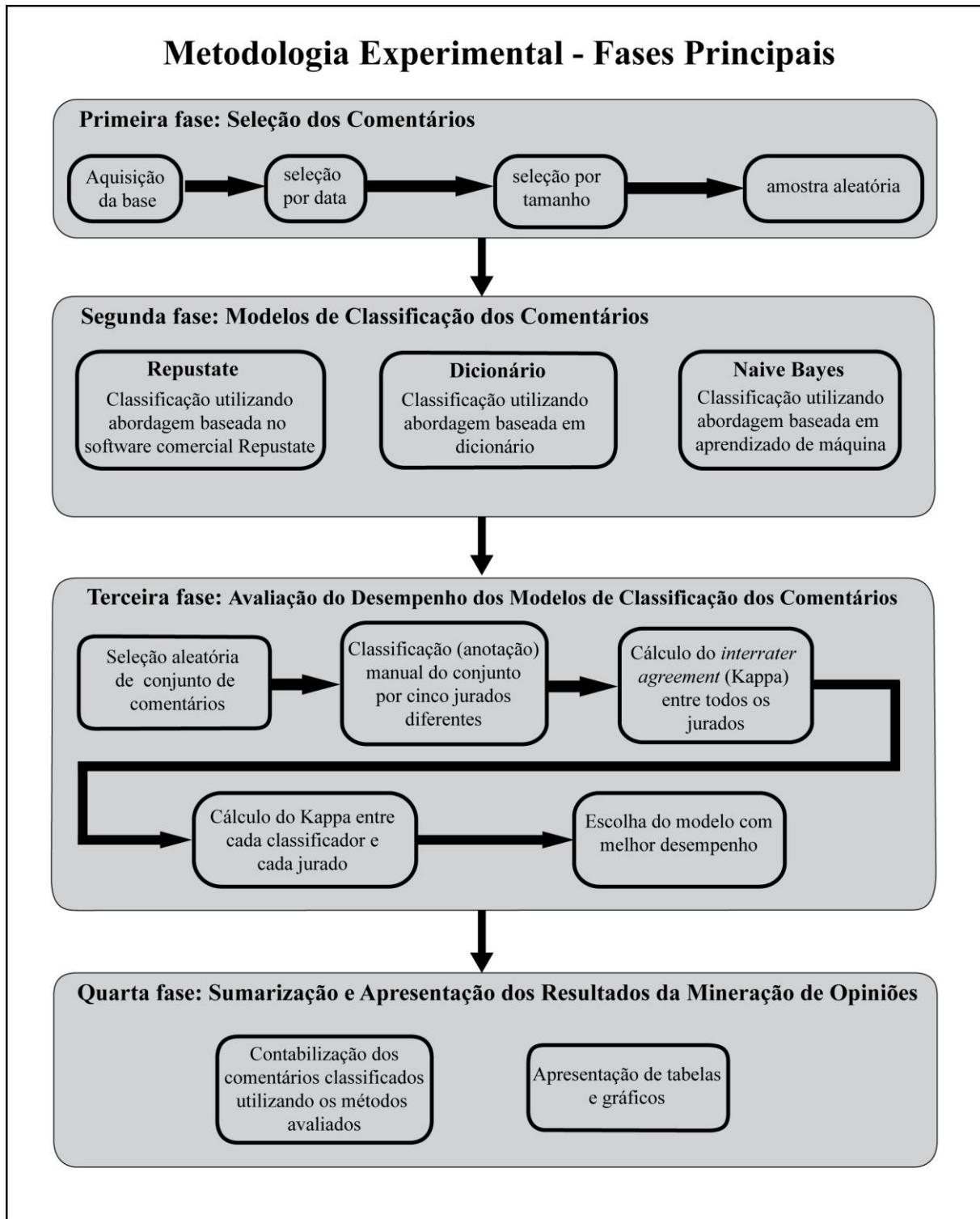


Figura 5 - Fases principais da Metodologia Experimental. Fonte: O Autor.

As quatro Fases são descritas a seguir:

- **Primeira Fase: Seleção dos Comentários**

Na primeira Fase foram utilizados dados de uma empresa de classificados *online* de empregos, que apesar de ter colaborado com a pesquisa autorizando a utilização de sua base

de dados, não permitiu a divulgação do seu nome. Os dados desta base foram coletados a partir de um formulário preenchido pelos clientes da empresa ao cancelar o serviço. Cabe ressaltar que o fato do cliente cancelar o serviço não significa necessariamente que seu sentimento seja negativo em relação ao serviço, pois como se trata de um serviço de classificados *online* de empregos, é comum que o cliente cancele o serviço ao se recolocar no mercado de trabalho.

Os parâmetros desta base são os seguintes:

- Código identificador do comentário;
- Código identificador do cliente;
- Data em que o comentário foi escrito;
- Comentário escrito pelo cliente;
- Nota do serviço, atribuída pelo cliente.

Os primeiros quatro parâmetros podem ser relacionados à quíntupla definida em Liu (2012). Neste trabalho, os comentários foram tratados de forma a expressar os sentimentos dos clientes sobre os serviços de forma geral, ou seja, no nível de documento, e não no nível de aspectos. Desta maneira, o valor dos parâmetros  $e_i$  e  $a_{ij}$  foi igual. Além desses parâmetros, o quinto parâmetro da base representou uma nota de 1 a 10, associada aos comentários, atribuída pelos clientes aos serviços prestados pela empresa. A Tabela 2 relaciona esses parâmetros com a quíntupla de opinião.

Tabela 2 - Parâmetros relacionados às opiniões dos usuários.

<b>Parâmetro</b>		<b>Descrição</b>
$e_i$	Serviços	A entidade é o serviço prestado pela empresa
$a_{ij}$	Serviços	GENERAL - Refere-se também à entidade, pois não se está utilizando nível de aspectos
$s_{ijkl}$	Sentimento	Refere-se ao sentimento associado ao serviço. Parâmetro a ser computado
$h_k$	Usuário	Usuário que emitiu a opinião. Identificado pelo seu código
$t_i$	Data	Data em que foi emitida a opinião

Fonte: O Autor

A base com os dados está armazenada em um banco de dados relacional, e possui mais de seis milhões de comentários. Neste trabalho, porém, foram considerados somente os comentários mais recentes, compreendendo o período de Janeiro a Junho de 2013, e um total de 594.788 comentários.

Com o objetivo de eliminar comentários espúrios e ao mesmo tempo selecionar aqueles que tenham pelo menos uma sentença completa, dentre os comentários selecionados anteriormente, selecionou-se aqueles que continham 50 ou mais caracteres. O número de comentários caiu, então, para 215.462. Parte desses comentários podem ser encontrados no Apêndice D.

Nesse número foram incluídos comentários curtos que ocorreram com frequência e que podem indicar opinião, como, por exemplo, comentários que se resumiram somente às seguintes palavras:

- “bom”;
- “muito bom”;
- “ruim”;
- “muito ruim”;
- “ótimo”;
- “excelente”;
- “péssimo”.

Porém, como ficará evidenciado nas próximas seções, devido a limitações na utilização dos *softwares* de tradução para a língua inglesa e para classificação dos comentários em um dos classificadores, do conjunto anterior de 215.462, selecionou-se aleatoriamente uma amostra de 2.500 comentários. Amostra essa, segundo (SAMOHYL, 2009), suficiente para representar o conjunto de comentários dos clientes da empresa.

- **Segunda Fase:** Modelos de Classificação dos Comentários

A fim de explorar os **Modelos de Classificação ou Classificadores** dos comentários que melhor se adaptaram ao problema proposto, optou-se por aplicar três deles:

- O primeiro Modelo de Classificação dos comentários aplicado utilizou um *software* comercial chamado Repustate (Repustate, 2014).
- O segundo Modelo de Classificação dos comentários aplicado utilizou um Dicionário chamado SentiLex (SILVA, CARVALHO e SARMENTO, 2012).
- O terceiro Modelo de Classificação dos comentários aplicado utilizou o aprendizado de máquina por meio da técnica Naive Bayes (PANG et. al., 2002; GAMALLO e GARCIA, 2014).

Os Modelos de Classificação são detalhados a seguir:

- **Modelo de Classificação dos Comentários com o Classificador baseado no Repustate**

Escolheu-se aplicar o Modelo de Classificação dos comentários baseado no Repustate, por se tratar de uma ferramenta consolidada no mercado (REPUSTATE, 2014a). Porém a ferramenta não oferece suporte à língua portuguesa. Por esse motivo, utilizou-se o artifício de realizar a tradução automática dos comentários antes de enviá-los à ferramenta, em um processo conhecido com Análise de Sentimentos entre Linguagens (*cross-language Sentiment Analysis*) (WAN, 2009; LIU, 2012). A Figura 6 apresenta o Modelo de Classificação dos comentários baseado no Repustate.

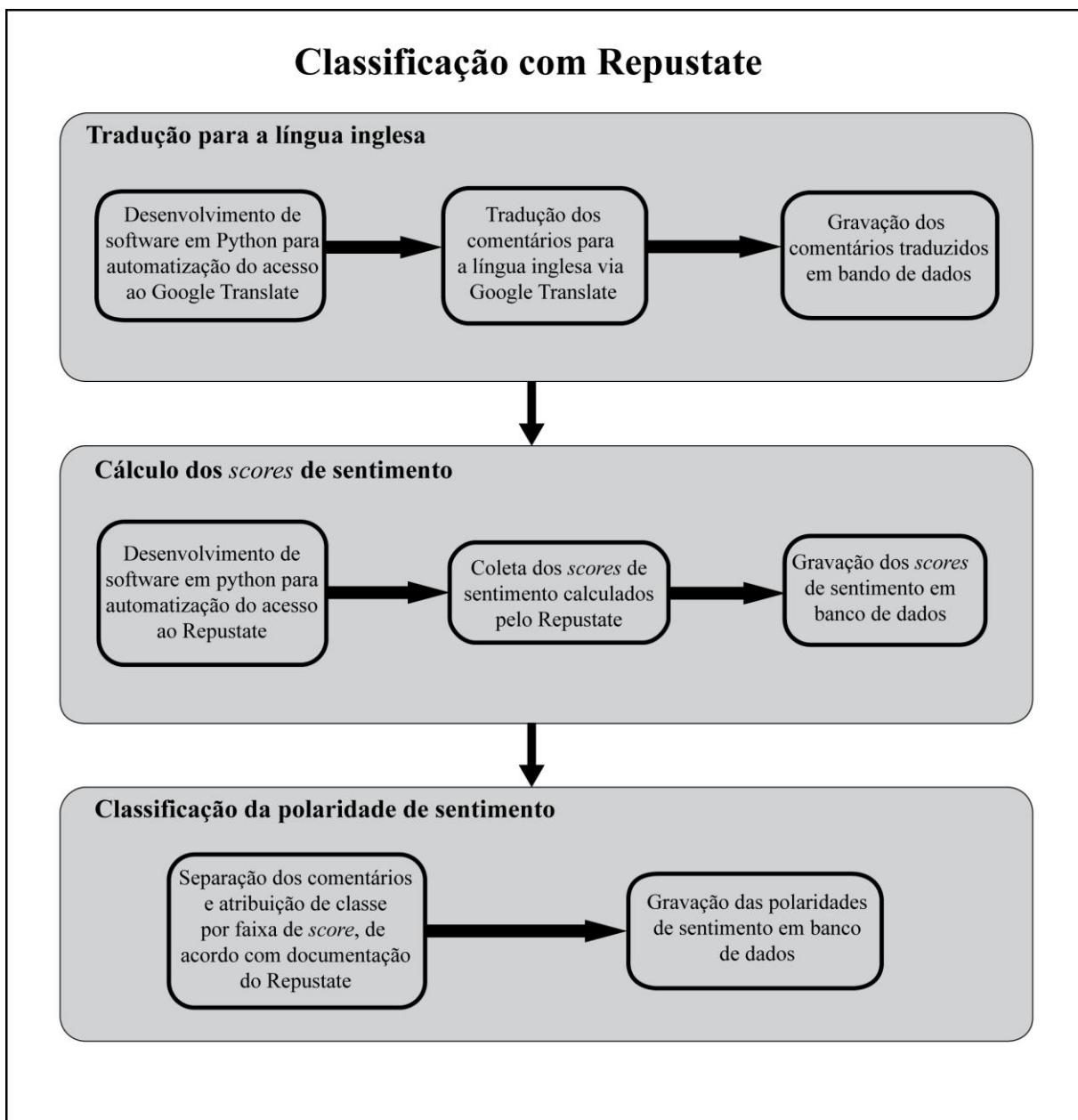


Figura 6 – Modelo de Classificação dos Comentários baseado no Repustate. Fonte: O Autor

O primeiro passo para a classificação dos comentários com o Repustate é a tradução para a língua inglesa. Utilizou-se para este fim, o *software* Google Translate (GOOGLE, 2013a).

Como a API (*Application Programming Interface*) fornecida pelo Google para tradução é um serviço pago (GOOGLE, 2013b), utilizou-se o serviço de tradução através do formulário disponível gratuitamente em Google (2013a). Porém, como este é um trabalho tedioso e propenso a erros se feito manualmente, desenvolveu-se um programa na linguagem Python (PYTHON SOFTWARE FOUNDATION, 2014), utilizando a ferramenta Selenium (SELENIUM, 2014) para automatizar esta tarefa (Apêndice A).

A ferramenta Selenium é normalmente utilizada para automação de testes funcionais para garantia de qualidade em *softwares* desenvolvidos para a *Web*. O *software* desenvolvido leu os comentários armazenados em banco de dados, simulando em seguida a entrada de cada um desses comentários no formulário do Google Translate. Após isso, colheu os resultados dessa mesma página, gravando-os em banco de dados.

Após a tradução dos comentários, realizou-se a obtenção dos *scores* de sentimento fornecidos pelo Repustate. Esses *scores*, de acordo com a documentação do *software* (REPUSTATE, 2013b) variam de -1 a 1, com valores positivos representando sentimento positivo, valores negativos, representando sentimentos negativos, e valores próximos de zero representando sentimentos neutros.

Novamente, pelo fato do acesso à API do Repustate ser pago, desenvolveu-se um *software* em Python com a ferramenta Selenium para automatizar as tarefas de ler os comentários traduzidos do banco de dados, enviá-los ao formulário, capturar os resultados e armazená-los em banco de dados. O software é apresentado no Apêndice B.

Finalmente, separou-se os comentários pela polaridade de sentimentos. Para a classificação dos comentários em relação à polaridade, foram consideradas três classes, para representar os comentários positivos, negativos e neutros.

*Scores* negativos representam sentimentos negativos, ou opinião desfavorável. *Scores* positivos representam sentimentos positivos ou favoráveis, e scores próximos de 0,0 representam neutralidade em relação aos sentimentos (REPUSTATE, 2013b). A Tabela 3 indica quais intervalos foram utilizados para considerar o sentimento expresso em um comentário como negativo, positivo ou neutro.

Tabela 3 - Intervalos de *score* para classificação de sentimentos com o Repustate

Intervalo	Sentimento
$-1,00 \leq score < -0,10$	Negativo
$-0,10 \leq score \leq 0,10$	Neutro
$0,10 < score \leq 1,00$	Positivo

Fonte: O Autor.

- **Modelo de Classificação dos Comentários com o Classificador Baseado em Dicionário.**

A outra abordagem utilizada para classificação foi baseada em dicionário, que se valeu de um léxico de sentimentos em português chamado SentiLex (SILVA, CARVALHO e SARMENTO, 2012). A Figura 7 apresenta o Modelo de Classificação dos comentários baseado em Dicionário.

O primeiro passo para a classificação dos comentários utilizando o método de classificação baseado em dicionário foi a obtenção de um léxico de sentimentos na língua portuguesa. Escolheu-se utilizar o SentiLex, um léxico para o português que possui, em sua versão atual, 7.014 lemas e 82.347 formas flexionadas. Em seguida, foi desenvolvido um *software* em Python para representar computacionalmente o SentiLex. O *software* leu o arquivo de formas flexionadas do SentiLex e criou uma estrutura de dados na qual foi possível realizar consultas por formas flexionadas.

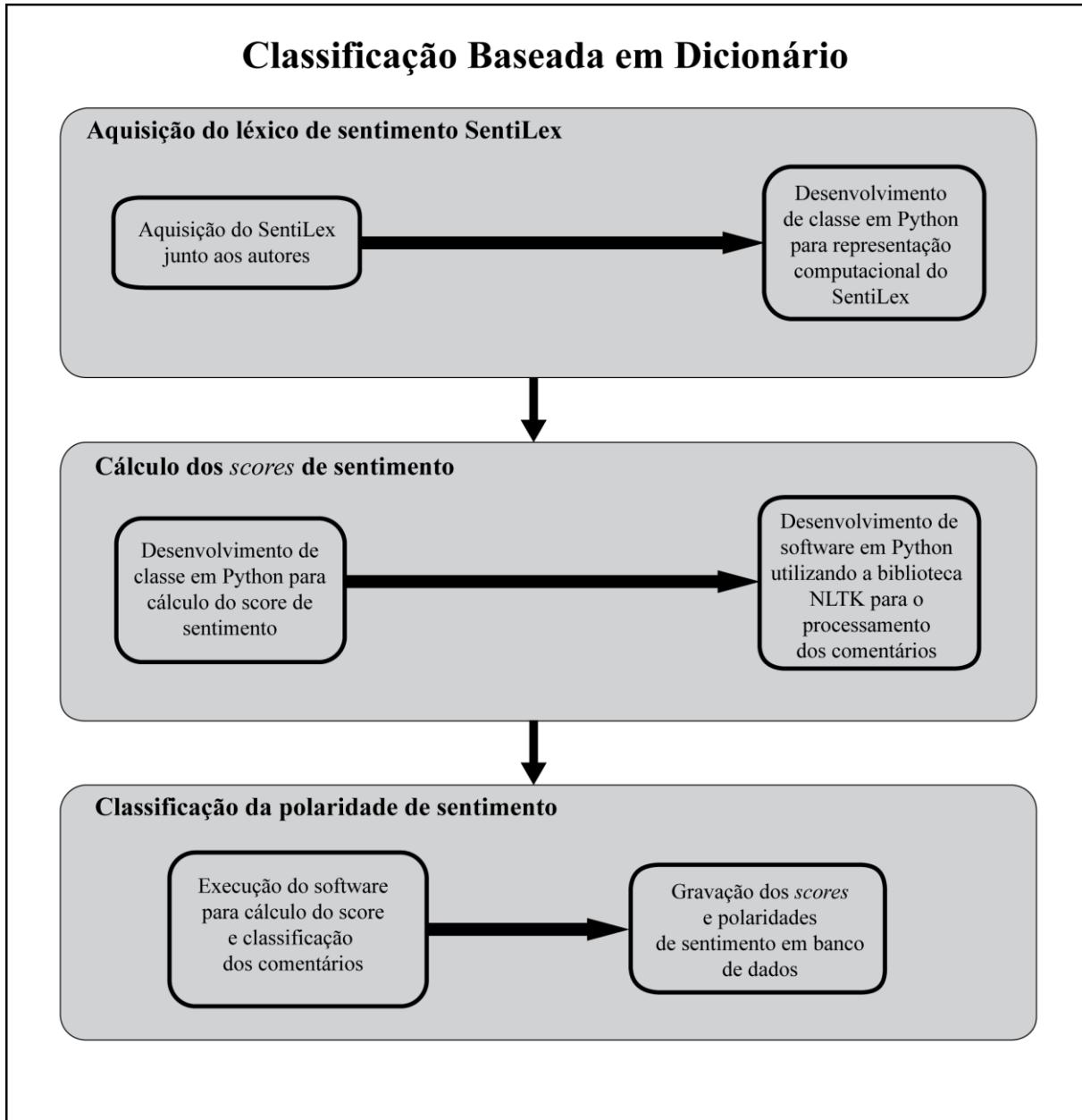


Figura 7 - Modelo de Classificação dos Comentários baseado em Dicionário. Fonte: O Autor.

O próximo passo foi calcular os *scores* de sentimento associados a cada comentário. Vale notar que esse *score* de sentimento não tem relação com o *score* de sentimento fornecido pelo Repustate. O processo para o cálculo dos *scores* de sentimentos está descrito a seguir:

- a. O comentário é dividido em sentenças;
- b. Para cada palavra na sentença, é feita uma busca no SentiLex-PT para ver se ela se encontra entre as formas flexionadas;
- c. Caso a forma flexionada seja encontrada, guarda-se o *score* de sentimentos referente àquela forma;

- d. Para a palavra onde se encontrou correspondência no SentiLex, verifica-se se existe palavra que representa negação, até duas posições antes de sua ocorrência. Caso a palavra representando negação seja encontrada, multiplica-se o score encontrado por -1;
- e. Somam-se todos os scores encontrados para a sentença;
- f. Repete-se o processo para todas as sentenças;
- g. Realiza-se a soma dos *scores* de todas as sentenças, sendo esse o *score* associado ao comentário.

Após o cálculo dos *scores* de sentimento foi realizada a classificação dos comentários da seguinte maneira:

- Comentários com  $score > 0$  foram classificados como positivos;
- Comentários com  $score < 0$  foram classificados como negativos;
- Comentários com  $score = 0$  foram classificados como neutro.

Após o cálculo do *score* e a classificação, esses resultados são armazenados em banco de dados.

- **Modelo de Classificação dos Comentários com o Classificador Baseado em Naive Bayes.**

O terceiro Método de Classificação dos comentários foi baseada em aprendizado de máquina foi o Naive Bayes. A Figura 8 apresenta o Modelo de Classificação dos comentários baseado no Naive Bayes.

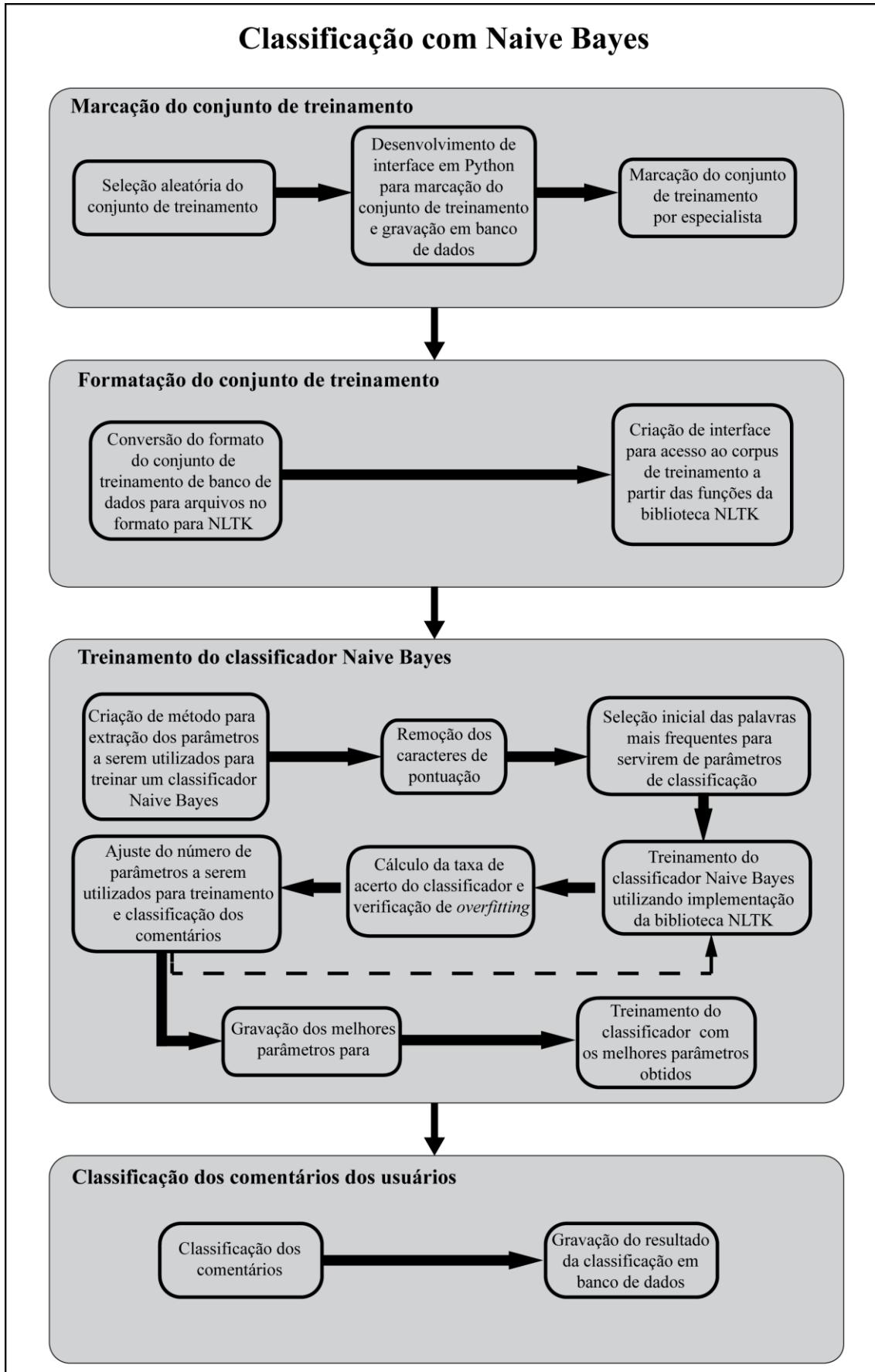


Figura 8 - Modelo de Classificação dos Comentários baseado no Naive Bayes. Fonte: O Autor.

Foi realizada inicialmente uma seleção aleatória na base de dados, separando seiscentos comentários. Foi então criado um *software* em Python para auxiliar a marcação desses comentários. O *software* apresentou cada um dos comentários a um especialista, oferecendo a opção de marcá-lo como positivo, negativo ou neutro. O próprio *software* foi encarregado de gravar a marcação em banco de dados.

Para possibilitar a utilização dos recursos da biblioteca NLTK (*Natural Language Toolkit*) (BIRD, LOPER e KLEIN, 2009), o conjunto de treinamento foi convertido para o formato utilizado pela NLTK.

Esse formato consiste em um diretório no sistema de arquivos do computador, contendo três subdiretórios para armazenar respectivamente os comentários positivos, neutros e negativos. Cada arquivo dentro desses diretórios contém somente um comentário, e o nome dos arquivos representam o código identificador do comentário. Depois de realizado esse procedimento, foi necessária a implementação de uma *interface* para tornar acessíveis os comentários como se fossem um *corpus* (conjunto de textos) da NLTK.

A fase seguinte consistiu em se realizar o treinamento do classificador Naive Bayes utilizando o conjunto de treinamento concebido nas etapas anteriores. Utilizou-se a implementação do classificador existente na própria biblioteca NLTK.

Criou-se então uma rotina computacional para extração dos parâmetros a serem utilizados pelo classificador. Em seguida, criou-se um filtro para desconsiderar caracteres de pontuação como parâmetro para classificação. Então, selecionaram-se as palavras que ocorreram com maior frequência no *corpus* de treinamento. Os parâmetros para treinamento e classificação passados ao classificador consistem em uma lista com essas palavras e a informação se cada palavra foi encontrada ou não no texto que está sendo classificado.

Iniciou-se o treinamento com as 100 palavras mais frequentes no *corpus*, separando-se uma parte do conjunto de treinamento para os testes. O conjunto de testes foi definido com 100 comentários, enquanto o conjunto de treinamento foi definido com 500 comentários. Foi então realizado o treinamento propriamente dito, iniciando-se um processo iterativo para determinação da quantidade ótima de parâmetros, aumentando-se a quantidade de parâmetros em passos de 50 palavras e observando-se o valor da acurácia da classificação para se determinar o ponto de *overfitting* (a memorização dos dados pelo modelo, ao invés da generalização).

Após encontrar a quantidade ótima de parâmetros para o classificador e o *corpus* utilizado, realizou-se o treinamento do classificador Naive Bayes com esses parâmetros. Então, efetuou-se a classificação dos comentários e a gravação dos resultados em banco de dados através de *software* desenvolvido na linguagem Python utilizando a biblioteca NLTK.

- **Terceira Fase:** Avaliação do Desempenho dos Modelos de Classificação dos Comentários

Para saber se uma dada caracterização/classificação de um objeto é confiável, é necessário ter este objeto caracterizado ou classificado várias vezes, por exemplo, por mais de um jurado ou juiz.

Assim, a Terceira Fase teve por objetivo medir o desempenho dos três classificadores. Para este fim, inicialmente foi solicitado a cinco jurados voluntários que examinassem um conjunto de 150 comentários e os classificassem como positivos, negativos ou neutros. Esse conjunto não possui intersecção com o conjunto de treinamento. O texto com as instruções aos jurados pode ser encontrado no Apêndice C, e os cinco jurados possuem o seguinte perfil:

- São brasileiros;
- São maiores de 21 anos de idade;
- Possuem curso superior;
- Possuem familiaridade com tecnologia.

O resultado da classificação realizada pelos cinco jurados foi armazenado em banco de dados.

Antes de verificar a concordância dos classificadores com os jurados, é importante verificar a concordância entre os próprios jurados. Isso é necessário, pois por se tratar de uma tarefa subjetiva, pode ser que os jurados não concordem em todas as classificações. Essa verificação serve para avaliar o quanto perto os classificadores chegam de um jurado humano.

Para avaliar o desempenho dos Modelos de Classificação dos comentários, ou seja, verificar a concordância (*interrater agreement*) entre os jurados e entre os jurados e os classificadores, foi utilizado o índice Kappa baseado na construção da Matriz de Confusão.

O índice Kappa é usado para descrever a concordância entre dois ou mais juízes, ou entre dois métodos de classificação (SIEGEL e CASTELLAN, 1988; BANERJEE et. al., 1999; FLEISS, LEVIN e PAIK, 2003). O cálculo de Kappa leva em consideração além da concordância percentual entre os jurados, também a concordância que seria esperada ao acaso, de acordo com a Equação (2).

$$K = \frac{P(c) - P(a)}{1 - P(a)} \quad (2)$$

Onde  $P(c)$  é a concordância observada e  $P(a)$  é a concordância esperada ao acaso, baseando-se nas observações para calcular as probabilidades de cada jurado atribuir aleatoriamente uma das classes. Quando os jurados concordam em todas as observações, o valor de  $K$  é igual a 1. Por outro lado, caso a concordância entre os jurados seja igual à concordância esperada ao acaso,  $K$  é igual a zero.

A Tabela 4 apresenta fornece a interpretação para os valores de Kappa, de acordo com os valores obtidos.

Tabela 4 - Interpretação dos valores de Kappa.

Intervalo	Nível de Concordância
< 0,00	Sem concordância
0,00 a 0,20	Fraca
0,21 a 0,40	Leve
0,41 a 0,60	Moderado
0,61 a 0,80	Bom
0,81 a 1,00	Muito bom

Fonte: Landis e Koch (1997)

A Matriz de Confusão quantifica quantos exemplos do conjunto seriam classificados bem pelo classificador utilizado. Os exemplos na diagonal principal desta matriz são os exemplos bem classificados, enquanto os exemplos fora dessa diagonal são os exemplos mal classificados (BERNARDES, 2006).

A Matriz de Confusão permite visualizar não somente a acurácia global de um dado classificador, mas também como este se comporta ao classificar os exemplos em cada categoria (CAMPBELL, 2002).

Na Tabela 5 é apresentado um exemplo de Matriz de Confusão entre um classificador e um jurado humano. Na diagonal principal da matriz (na cor cinza), estão representados os elementos que foram classificados da mesma forma tanto pelo jurado quanto pelo classificador. Os demais elementos indicam onde houve divergência na classificação.

Tabela 5–Exemplo de Matriz de confusão.  
Classificador

		Negativo	Neutro	Positivo	Total
Jurado	Negativo	17	17	3	37
	Neutro	10	13	12	35
	Positivo	5	17	56	78
	Total	32	47	71	<b>150</b>

Fonte: O Autor

Portanto, para a avaliação da concordância, montou-se as Matrizes de Confusão e calculou-se o índice Kappa para cada par de voluntários, na seguinte sequência: jurado 1 com jurado 2, jurado 1 com jurado 3, jurado 1 com jurado 4, jurado 1 com jurado 5, jurado 2 com jurado 3, jurado 2 com jurado 4, jurado 2 com jurado 5, jurado 3 com jurado 4, jurado 3 com jurado 5 e jurado 4 com jurado 5.

Após obter a concordância entre os jurados as Matrizes de Confusão foram montadas e o índice Kappa calculado para avaliar a concordância entre os jurados e os classificadores. Escolheu-se então o classificador com melhor desempenho, ou seja, o classificador cuja resposta mais se aproximou dos jurados humanos.

Em seguida foi realizada a análise da divergência entre os classificadores e os jurados. Para tanto, escolheu-se os comentários que foram classificados unanimemente pelos jurados, e que ao mesmo tempo divergiram dos classificadores. Dentre esses comentários, foram escolhidos exemplos, e realizada análise de cada um desses comentários a fim de explicar a divergência.

- **Quarta Fase:** Sumarização e Apresentação dos Resultados da Mineração de Opiniões

Nesta Fase os resultados foram sumarizados e apresentados. Foi criada uma tabela apresentando a porcentagem dos comentários que foram classificados em cada faixa pelo classificador que apresentou melhor desempenho.

Além disso, foram criados gráficos com o resultado da Mineração de Opiniões, utilizando cada um dos classificadores. Os gráficos mostraram a distribuição dos sentimentos agrupados mês a mês, no período estudado. Para cada classificador foram criados dois

gráficos, um com a distribuição absoluta dos sentimentos, e outro com a distribuição percentual.

No próximo capítulo são apresentados e discutidos os resultados experimentais obtidos pela classificação realizada pelos jurados e pelos Modelos de Classificação.

## 4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Neste capítulo são apresentados os resultados obtidos a partir dos experimentos realizados. Primeiramente são apresentados os resultados de cada um dos Modelos de Classificação dos comentários (Classificadores). Em seguida, são apresentados os resultados da classificação do conjunto de validação, realizada pelos jurados.

A etapa seguinte apresenta as Matrizes de Confusão e a concordância entre cada um dos jurados. A seguir são apresentados o cálculo da concordância entre cada um dos jurados e cada um dos métodos de classificação utilizados, além da análise e discussão sobre o desempenho dos classificadores.

Após esta etapa, foi realizada a análise dos comentários em que houve divergência entre os jurados e os classificadores. Finalmente a última seção apresenta a Mineração de Opiniões.

### 4.1 RESULTADOS DOS MODELOS DE CLASSIFICAÇÃO

A partir dos *scores* obtidos por meio do programa de computador desenvolvido pelo autor deste trabalho com o auxílio do Repustate, conforme descrito no capítulo de Materiais e Métodos, os comentários foram classificados entre positivos, negativos ou neutros. Esses comentários foram classificados de acordo com os intervalos definidos no capítulo anterior, na Tabela 4, conforme Miranda e Sassi (2014), e foram obtidos os resultados apresentados na Tabela 6.

Tabela 6 - Classificação dos Comentários com o Classificador baseado no Repustate

Sentimento	Quantidade	Porcentagem
Negativo	595	23,80
Neutro	650	26,00
Positivo	1255	50,20
<b>Total</b>	<b>2500</b>	<b>100,00</b>

Fonte: O Autor

Nos resultados obtidos com o classificador baseado no Repustate, observa-se uma predominância de comentários classificados como positivos, seguidos pelos comentários neutros e por último os comentários negativos.

Os mesmos comentários foram classificados pelo classificador baseado em Dicionário (SentiLex), conforme descrito no capítulo de Materiais e Métodos. Os resultados desta classificação são apresentados na Tabela 7.

Tabela 7 - Classificação dos comentários com o Classificador baseado em Dicionário (SentiLex)

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	573	22,92
Neutro	733	29,32
Positivo	1194	47,76
<b>Total</b>	<b>2500</b>	<b>100,00</b>

Fonte: O Autor

Observa-se uma distribuição semelhante com os resultados obtidos com o classificador baseado no Repustate, com predominância dos positivos, seguido dos neutros e por último os negativos.

Por último, os comentários foram classificados utilizando o classificador baseado em aprendizado de máquina usando a técnica Naive Bayes (conforme descrito no capítulo de Materiais e Métodos). Os resultados da classificação são apresentados na Tabela 8.

Tabela 8 - Classificação dos Comentários com Naive Bayes

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	672	26,88
Neutro	33	1,32
Positivo	1795	71,80
<b>Total</b>	<b>2500</b>	<b>100,00</b>

Fonte: O Autor

Ao observar os resultados da classificação com Naive Bayes, verifica-se que a distribuição das classes é bem diferente da obtida por meio das outras duas abordagens. Apesar da predominância de positivos, esta é muito mais acentuada. Observa-se ainda que os

comentários neutros chegam somente a pouco mais de 1,32%, número esse que acaba contrastando muito com os obtidos pelos outros classificadores.

#### 4.2 RESULTADOS DA VALIDAÇÃO PELOS JURADOS

Nesta seção são apresentados os resultados da classificação do conjunto de validação pelos jurados. A Tabela 9 até a Tabela 13 apresentam os resultados da classificação realizada por cada um dos jurados.

Tabela 9 - Classificação dos Comentários pelo Jurado 1

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	40	26,67
Neutro	36	24,00
Positivo	74	49,33
<b>Total</b>	150	100,00

Fonte: O Autor

Tabela 10 - Classificação dos Comentários pelo Jurado 2

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	41	27,33
Neutro	34	22,67
Positivo	75	50,00
<b>Total</b>	150	100,00

Fonte: O Autor

Tabela 11 - Classificação dos Comentários pelo Jurado 3

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	30	20,00
Neutro	40	26,67
Positivo	80	53,33
<b>Total</b>	<b>150</b>	<b>100,00</b>

Fonte: O Autor

Tabela 12 - Classificação dos Comentários pelo Jurado 4

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	46	30,67
Neutro	22	14,67
Positivo	82	54,67
<b>Total</b>	<b>150</b>	<b>100,00</b>

Fonte: O Autor

Tabela 13 - Classificação dos Comentários pelo Jurado 5

<b>Sentimento</b>	<b>Quantidade</b>	<b>Porcentagem</b>
Negativo	37	24,67
Neutro	35	23,33
Positivo	78	52,00
<b>Total</b>	<b>150</b>	<b>100,00</b>

Fonte: O Autor

Apesar do conjunto classificado pelos jurados se tratar da mesma amostra extraída aleatoriamente do conjunto inicial de 2500 comentários, conforme descrito em Materiais e Métodos, fica evidente que houve divergência na classificação, com se pode observar

comparando-se o número de elementos atribuídos a cada uma das classes, por cada jurado. Porém, ainda se observa que houve predominância dos comentários positivos.

#### 4.3 AVALIAÇÃO DA CONCORDÂNCIA ENTRE OS JURADOS

Observando-se os resultados da classificação entre os jurados, percebe-se que a concordância entre eles não é total. Porém, não fica claro, pela distribuição percentual mostrada na seção anterior, qual o nível de concordância entre eles, pois não foi evidenciado, para cada amostra, se houve classificação divergente ou não, ou qual foi o número de classificações divergentes. Nesta seção, por meio das Matrizes de Confusão, e do cálculo do Kappa, quantificou-se esta concordância.

São apresentadas a seguir as Tabelas contendo as Matrizes de Confusão, e o cálculo da concordância entre os jurados.

A Tabela 14 apresenta a Matriz de Confusão entre o Jurado 1 e Jurado 2.

Tabela 14 - Matriz de Confusão entre o Jurado 1 e o Jurado 2  
Jurado 2

		Negativo	Neutro	Positivo	Total
Jurado 1	Negativo	35	3	2	40
	Neutro	6	27	3	36
	Positivo	0	4	70	74
	Total	41	34	75	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 1 e 2 foram:

1. Concordância entre os jurados: 132 (88%)
2. Concordância esperada somente pelo acaso: 56,1 (37,4%)
3. Kappa = 0,808
4. Erro padrão do Kappa = 0,042
5. Intervalo de confiança 95%: 0,727 a 0,890
6. Concordância considerada **muito boa**

A Tabela 15 apresenta a Matriz de Confusão entre o Jurado 1 e Jurado 3.

Tabela 15 - Matriz de Confusão entre o Jurado 1 e o Jurado 3  
Jurado 3

		Negativo	Neutro	Positivo	Total
Jurado 1	Negativo	28	11	1	40
	Neutro	2	27	7	36
	Positivo	0	2	72	74
	Total	30	40	80	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 1 e 3 foram:

1. Concordância entre os jurados: 127 (84, 67%)
2. Concordância esperada somente pelo acaso: 57,1 (38,04%)
3. Kappa = 0,753
4. Erro padrão do Kappa = 0,046
5. Intervalo de confiança 95%: 0,663 a 0,842
6. Concordância considerada **boa**

A Tabela 16 apresenta a Matriz de Confusão entre o Jurado 1 e Jurado 4.

Tabela 16 - Matriz de Confusão entre o Jurado 1 e o Jurado 4  
Jurado 4

		Negativo	Neutro	Positivo	Total
Jurado 1	Negativo	38	1	1	40
	Neutro	8	20	8	36
	Positivo	0	1	73	74
	Total	46	22	82	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 1 e 4 foram:

1. Concordância entre os jurados: 131 (87,33%)
2. Concordância esperada somente pelo acaso: 58 (38,67%)
3. Kappa = 0,793
4. Erro padrão do Kappa = 0,043
5. Intervalo de confiança 95%: 0,727 a 0,890
6. Concordância considerada **muito boa**

A Tabela 17 apresenta a Matriz de Confusão entre o Jurado 1 e Jurado 5.

Tabela 17 - Matriz de Confusão entre o Jurado 1 e o Jurado 5  
Jurado 5

		Negativo	Neutro	Positivo	Total
Jurado 1	Negativo	34	4	2	40
	Neutro	3	29	4	36
	Positivo	0	2	72	74
	Total	37	35	78	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 1 e 5 foram:

1. Concordância entre os jurados: 135 (90%)

2. Concordância esperada somente pelo acaso: 56,7 (37,83%)
3. Kappa = 0,839
4. Erro padrão do Kappa = 0,037
5. Intervalo de confiança 95%: 0,763 a 0,915
6. Concordância considerada **muito boa**

A Tabela 18 apresenta a Matriz de Confusão entre o Jurado 2 e Jurado 3.

Tabela 18 - Matriz de Confusão entre o Jurado 2 e o Jurado 3  
Jurado 3

		Negativo	Neutro	Positivo	Total
Jurado 2	Negativo	28	13	0	41
	Neutro	2	24	8	34
	Positivo	0	3	72	75
	Total	30	40	80	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 2 e 3 foram:

1. Concordância entre os jurados: 124 (82,67%)
2. Concordância esperada somente pelo acaso: 57 (38,18%)
3. Kappa = 0,720
4. Erro padrão do Kappa = 0,048
5. Intervalo de confiança 95%: 0,626 a 0,813
6. Concordância considerada **boa**

A Tabela 19 apresenta a Matriz de Confusão entre o Jurado 2 e Jurado 4.

Tabela 19 - Matriz de Confusão entre o Jurado 2 e o Jurado 4  
Jurado 4

		Negativo	Neutro	Positivo	Total
Jurado 2	Negativo	38	2	1	41
	Neutro	7	19	8	34
	Positivo	1	1	73	75
	Total	46	22	82	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 2 e 4 foram:

1. Concordância entre os jurados: 130 (86,67%)
2. Concordância esperada somente pelo acaso: 58,6 (39,04%)
3. Kappa = 0,781
4. Erro padrão do Kappa = 0,044
5. Intervalo de confiança 95%: 0,695 a 0,867
6. Concordância considerada **boa**

A Tabela 20 apresenta a Matriz de Confusão entre o Jurado 2 e Jurado 5.

Tabela 20 - Matriz de Confusão entre o Jurado 2 e o Jurado 5  
Jurado 5

		Negativo	Neutro	Positivo	Total
Jurado 2	Negativo	34	6	1	41
	Neutro	3	27	4	34
	Positivo	0	2	73	75
	Total	37	35	78	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 2 e 5 foram:

1. Concordância entre os jurados: 132 (88%)
2. Concordância esperada somente pelo acaso: 56,1 (37,4%)
3. Kappa = 0,808
4. Erro padrão do Kappa = 0,042
5. Intervalo de confiança 95%: 0,727 a 0,890
6. Concordância considerada **muito boa**

A Tabela 21 apresenta a Matriz de Confusão entre o Jurado 3 e Jurado 4.

Tabela 21 - Matriz de Confusão entre o Jurado 3 e o Jurado 4  
Jurado 4

		Negativo	Neutro	Positivo	Total
Jurado 3	Negativo	29	1	0	30
	Neutro	16	16	8	40
	Positivo	1	5	74	80
	Total	46	22	82	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 3 e 4 foram:

1. Concordância entre os jurados: 119 (79,33%)
2. Concordância esperada somente pelo acaso: 58,8 (39,20%)

3.  $\text{Kappa} = 0,660$
4. Erro padrão do Kappa = 0,050
5. Intervalo de confiança 95%: 0,562 a 0,758
6. Concordância considerada **boa**

A Tabela 22 apresenta a Matriz de Confusão entre o Jurado 3 e Jurado 5.

Tabela 22 - Matriz de Confusão entre o Jurado 3 e o Jurado 5  
Jurado 5

		Negativo	Neutro	Positivo	Total
Jurado 3	Negativo	26	3	1	30
	Neutro	11	24	5	40
	Positivo	0	8	72	80
	Total	37	35	78	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 3 e 5 foram:

1. Concordância entre os jurados: 122 (81,33%)
2. Concordância esperada somente pelo acaso: 58,3 (38,89%)
3.  $\text{Kappa} = 0,695$
4. Erro padrão do Kappa = 0,050
5. Intervalo de confiança 95%: 0,596 a 0,793
6. Concordância considerada **boa**

A Tabela 23 apresenta a Matriz de Confusão entre o Jurado 4 e Jurado 5.

Tabela 23 - Matriz de Confusão entre o Jurado 4 e o Jurado 5  
Jurado 5

		Negativo	Neutro	Positivo	Total
Jurado 4	Negativo	37	8	1	46
	Neutro	0	19	3	22
	Positivo	0	8	74	82
	Total	37	35	78	<b>150</b>

Fonte: O Autor

Os dados calculados para os jurados 4 e 5 foram:

1. Concordância entre os jurados: 130 (86,67%)
2. Concordância esperada somente pelo acaso: 59,1 (39,41%)
3.  $\text{Kappa} = 0,780$
4. Erro padrão do Kappa = 0,044

5. Intervalo de confiança 95%: 0,693 a 0,866

6. Concordância considerada **boa**

Como observado, a concordância entre os jurados não foi total, fato que fica evidente quando se verifica a porcentagem de classificações coincidentes entre os jurados que varia de 79,33% a 90%. Essa divergência fica também evidente ao se examinar as Matrizes de Confusão.

Ao se observar os valores do índice Kappa calculados para os jurados, é possível quantificar esta concordância. Verifica-se que estes valores variam de 0,660 a 0,839, ou seja, indicam concordância variando de **boa** a **muito boa**.

Esses valores percentuais de classificações coincidentes e os valores do Kappa demonstram que classificar os comentários dos clientes da empresa em questão não é tarefa trivial, mesmo para seres humanos. Na próxima seção são apresentados os cálculos de concordância entre os jurados e os classificadores.

#### 4.4 AVALIAÇÃO DA CONCORDÂNCIA ENTRE OS JURADOS E OS CLASSIFICADORES

São apresentadas nesta seção as Tabelas contendo as Matrizes de Confusão e o cálculo do Kappa entre cada Jurado e os três classificadores (classificador baseado no Repustate, classificador baseado em Dicionário e classificador baseado em Naive Bayes). Em seguida, determinou-se o desempenho dos classificadores quando comparados a classificação feita pelo ser humano.

##### 4.4.1 Classificador Repustate com os Jurados

As Tabelas a seguir apresentam os resultados da classificação com o Repustate.

A Tabela 24 apresenta a Matriz de Confusão entre o Jurado 1 e o Repustate.

Tabela 24 - Matriz de Confusão entre o Jurado 1 e o Repustate.  
Repustate

		Negativo	Neutro	Positivo	Total
Jurado 1	Negativo	11	11	18	40
	Neutro	14	12	10	36
	Positivo	6	10	58	74
	Total	31	33	86	<b>150</b>

Fonte: O Autor

Os dados calculados para o Jurado 1 e o classificador Repustate foram:

1. Concordância entre classificador e o jurado: 81 (54%)
2. Concordância esperada somente pelo acaso: 58,6 (39,08%)
3.  $Kappa = 0,245$
4. Erro padrão do  $Kappa = 0,059$
5. Intervalo de confiança 95%: 0,130 a 0,360
6. Concordância considerada **leve**

A Tabela 25 apresenta a Matriz de Confusão entre o Jurado 2 e o Repustate.

Tabela 25 - Matriz de Confusão entre o Jurado 2 e o Repustate.

		Negativo	Neutro	Positivo	Total
Jurado 2	Negativo	13	11	17	41
	Neutro	13	9	12	34
	Positivo	5	13	57	75
	Total	31	33	86	<b>150</b>

Fonte: O Autor

Os dados calculados para o Jurado 2 e o classificador Repustate foram:

1. Concordância entre classificador e o jurado: 79 (54%)
2. Concordância esperada somente pelo acaso: 59 (39,30%)
3.  $Kappa = 0,220$
4. Erro padrão do  $Kappa = 0,059$
5. Intervalo de confiança 95%: 0,105 a 0,335
6. Concordância considerada **leve**

A Tabela 26 apresenta a Matriz de Confusão entre o Jurado 3 e o Repustate.

Tabela 26 - Matriz de Confusão entre o Jurado 3 e o Repustate.

		Repustate			Total
		Negativo	Neutro	Positivo	
Jurado 3	Negativo	10	7	13	30
	Neutro	11	13	16	40
	Positivo	10	13	57	80
	Total	31	33	86	<b>150</b>

Fonte: O Autor

Os dados calculados para o Jurado 3 e o classificador Repustate foram:

1. Concordância entre classificador e o jurado: 80 (53,33%)
2. Concordância esperada somente pelo acaso: 60,9 (40,58%)
3. Kappa = 0,215
4. Erro padrão do Kappa = 0,062
5. Intervalo de confiança 95%: 0,093 a 0,336
6. Concordância considerada **leve**

A Tabela 27 apresenta a Matriz de Confusão entre o Jurado 4 e o Repustate.

Tabela 27 - Matriz de Confusão entre o Jurado 4 e o Repustate.

		Repustate			Total
		Negativo	Neutro	Positivo	
Jurado 4	Negativo	12	16	18	46
	Neutro	10	5	7	22
	Positivo	9	12	61	82
	Total	31	33	86	<b>150</b>

Fonte: O Autor

Os dados calculados para o Jurado 4 e o classificador Repustate foram:

1. Concordância entre classificador e o jurado: 78 (52%)
2. Concordância esperada somente pelo acaso: 61,4 (40,91%)
3. Kappa = 0,188
4. Erro padrão do Kappa = 0,057
5. Intervalo de confiança 95%: 0,077 a 0,299
6. Concordância considerada **fraca**

A Tabela 28 apresenta a Matriz de Confusão entre o Jurado 5 e o Repustate.

Tabela 28 - Matriz de Confusão entre o Jurado 5 e o Repustate.

		Repustate			Total
		Negativo	Neutro	Positivo	
Jurado 5	Negativo	11	11	15	37
	Neutro	14	9	12	35
	Positivo	6	13	59	78
	Total	31	33	86	<b>150</b>

Fonte: O Autor

Os dados calculados para o Jurado 4 e o classificador Repustate foram:

1. Concordância entre classificador e o jurado: 78 (52,67%)
2. Concordância esperada somente pelo acaso: 60,1 (40,04%)
3.  $Kappa = 0,211$
4. Erro padrão do  $Kappa = 0,059$
5. Intervalos de confiança 95%: 0,096 a 0,325
6. Concordância considerada **leve**

Para o classificador baseado no Repustate, a concordância variou entre fraca e leve, com  $Kappa$  variando de 0,188 a 0,245.

#### 4.4.2 Classificador Baseado em Dicionário com os Jurados

As Tabelas a seguir apresentam os resultados da avaliação de concordância entre o classificador baseado em dicionário usando o SentiLex e os jurados.

A Tabela 29 apresenta a Matriz de Confusão entre o Jurado 1 e o classificador baseado em Dicionário.

Tabela 29 - Matriz de Confusão entre o Jurado 1 e o classificador baseado em Dicionário.

		Dicionário			Total
		Negativo	Neutro	Positivo	
Jurado 1	Negativo	16	20	4	40
	Neutro	11	10	15	36
	Positivo	5	17	52	74
	Total	32	47	71	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 1 e o classificador baseado em Dicionário foram:

1. Concordância entre classificador e o jurado: 78 (52%)

2. Concordância esperada somente pelo acaso: 54,8 (36,56%)
3. Kappa = 0,243
4. Erro padrão do Kappa = 0,059
5. Intervalo de confiança 95%: 0,128 a 0,358
6. Concordância considerada **leve**

A Tabela 30 apresenta a Matriz de Confusão entre o Jurado 2 e o classificador baseado em Dicionário.

Tabela 30 - Matriz de Confusão entre o Jurado 2 e o classificador baseado em Dicionário.

		Dicionário			Total
		Negativo	Neutro	Positivo	
Jurado 2	Negativo	18	19	4	41
	Neutro	9	10	15	34
	Positivo	5	18	52	75
	Total	32	47	71	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 2 e o classificador baseado em Dicionário foram:

1. Concordância entre classificador e o jurado: 80 (53,33%)
2. Concordância esperada somente pelo acaso: 53,9 (36,60%)
3. Kappa = 0,264
4. Erro padrão do Kappa = 0,059
5. Intervalo de confiança 95%: 0,148 a 0,380
6. Concordância considerada **leve**

A Tabela 31 apresenta a Matriz de Confusão entre o Jurado 3 e o classificador baseado em Dicionário.

Tabela 31 - Matriz de Confusão entre o Jurado 3 e o classificador baseado em Dicionário.

		Dicionário			Total
		Negativo	Neutro	Positivo	
Jurado 3	Negativo	15	12	3	30
	Neutro	11	16	13	40
	Positivo	6	19	55	80
	Total	32	47	71	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 3 e o classificador baseado em Dicionário foram:

1. Concordância entre classificador e o jurado: 86 (57,33%)

2. Concordância esperada somente pelo acaso: 56,8 (37,87%)
3. Kappa = 0,313
4. Erro padrão do Kappa = 0,061
5. Intervalo de confiança 95%: 0,194 a 0,433
6. Concordância considerada **leve**

A Tabela 32 apresenta a Matriz de Confusão entre o Jurado 4 e o classificador baseado em Dicionário.

Tabela 32 - Matriz de Confusão entre o Jurado 4 e o classificador baseado em Dicionário.

		Dicionário			
		Negativo	Neutro	Positivo	Total
Jurado 4	Negativo	21	20	5	46
	Neutro	6	8	8	22
	Positivo	5	19	58	82
	Total	32	47	71	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 4 e o classificador baseado em Dicionário foram:

1. Concordância entre classificador e o jurado: 87 (58,00%)
2. Concordância esperada somente pelo acaso: 55,5(37,01%)
3. Kappa = 0,333
4. Erro padrão do Kappa = 0,056
5. Intervalo de confiança 95%: 0,223 a 0,444
6. Concordância considerada **leve**

A Tabela 33 apresenta a Matriz de Confusão entre o Jurado 5 e o classificador baseado em Dicionário.

Tabela 33 - Matriz de Confusão entre o Jurado 5 e o classificador baseado em Dicionário.

		Dicionário			
		Negativo	Neutro	Positivo	Total
Jurado 5	Negativo	17	17	3	37
	Neutro	10	13	12	35
	Positivo	5	17	56	78
	Total	32	47	71	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 5 e o classificador baseado em Dicionário foram:

1. Concordância entre classificador e o jurado: 86 (57,33%)
2. Concordância esperada somente pelo acaso: 55,8 (37,19%)

3.  $\text{Kappa} = 0,321$
4. Erro padrão do Kappa = 0,059
5. Intervalo de confiança 95%: 0,205 a 0,437
6. Concordância considerada **leve**

Para o classificador baseado em Dicionário, a concordância foi leve, com Kappa variando de 0,243 a 0,333. Portanto, observou-se que o classificador baseado em Dicionário apresentou performance melhor do que o classificador baseado no Repustate.

#### 4.4.3 Classificador Baseado em Naive Bayes com os Jurados

As Tabelas a seguir apresentam os resultados da avaliação de concordância entre o classificador baseado em Naive Bayes e os jurados.

A Tabela 34 apresenta a Matriz de Confusão entre o Jurado 1 e o classificador baseado em Naive Bayes.

Tabela 34 - Matriz de Confusão entre o Jurado 1 e o classificador baseado em Naive Bayes.

		Naive Bayes			
		Negativo	Neutro	Positivo	Total
Jurado 1	Negativo	20	0	20	40
	Neutro	11	1	24	36
	Positivo	2	1	71	74
	Total	33	2	115	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 1 (linhas) e Naive Bayes (colunas) foram:

1. Concordância entre classificador e o jurado: 92 (61,33%)
2. Concordância esperada somente pelo acaso: 66 (44,01%)
3.  $\text{Kappa} = 0,309$
4. Erro padrão do Kappa = 0,054
5. Intervalo de confiança 95%: 0,204 a 0,415
6. Concordância considerada **leve**

A Tabela 35 apresenta a Matriz de Confusão entre o Jurado 2 e o classificador baseado em Naive Bayes.

Tabela 35 - Matriz de Confusão entre o Jurado 2 e o classificador baseado em Naive Bayes.

		Naive Bayes			
		Negativo	Neutro	Positivo	Total
Jurado 2	Negativo	23	0	18	41
	Neutro	9	1	24	34
	Positivo	1	1	73	75
	Total	33	2	115	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 2 (linhas) e Naive Bayes (colunas) foram:

1. Concordância entre classificador e o jurado: 97 (64,67%)
2. Concordância esperada somente pelo acaso: 67 (44,65%)
3.  $Kappa = 0,362$
4. Erro padrão do  $Kappa = 0,055$
5. Intervalo de confiança 95%: 0,254 a 0,469
6. Concordância considerada **leve**

A Tabela 36 apresenta a Matriz de Confusão entre o Jurado 3 e o classificador baseado em Naive Bayes.

Tabela 36 - Matriz de Confusão entre o Jurado 3 e o classificador baseado em Naive Bayes.

		Naive Bayes			
		Negativo	Neutro	Positivo	Total
Jurado 3	Negativo	19	0	11	30
	Neutro	10	1	29	40
	Positivo	4	1	75	80
	Total	33	2	115	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 3 (linhas) e Naive Bayes (colunas) foram:

1. Concordância entre classificador e o jurado: 95 (63,33%)
2. Concordância esperada somente pelo acaso: 68,5 (45,64%)
3.  $Kappa = 0,325$
4. Erro padrão do  $Kappa = 0,056$
5. Intervalo de confiança 95%: 0,216 a 0,435
6. Concordância considerada **leve**

A Tabela 37 apresenta a Matriz de Confusão entre o Jurado 4 e o classificador baseado em Naive Bayes.

Tabela 37 - Matriz de Confusão entre o Jurado 4 e o classificador baseado em Naive Bayes.

		Naive Bayes			
		Negativo	Neutro	Positivo	Total
Jurado 4	Negativo	23	0	23	46
	Neutro	7	2	13	22
	Positivo	3	0	79	82
	Total	33	2	115	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 4 e o classificador baseado em Naive Bayes foram:

1. Concordância entre classificador e o jurado: 104 (69,33%)
2. Concordância esperada somente pelo acaso: 73,3 (48,85%)
3. Kappa = 0,400
4. Erro padrão do Kappa = 0,061
5. Intervalo de confiança 95%: 0,280 a 0,520
6. Concordância considerada **moderada**

A Tabela 38 apresenta a Matriz de Confusão entre o Jurado 5 e o classificador baseado em Naive Bayes.

Tabela 38 - Matriz de Confusão entre o Jurado 5 e o classificador baseado em Naive Bayes.

		Naive Bayes			
		Negativo	Neutro	Positivo	Total
Jurado 5	Negativo	20	0	17	37
	Neutro	11	1	23	35
	Positivo	2	1	75	78
	Total	33	2	115	<b>150</b>

Fonte: O Autor

Os dados calculados entre o Jurado 5 (linhas) e Naive Bayes (colunas) foram:

1. Concordância entre classificador e o jurado: 96 (64%)
2. Concordância esperada somente pelo acaso: 68,4 (45,60%)
3. Kappa = 0,338
4. Erro padrão do Kappa = 0,055
5. Intervalo de confiança 95%: 0,230 a 0,446
6. Concordância considerada **leve**

Para o classificador baseado em Naive Bayes, a concordância variou de leve a moderada, com os valores de Kappa variando de 0,309 a 0,400. Portanto, o classificador Naive Bayes superou os outros classificadores no quesito concordância.

A seguir são apresentados na Tabela 39 os valores percentuais de classificações coincidentes entre os classificadores e os jurados. São apresentados os valores mínimo e máximo para cada classificador.

Tabela 39–Percentuais de coincidência dos classificadores com os jurados

Classificador	Valor mínimo	Valor máximo
Repustate	52%	54%
Dicionário	52%	58%
NaiveBayes	61,33%	69,33%

Fonte: O Autor

O classificador baseado em Naive Bayes claramente se sobressaiu aos demais em relação ao percentual de classificações coincidentes com os jurados. A seguir, na Tabela 40, são apresentadas as médias dos valores de Kappa para cada um dos três classificadores.

Tabela 40 - Médias do Valor de Kappa para cada um dos Três Classificadores

Classificador	Valor médio do Kappa
Repustate	0,2158
Dicionário	0,2948
Naive Bayes	0,3468

Fonte: O Autor

Apesar da concordância indicada pelos valores médios das três abordagens serem consideradas leves, a abordagem baseada em Naive Bayes claramente se sobressaiu. Porém, observando-se as Matrizes de Confusão obtidas ao se comparar o classificador Naive Bayes e os jurados percebe-se que o classificador Naive Bayes encontrou dificuldades em classificar comentários neutros. Esta dificuldade foi observada em outros trabalhos que utilizaram Naive Bayes para análise de sentimentos (GAMALLO, GARCIA e FERNÁNDEZ-LANZA, 2013; GAMALLO e GARCIA, 2014).

## 4.5 ANÁLISE DA DIVERGÊNCIA ENTRE JURADOS E CLASSIFICADORES

Foram analisados e discutidos os resultados em que os classificadores divergiram dos jurados, em busca das suas limitações. Para este fim, foram comparados os resultados dos classificadores com os comentários que tiveram classificação unânime entre os jurados, pois foi considerado que os comentários que não foram classificados unanimemente entre os jurados não são confiáveis para comparação.

Vale ressaltar que os comentários sublinhados apresentam a mesma grafia, inclusive os erros, como foram encontrados na base de dados da empresa. Somente menções ao nome da empresa, aos preços praticados ou às palavras inapropriadas foram removidas.

### 4.5.1 Classificador Repustate com os Jurados

A análise dos resultados do classificador baseado no Repustate se atreve à estrutura e complexidade dos comentários. O motivo disso é que por se tratar de uma ferramenta comercial, o conhecimento do funcionamento interno do classificador é restrito à empresa que desenvolveu a ferramenta.

Os comentários a seguir foram classificados como **neutros** pelo classificador baseado em Repustate, e como **negativos** pelos jurados. Foram encontrados cinco comentários seguindo esse critério. Dos cinco comentários encontrados, três são apresentados a seguir:

- “não vejo verdade neste serviço, e particularmente não é confiavel.” : Este comentário possui erro de grafia na palavra “confiável”. Além disso, as duas construções que denotam negatividade são expressões positivas com negação.
- “enviamos curriculo e nao temos retorno das empresas”: Além dos erros de grafia, temos a construção que denota negatividade (“não temos retorno”) é específica do domínio.
- “Por se tratar de trabalhos on-lines acho que a Empresa deveria entrar em contato com o e ajuda-lo melhor. voces colocam um site, pagamos e olha agora é com voces não 'bem por aih'”: As construções linguísticas que denotam negatividade possuem muitos erros de grafia.

Os comentários a seguir foram classificados como **positivos** pelo classificador baseado em Repustate, e como **negativos** pelos jurados. Foram encontrados doze comentários seguindo esse critério. Dos doze comentários encontrados, quatro são apresentados a seguir:

- “Por ser um site pago deveria haver mais qualidades nas empresas anunciantes. Todo esse tempo de assinante e com um curriculum interessante não obtive sucesso com a Empresa.”: Comentário apresenta uma opinião indireta. Além disso, o fato do adjetivo “interessante” ser relacionado ao substantivo “currículo” e não ao serviço da empresa, interferiu no resultado.
- “Recebi recomendação de uma amiga, porém fiquei insatisfeita com os serviços da Empresa. Sugiro atendimento online para o assinante assim é possível uma comunicação mais efetiva. Obrigada.”: As palavras “recomendação” e “efetiva” indicam positividade, porém a palavra “insatisfeita”, que indica negatividade deveria ter tido um peso maior na definição da polaridade do comentário.
- “Consegui mai entrevistas em sites grátis que na catho.”: Além de erro de grafia (“mais” ao invés de “mais”), é uma opinião comparativa, que segundo a literatura (Liu, 2012) é mais difícil de ser classificada.
- “tem sites gratuitos melhores que esse na minha opinião”: É uma opinião comparativa.

Os comentários a seguir foram classificados como **negativos** pelo classificador baseado em Repustate, e como **neutros** pelos jurados. Dos cinco comentários encontrados, dois são apresentados a seguir:

- “Empresa boa e preço salgado. O serviço é otimo pena que seja tão caro.”: Erro de grafia na palavra “ótimo”, que é fortemente positiva.
- “Site que auxilia, mas, deveria apresentar status de vaga...ser urgente...contratação imedita..informar local, pois nao sao todos oscasos que apresenta estas informações...”: Muitos erros de grafia, além de possuir redação confusa.

Os comentários a seguir foram classificados como **positivos** pelo classificador baseado em Repustate, e como **neutros** pelos jurados. Os três comentários encontrados são apresentados a seguir:

- “Conheço muitas pessoas que se recolocaram através da Empresa, porém como já informei o material que eu vendo é mais técnico.”: O comentário expressa somente fatos, e não opiniões.
- “o site é bom mas as empresas não retornam quando vc manda seu cv.”: O comentários contém muitas abreviações.
- “Acervo muito bom de vagas! Mas parece que existem anuncios que nao sao levados em consideracao pelas empresas!”: O classificador detectou somente a opinião positiva, e não a opinião negativa contida no comentário.

Os comentários a seguir foram classificados como **negativos** pelo classificador baseado em Repustate, e como **positivos** pelos jurados. Dos cinco comentários encontrados, dois são apresentados a seguir:

- “É legal, tem bastante vaga e manda todos os dias a relação”: Comentário aparentemente simples, com opinião direta. O modelo errou ao classificá-lo como negativo.
- “É uma empresa legal, que ajuda bastante aos desempregados com os 7 dias gratis.”: Comentário positivo, que possui a palavra desempregado, que em geral possui significado negativo.

Os comentários a seguir foram classificados como **neutros** pelo classificador baseado em Repustate, e como **positivos** pelos jurados. Dos dez comentários encontrados, três são apresentados a seguir:

- “Otima maneira de encontrar vagas para meu perfil profissional”: Existe erro de grafia na única palavra que com sentido positivo (ótima).
- “Foi o site em que obtive contato com as melhores vagas, das empresas maiores.”: O comentário expressa uma opinião comparativa, que segundo a literatura (Liu, 2012) apresenta maior dificuldade para determinação da polaridade. Além disso, o comentário contém erro de grafia na palavra “maiores”.
- “Muito bom site de procura de emprego me ajudou em momentos difíceis”: Palavra “bom” e palavra “difícil” na mesma frase dificultaram a determinação da polaridade pelo modelo.

Verificou-se que o modelo baseado no Repustate apresentou dificuldade ao classificar comentários com erros de grafia. Além disso, opiniões comparativas ou opiniões indiretas

também não foram classificadas corretamente. Este foi o modelo que apresentou pior desempenho (Tabela 40). Porém, a vantagem de se utilizar esse modelo é que ao se utilizar uma ferramenta externa para avaliação da polaridade de sentimento, não é necessário implementar e ajustar o classificador, sendo o modelo que pode ser implementado mais rapidamente.

#### 4.5.2 Classificador Baseado em Dicionário com os Jurados

A análise dos resultados do classificador baseado em Dicionário consistiu na verificação do score calculado pelo modelo, cujo cálculo é descrito no capítulo 3, e nas palavras encontradas no SentiLex. Quando forem apresentadas as palavras encontradas no SentiLex, serão descritas no formato “palavra” (polaridade).

Os comentários a seguir foram classificados como **neutros** pelo classificador baseado em Dicionário, e como **negativos** pelos jurados. Dos dez comentários encontrados, quatro são apresentados a seguir:

- “Perda de tempo, e quando optei por cancelar me ligaram insistindo que permanecesse. Resultado, perdi mais dinheiro. NÃO DESEJO SER CONTATADA POR TELEFONE PARA REATIVAÇÃO DA ASSINATURA!”: Score 0. O sistema encontrou somente a palavra “perdi”. O SentiLex apresenta esta palavra com dois possíveis valores de sentimento: -1 ou 0. No contexto deste comentário o correto seria o valor -1, porém, por um detalhe de implementação o sistema escolheu o valor 0.
- “Acho que falam muito em oportunidade, porém se as pessoas estão procurando emprego precisam de dinheiro, e tem que pagar pra conseguir enviar um curriculo, não dá.”: Score 0. O sistema não encontrou nenhuma palavra no SentiLex.
- “Esperava maior interação da Empresa com os assinantes. Não adianta enviar centenas de vagas, que na verdade são pseudo-vagas, pois não se encaixam no perfil do candidato.”: Score 0. O sistema encontrou a flexão “vagas”(-1). No caso deste comentário a palavra “vagas” é um substantivo, e portanto não detectar esta diferença é uma deficiência do sistema. Foi encontrada também a palavra “verdade”(1).

- “me cobraram XX,XX reais sendo que eu não autorizei essa m.....”: Score 0. Encontrou as palavras “reais”(1) e m....(-1). A palavra reais não deveria ser considerada como palavra de opinião neste caso, pois se refere à moeda.

Os comentários a seguir foram classificados como **positivos** pelo classificador baseado em Dicionário, e como **negativos** pelos jurados. Os dois comentários encontrados são apresentados a seguir:

- “Por ser um site pago deveria haver mais qualidades nas empresas anunciantes. Todo esse tempo de assinante e com um curriculum interessante não obtive sucesso com a Empresa..”: Score 2. Achou a palavra “sucesso” (1), porém o modelo não considerou a negação dessa palavra, pois está a duas palavras de distância. A palavra “interessante” (1) refere-se ao currículo profissional do assinante, e não ao serviço prestado pela empresa.
- “não vejo verdade neste serviço, e particularmente não é confiavel..”: Score 1. Encontrou a palavra “verdade” (1), mas não computou a negação. A palavra confiável não foi encontrada devido ao erro de grafia.

Os comentários a seguir foram classificados como **negativos** pelo classificador baseado em Dicionário, e como **neutros** pelos jurados. Os três comentários encontrados são apresentados a seguir:

- “Conheço muitas pessoas que se recolocaram através da Empresa, porém como já informei o material que eu vendo é mais técnico..”: Score -1. Somente a palavra “vendo” (-1) foi encontrada no SentiLex. Porém, esta palavra, neste contexto, não possui polaridade negativa.
- “Fiquei pouco tempo, não tenho muito á dizer, mas obrigada..”: Score -1. Somente a palavra “obrigada” (-1) foi encontrada no SentiLex. No SentiLex, a classe grammatical da palavra “obrigada” é um adjetivo, e nesse caso significa coagido. No caso do comentário, é uma interjeição.
- “Serviços satisfatórios, porém ineficiente no cancelamento..”: Score -1. Somente a palavra “ineficiente” (-1) foi encontrada no SentiLex.

Os comentários a seguir foram classificados como **positivos** pelo classificador baseado em Dicionário, e como **neutros** pelos jurados. Os três comentários encontrados são apresentados a seguir:

- “o site é bom mas as empresas não retornam quando vc manda seu cv.”: Score 1. Somente a palavra “bom” (1) foi encontrada no SentiLex.
- “Gostei bastante da Empresa, porém não tive ainda nenhum retorno das empresas em que enviei o meu curriculo.”: Score 1. Somente a palavra “gostei” (1) foi encontrada no SentiLex.
- “Acho que a Empresa pode ser um bom local para se procurar emprego, porém não me satisfez.”: Score 1. Somente a palavra “bom” foi encontrada no SentiLex.

Os comentários a seguir foram classificados como **negativos** pelo classificador baseado em Dicionário, e como **positivos** pelos jurados. Dos cinco comentários encontrados, quatro são apresentados a seguir.

- “É um site com grande facilidade de efetuar diferentes buscas de vagas.”: Score -1. Foram encontradas no SentiLex as palavras “diferentes” (0) e “vagas” (-1). A palavra vagas no SentiLex existe como adjetivo, e nesse caso sua polaridade é -1. No texto, a palavra vaga é um substantivo, e deveria, neste contexto, ter a polaridade 0. Além disso, a palavra “facilidade”, não encontrada no SentiLex, indica polaridade positiva.
- “Foi o site em que obtive contato com as melhores vagas, das empresas maiores.”: Score -1. Foram encontradas no SentiLex as palavras “melhores” (0) e “vagas” (-1). Novamente a palavra “vaga” no comentário é um substantivo, enquanto que no SentiLex esta palavra existe somente como adjetivo.
- “é um ótimo site. Sempre mandam as vagas de acordo com o meu perfil, tem um acesso fácil e consigo buscar vagas do meu interesse facilmente.”: Score -1. Foram encontradas no SentiLex as palavras “ótimo” (1), “vagas” (-1), “fácil” (0) e novamente “vagas” (-1). A polaridade atribuída à palavra vaga, neste caso, teve um peso decisivo para a definição da polaridade do comentário como negativa ao invés de positiva.
- “Uma ótima empresa relacionada a oportunidades de um novo trabalho, mantendo informado sobre as vagas oferecidas por empresas de todo o Brasil.”: Score -1. Foram encontradas as palavras “ótima” (1), “vagas” (-1) e “oferecidas” (-1).

Os comentários a seguir foram classificados como **neutros** pelo classificador baseado em Dicionário, e como **positivos** pelos jurados. Dos quinze comentários encontrados, três são apresentados a seguir.

- “um site que ajuda bastante, interativo de facil manuseio.”: Score 0. Foi encontrada a palavra “fácil” (0) no SentiLex. Porém, no comentário, as palavras “fácil manuseio” apresentam polaridade positiva.
- “Obrigada, primeiramente por me ajudarem, de certa forma, na minha nova fase e parabenizo a Empresa por todas essas ferramentas oferecidas e pelo sucesso.”: Score 0. Foram encontradas no SentiLex as seguintes palavras: “obrigada” (-1), “certa” (1), “oferecidas” (-1) e “sucesso” (1). As palavras “obrigada” e “oferecidas” no SentiLex, existem como adjetivos, e nesse caso teriam polaridade negativa. Porém, no comentário, a palavra “obrigada” é uma interjeição, com polaridade positiva, e a palavra “oferecidas” é verbo, com polaridade neutra.
- “Excelente modo de pesquisa e aviso de vagas e muitas oportunidades.”: Score 0. Foram encontradas no SentiLex as palavras “excelente” (1) e “vagas” (-1). Novamente a palavra “vagas” no comentário é um substantivo, enquanto que no SentiLex a palavra é um adjetivo, com polaridade negativa.

Foi possível notar que o classificador baseado em Dicionário encontrou dificuldades para classificação devido ao fato do modelo implementado não levar em consideração a classe gramatical das palavras. Outra dificuldade encontrada refere-se à palavra “vaga”. No dicionário utilizado, o SentiLex, esta palavra existe como adjetivo e possui sentido negativo. Porém é uma palavra que ocorre com bastante frequência como substantivo no domínio estudado, tendo neste caso sentido neutro. Como vantagem, destaca-se o melhor desempenho em relação ao classificador baseado em Repustate (Tabela 40), e a possibilidade de analisar o funcionamento interno do classificador.

#### 4.5.3 Classificador Baseado em Naive Bayes com os Jurados

A análise dos resultados do classificador baseado em Naive Bayes consistiu em identificar quais palavras foram utilizadas pelo classificador para decidir a classe a ser atribuída. Essas palavras são as que ocorrem no comentário e entre as 850 palavras mais frequentes no conjunto de treinamento.

Não foram encontrados comentários classificados como **neutros** pelo classificador e **positivos** pelos jurados, nem comentários classificados como **neutros** pelo classificador e **negativos** pelos jurados.

Os comentários a seguir foram classificados como **positivo** pelo classificador baseado em Naive Bayes, e como **negativos** pelos jurados. Dos nove comentários encontrados, três são apresentados a seguir:

- “me cobraram XX,XX reais sendo que eu não autorizei essa m....”. As seguintes palavras estão entre as 850 mais frequentes: "me", "essa", "sendo", "eu", "que".
- “Consegui mai entrevistas em sites gráts que na catho.”. As seguintes palavras estão entre as 850 mais frequentes: "em", "na", "entrevistas", "sites", "consegui", "que", "catho".
- “Ruim”. A palavra “ruim” está entre as 850 palavras mais frequentes. No conjunto de treinamento não há nenhuma ocorrência desta palavra entre os comentários positivos. Uma investigação mais aprofundada, que será realizada em trabalhos futuros, é necessária para determinar o motivo desta classificação errônea.

Os comentários a seguir foram classificados como **negativos** pelo classificador baseado em Naive Bayes, e como **neutros** pelos jurados. Os três comentários encontrados são apresentados a seguir:

- “Site que auxilia, mas, deveria apresentar status de vaga...ser urgente...contratação imedita..informar local, pois nao sao todos oscasos que apresenta estas informações...”. As seguintes palavras estão entre as 850 mais frequentes: "ser", "site", "deveria", "nao", "estas", "todos", "...", "mas", "de", "pois", "que", "vaga".
- “Acervo muito bom de vagas! Mas parece que existem anuncios que nao sao levados em consideracao pelas empresas!”. As seguintes palavras estão entre as 850 mais frequentes: "em", "mas", "existem", "de", "nao", "parece", "muito", "que", "bom", "vagas", "empresas", "pelas".
- “Gostei bastante da Empresa, porém não tive ainda nenhum retorno das empresas em que enviei o meu curriculo.”. As seguintes palavras estão entre as 850 mais frequentes: "em", "nenhum", "gostei", "catho", "da", "meu", "retorno", "curriculo", "empresas", "o", "das", "ainda", "que", "tive", "enviei", "bastante".

Os comentários a seguir foram classificados como **positivos** pelo classificador baseado em Naive Bayes, e como **neutros** pelos jurados. Dos sete comentários encontrados, três são apresentados a seguir:

- “Empresa boa e preço salgado. O serviço é otimo pena que seja tão caro.”. As seguintes palavras estão entre as 850 mais frequentes: "empresa", "e", "otimo", "o", "caro", "que", "pena", "boa", "seja".
- “boa, porém não eficiente nas informações sobre as cobranças”. As seguintes palavras estão entre as 850 mais frequentes: "nas", "as", "eficiente", "sobre", "boa".
- “o site é bom mas as empresas não retornam quando vc manda seu cv.”. As seguintes palavras estão entre as 850 mais frequentes: "mas", "seu", "quando", "o", "site", "as", "vc", "empresas", "bom", "cv".

O comentário a seguir foi classificado como **negativo** pelo classificador baseado em Naive Bayes, e como **neutro** pelos jurados. Foi encontrado um comentário segundo esse critério, apresentado a seguir:

- “Foi o site em que obtive contato com as melhores vagas, das empresas maiores.”. As seguintes palavras estão entre as 850 mais frequentes: "em", "das", "site", "o", "contato", "as", "obtive", "que", "melhores", "vagas", "empresas", "com", "foi".

Observa-se pela análise, que entre os comentários que foram classificados incorretamente, muitas das palavras que são consideradas como parâmetros de classificação não são palavras que expressam opinião. Além disso, havia ficado claro pelos resultados apresentados na seção anterior, que o modelo baseado em Naive Bayes teve sérias dificuldades em classificar comentários neutros. Por outro lado, foi o classificador que obteve o melhor desempenho (Tabela 40).

Cada um dos três Modelos de Classificação apresentou vantagens e desvantagens. Por exemplo, enquanto o modelo baseado em Repustate apresentou o pior desempenho de classificação, foi muito mais rápido implementá-lo do que os outros dois Modelos.

A seguir, na Tabela 41, é apresentado um resumo comparativo com as vantagens e desvantagens encontradas nos três modelos de classificação.

Tabela 41 - Vantagens e desvantagens dos três Modelos de Classificação

Modelo	Vantagens	Desvantagens
Repustate	Rápida implementação	Dificuldade em classificar comentários com erros de grafia.
Dicionário	Facilidade em examinar o funcionamento interno do classificador.	Dificuldade em classificar comentários com a palavra “vaga”. Depende do conhecimento da classe gramatical das palavras.
Naive Bayes	Melhor desempenho de classificação.	Dificuldade em classificar comentários neutros.

Fonte: O Autor

#### 4.6 APRESENTAÇÃO DA MINERAÇÃO DE OPINIÕES

A seguir são apresentados os resultados da Mineração de Opiniões obtidos com a classificação realizada pelos classificadores. Os experimentos foram realizados com cada um dos três classificadores, e dois gráficos foram gerados para cada caso.

Conforme definido em Materiais e Métodos na Tabela 3, a entidade refere-se ao serviço prestado pela empresa de classificados *online* de empregos, e o aspecto é o aspecto geral. Os emissores são os clientes da empresa, e são identificados somente pelo seu código.

O primeiro gráfico, apresentado na Figura 9, mostra a distribuição dos comentários agrupados por mês e por polaridade de opinião durante o período de Janeiro a Junho de 2013, utilizando o classificador baseado no Repustate.

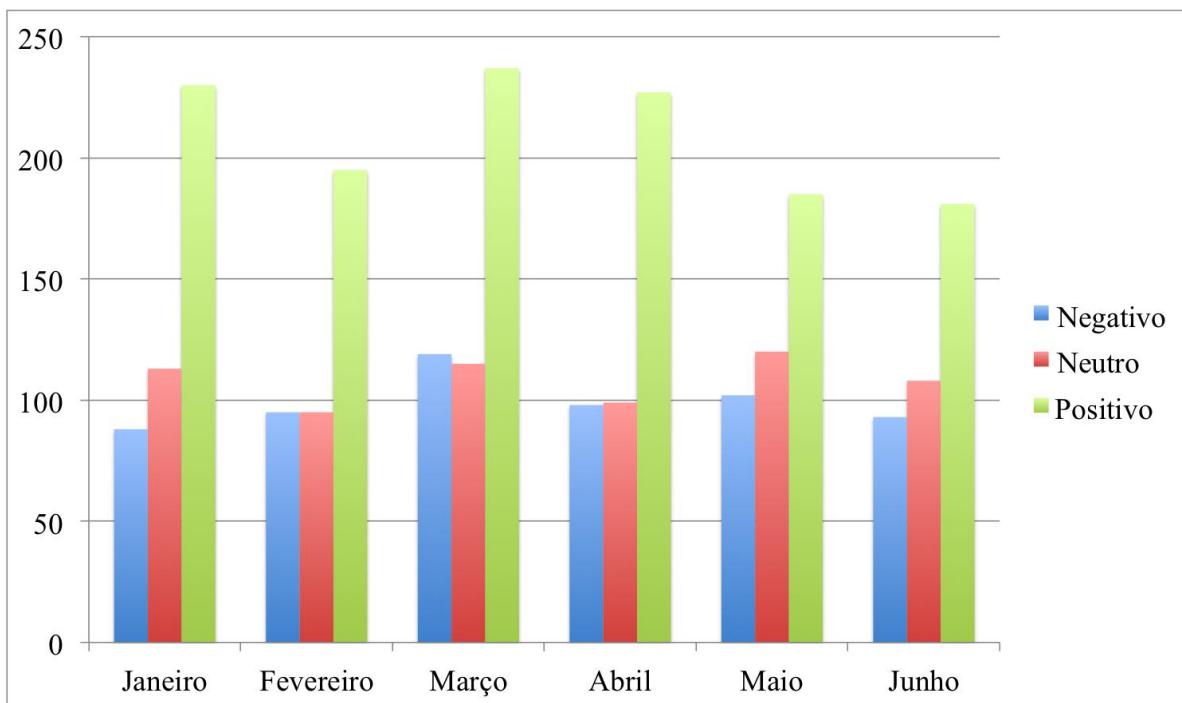


Figura 9 - Distribuição dos comentários com classificador baseado no Repustate. Fonte: O Autor

O gráfico da figura 10 apresenta o percentual de cada uma das polaridades, mês a mês, durante o período de Janeiro a Julho de 2013, utilizando o classificador baseado no Repustate.

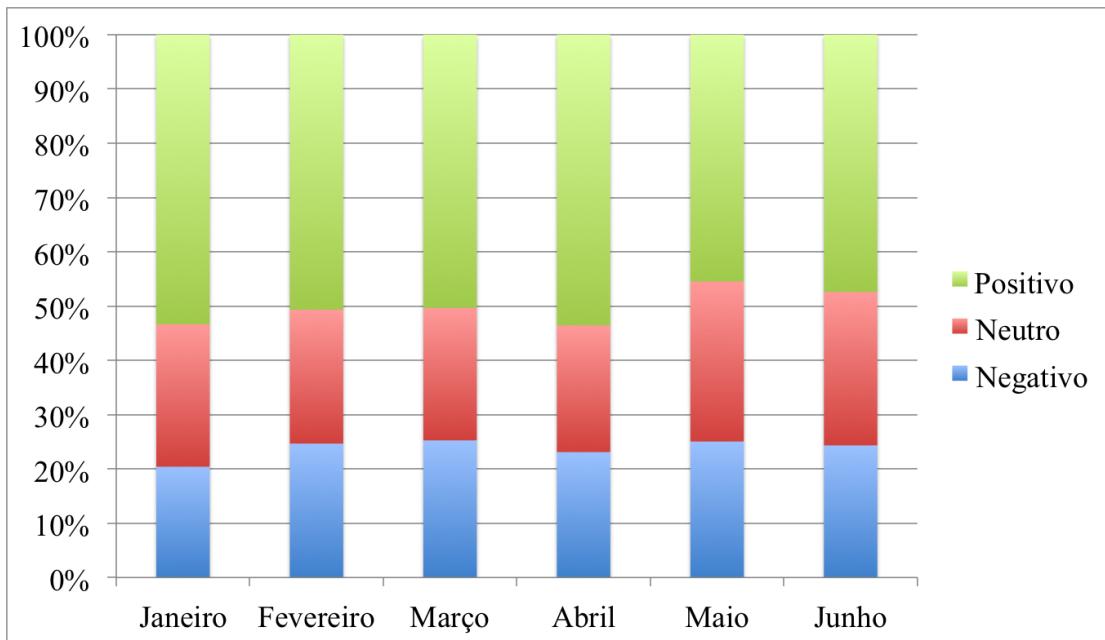


Figura 10 - Distribuição percentual com classificador baseado no Repustate. Fonte: O Autor.

O gráfico apresentado na Figura 11, e mostra a distribuição dos comentários agrupados por mês e por polaridade de opinião durante o período de Janeiro a Junho de 2013, utilizando o classificador baseado em Dicionário.

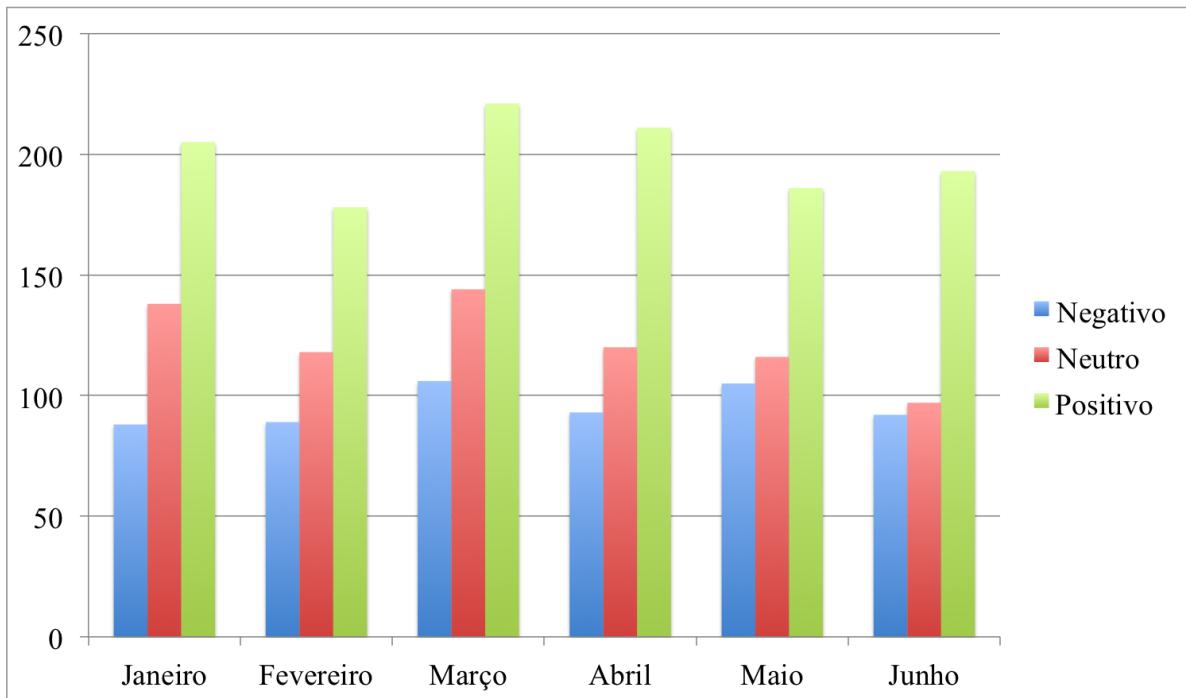


Figura 11- Distribuição dos comentários com classificador baseado em Dicionário. Fonte: O Autor

O gráfico da figura 12 apresenta o percentual de cada uma das polaridades, mês a mês, durante o período de Janeiro a Julho de 2013, utilizando o classificador baseado em Dicionário.

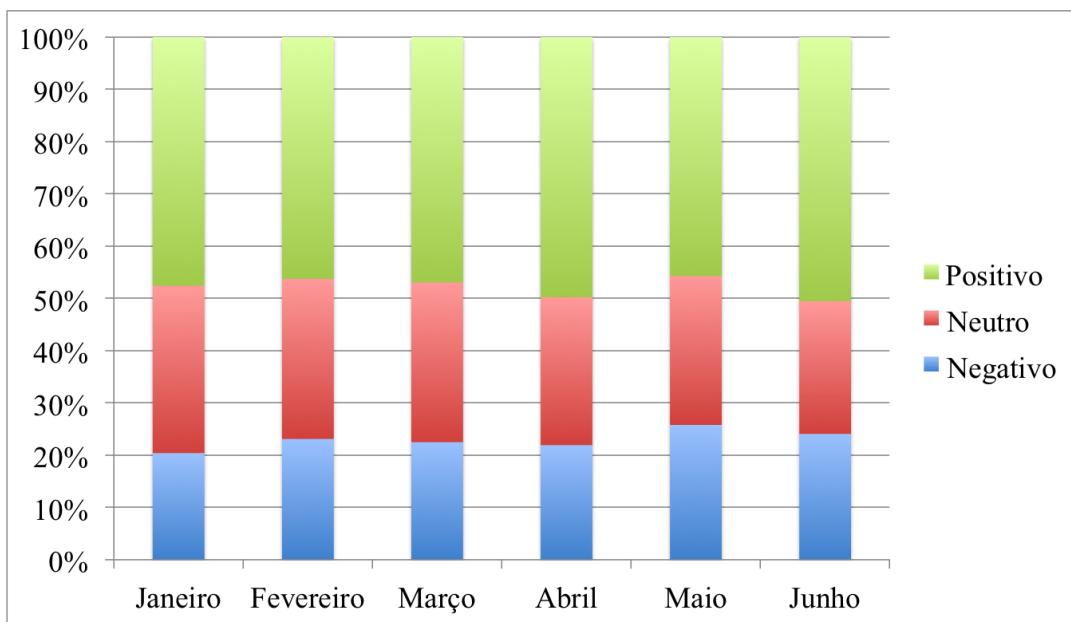


Figura 12 - Distribuição percentual com classificador baseado em Dicionário. Fonte: O Autor.

O gráfico apresentado na Figura 13, e mostra a distribuição dos comentários agrupados por mês e por polaridade de opinião durante o período de Janeiro a Junho de 2013, utilizando o classificador Naive Bayes.

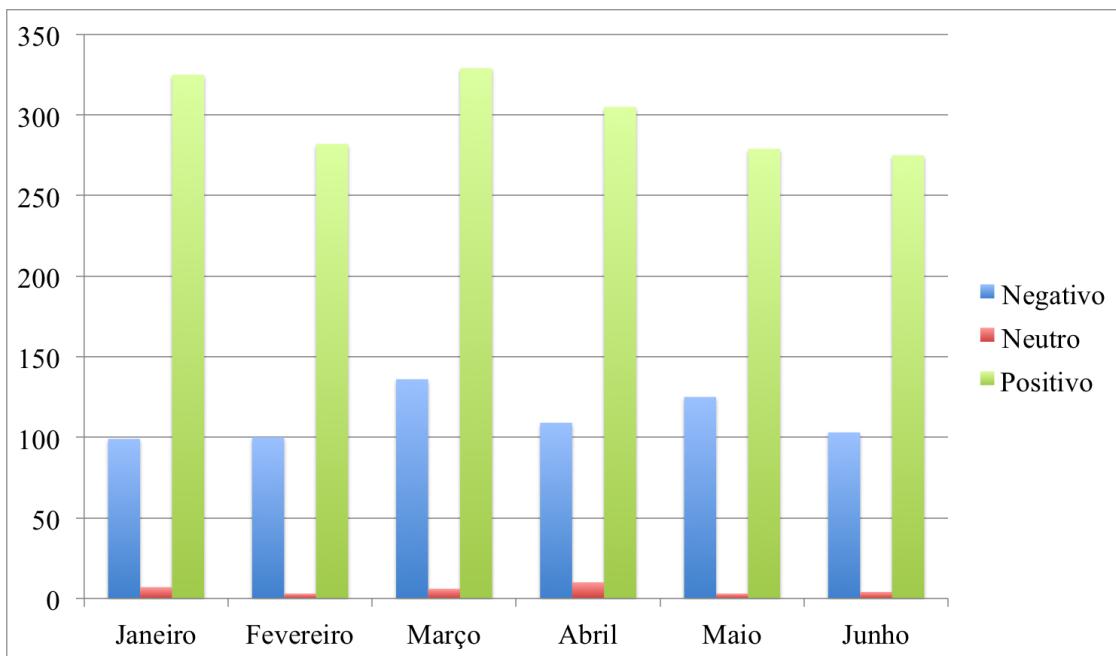


Figura 13 - Distribuição dos comentários com classificador Naive Bayes. Fonte: O Autor

O gráfico da figura 14 apresenta o percentual de cada uma das polaridades, mês a mês, durante o período de Janeiro a Julho de 2013, utilizando o classificador Naive Bayes.

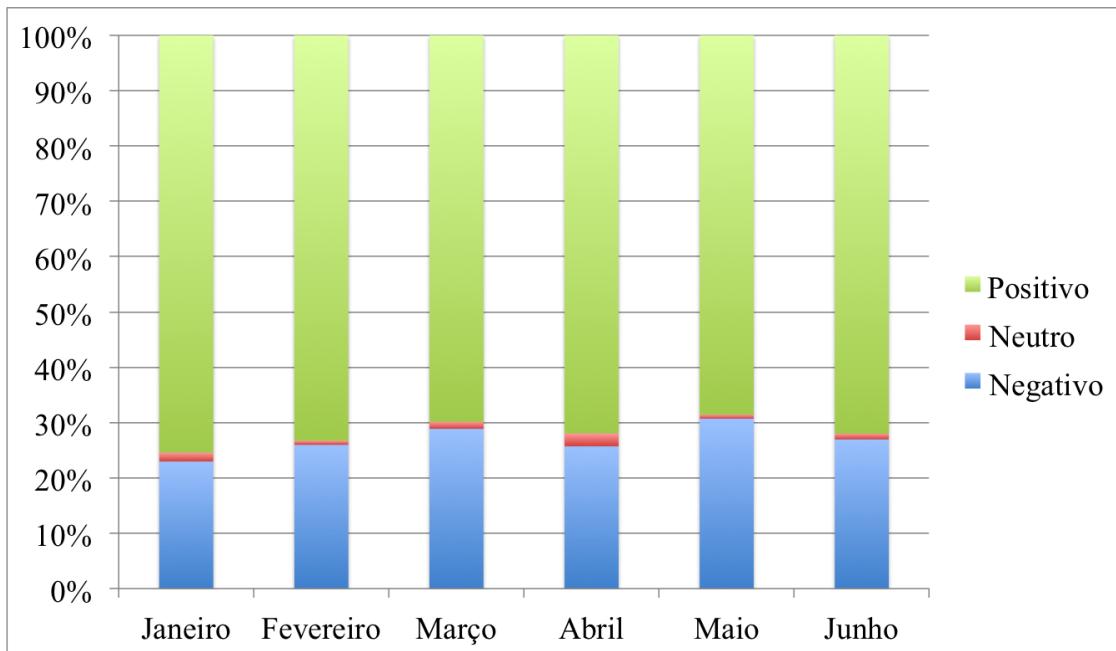


Figura 14 - Distribuição percentual com classificador Naive Bayes. Fonte: O Autor.

Apesar de observarmos variações nos resultados conforme o classificador utilizado, observa-se que o percentual de clientes insatisfeitos com os serviços fica estável entre os três

classificadores, variando entre 20% e 30%. Assim, os Modelos de Classificação dos comentários de clientes comparados neste trabalho podem minerar opiniões em empresa brasileira de classificados online de empregos.

Apresenta-se no próximo capítulo a conclusão do trabalho.

## 5 CONCLUSÃO

Neste trabalho foram comparados três Modelos de Classificação (classificadores) dos comentários de clientes para minerar opiniões em empresa brasileira de classificados *online* de empregos. Foram implementados três Modelos: baseado em *software* comercial Repustate, baseado em Dicionário e baseado em Naive Bayes.

A característica não estruturada dos comentários em língua portuguesa representou uma tarefa complexa para os classificadores, tornando os modelos desenvolvidos nesse trabalho relevantes para as pesquisas em Análise de Sentimentos e Mineração de Opiniões.

O desempenho dos três Modelos de Classificação dos comentários não estruturados de clientes em língua portuguesa foi analisado por meio de comparação com a classificação realizada por jurados (seres humanos). A avaliação do desempenho dos Modelos de Classificação foi feita usando o índice Kappa e a Matriz de Confusão.

Sensível dificuldade dos três Modelos em classificar os comentários foi percebida, fato também observado nos jurados, devido às características dos comentários. Quando se analisou a classificação destes comentários realizada por jurados, constatou-se que a concordância entre eles variou entre 79,33% a 90%. Portanto, ficou evidente que o problema de classificar o sentimento expresso em comentários de clientes não é uma tarefa fácil, nem mesmo para seres humanos.

Segue um resumo sobre a análise:

- Classificador baseado no Repustate: foi realizada a tradução dos comentários para a língua inglesa, pois o Repustate não suporta a língua portuguesa. A partir dos scores gerados pelo software, foi realizada a classificação e o resultado foi comparado aos jurados, apresentando concordância que variou entre fraca e leve, dependendo do jurado. Foi o modelo que apresentou pior dos três resultados obtidos, pois sua taxa de acerto variou entre 52% e 54%. Este resultado serviu como base de comparação entre os modelos.
- Classificador baseado em Dicionário: utilizou-se o SentiLex como dicionário de sentimentos e foi desenvolvido *software* na linguagem Python para classificação baseando-se nos *scores* fornecidos pelo dicionário. Tendo uma taxa de acerto que variou entre 52% e 58% e concordância **leve**, o modelo foi o segundo melhor em desempenho.

- Classificador baseado em Naive Bayes: desenvolveu-se um *software* na linguagem Python, utilizando a biblioteca NLTK para treinar um classificador Naive Bayes. O modelo conseguiu o melhor desempenho entre os três, obteve uma boa taxa de acerto em relação aos jurados, que variou de 61% a 69%, e a concordância ficou entre **leve** e **moderada**, dependendo do jurado.

Os modelos foram aplicados na base de dados contendo os comentários não estruturados em língua portuguesa e, apesar das características dos comentários, a Mineração de Opiniões foi realizada.

Em resposta ao problema de pesquisa pode-se considerar que os Modelos de Classificação dos comentários de clientes comparados neste trabalho podem minerar opiniões em empresa brasileira de classificados online de empregos.

Conclui-se, então, que o objetivo deste trabalho foi alcançado ao comparar Modelos de Classificação dos comentários de clientes para a Mineração de Opiniões em empresa brasileira de classificados *online* de empregos.

Vale destacar que a Mineração de Opiniões em comentários não estruturados em língua portuguesa é uma tarefa complexa, que exige pesquisa e que este cenário está aberto para novos estudos.

Os Modelos de Classificação desenvolvidos neste trabalho foram ajustados para o domínio estudado, ou seja, para comentários de clientes de uma empresa de classificados *online* de empregos. Apesar disso, esses modelos podem servir como base para aplicação em outros domínios, como por exemplo, minerar opiniões de clientes de empresas de outros ramos de atividade ou minerar opiniões de eleitores em período eleitoral. Sendo que a mineração pode ser feita tanto em bases de dados privadas quanto nas redes sociais.

Considera-se a principal contribuição deste trabalho para o campo de estudos da Análise de Sentimentos e Mineração de Opiniões o desenvolvimento dos três modelos para classificação de comentários não estruturados em língua portuguesa.

A análise da concordância dos modelos com os cinco jurados foi também uma contribuição importante para este campo de estudos, já que teve como objetivo comparar o desempenho dos classificadores com seres humanos, ou seja, comparar a classificação realizada no mundo real.

Outro aspecto importante que pode ser considerado uma contribuição se refere a aplicação da Mineração de Opiniões em língua portuguesa, pois apesar de já existirem pesquisas nesse sentido, ainda é uma área que necessita de exploração.

Finalmente, outra contribuição deste trabalho foi a criação de um *corpus* em português marcado em relação ao sentimento. Esse *corpus* foi utilizado para treinamento e testes dos modelos aplicados neste trabalho e será futuramente disponibilizado publicamente.

A principal limitação foi a escassez de referências e modelos sobre Mineração de Opiniões em língua portuguesa. Existem poucos trabalhos e modelos para a Mineração de Opiniões nessa língua nos quais esse trabalho pudesse se basear, o que revela a relevância e o caráter inovador da pesquisa.

Outras limitações referem-se a implementação e acuidade dos Modelos de Classificação. Os modelos de classificação não possuem a capacidade de identificar a classe gramatical das palavras, e como foi percebido durante a análise dos resultados, a mesma palavra, dependendo de sua classe gramatical pode apresentar polaridades de sentimento totalmente diferentes.

Verificou-se ainda outra limitação que diz respeito ao próprio dicionário utilizado, o SentiLex. Nele, uma palavra que é utilizada comumente no domínio estudado e ocorre como substantivo com polaridade neutra existe no dicionário somente como adjetivo com polaridade negativa. Este fato impactou negativamente os resultados do classificador baseado em Dicionário.

Para a continuidade da pesquisa planeja-se melhorar os modelos de classificação. Para isso, pretende-se implementar nos modelos baseados em Dicionário e em Naive Bayes a capacidade de identificar a classe gramatical das palavras levando em consideração este parâmetro na classificação.

A fim de tentar contornar a dificuldade do modelo baseado em Naive Bayes em classificar comentários neutros, propõe-se a implementação de um Modelo de Classificação de comentários híbrido, combinando-se os classificadores baseados em Dicionário e em Naive Bayes. Finalmente, propõe-se também explorar outras técnicas de Aprendizado de Máquina para a classificação dos comentários, entre elas, a utilização de *Support Vector Machines* (SVM).

Os estudos aqui realizados não têm a pretensão de esgotar o assunto, pelo contrário, buscou-se realizar uma contribuição com o desenvolvimento e aplicação dos Modelos de Classificação dos comentários de clientes.

Sabe-se que existe uma clara demanda por estudos sistematizados que possam estabelecer outros domínios de aplicação ainda mais adequados para os modelos. Este cenário oferece portanto amplo espaço para trabalhos de continuidade.

## 5.1 CO-ORIENTAÇÃO EM PROJETO DE INICIAÇÃO CIENTÍFICA

Foi realizada co-orientação em um projeto de Iniciação Científica, intitulado: “*Opinion Mining aplicado a E-mail Marketing: Criando campanhas baseadas na opinião dos consumidores de uma empresa de classificados on-line*”. Iniciado em 2012 e finalizado em 2013, o projeto gerou os seguintes resumos publicados:

*Opinion Mining como Apoio na Elaboração de Campanhas de E-Mail Marketing.* In: XX Simpósio Internacional de Iniciação Científica - USP, 2012, São Paulo. XX Simpósio Internacional de Iniciação Científica - USP, 2012.

Utilizando *Opinion Mining* para Segmentar Campanhas De *E-Mail Marketing*. In: XIV Simpósio de iniciação científica e tecnológica, 2012, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo. São Paulo: Boletim Técnico da FATEC-SP, 2012. v. BT/34. p. 149-149.

Opinion Mining no Apoio à Avaliação da Satisfação do Cliente. In: 15º Simpósio de Iniciação Científica e Tecnológica da Fatec (SICT-2013), 2013, São Paulo. Boletim Técnico da Faculdade de Tecnologia de São Paulo. São Paulo: FATEC, 2013. v. 1. p. 147-147.

## 5.2 PUBLICAÇÕES DO AUTOR

### - Periódico Internacional

MIRANDA, M. D. ; SASSI, R. J. . Using Sentiment Analysis to Assess Customer Satisfaction in a Online Job Search Company. In: 17th International Conference on Business Information Systems (BIS2014): 6th Workshop on Applications of Knowledge-Based Technologies in Business (AKTB 2014), 2014, Chipre. Lecture Notes in Business Information Processing, 2014. v. 183.

- Periódico Internacional

MIRANDA, M. D. ; Sassi, R. J. Using Sentiment Analysis to Aid Customer Satisfaction Assessment in a Brazilian Online Job Search Company. European Journal of Scientific Research, v. 119, p. 234-243, 2014.

- Trabalhos completos publicados em anais de congressos:

MIRANDA, M. D. ; Sassi, R. J. . Opinion Mining Aplicado à Avaliação Da Satisfação Do Cliente em Uma Empresa de Classificados Online de Empregos. In: XX SIMPEP - Simpósio Brasileiro de Engenharia de Produção, 2013, Bauru. XX SIMPEP - Engenharia de Produção & Objetivos de Desenvolvimento do Milênio, 2013.

MIRANDA, M. D. ; Farias, E. B. P. ; Sassi, R. J. . Satisfação do cliente em uma empresa de classificados online de empregos: *Sentiment Analysis* como ferramenta de avaliação. In: III Congresso Brasileiro de Engenharia de Produção (CONBREPRO 2013), 2013, Ponta Grossa. CONBREPRO - 2013, 2013.

MIRANDA, M. D. ; Romero, M. ; Sassi, R. J. . Eficácia em *E-mail Marketing*: Usando um *Data Mart* em uma Empresa de Classificados on-line de Empregos. In: 9o. Congresso Internacional de Gestão da Tecnologia e Sistemas de Informação - CONTECSI, 2012, São Paulo. *Proceedings to 9th CONTECSI International Conference on Information Systems and Technology Management*. São Paulo: TECSI EAC FEA USP, 2012. v. 1. p. 1-18.

MIRANDA, M. D. ; Sassi, R. J. . Criando Campanhas de E-mail Marketing com o Apoio do Opinion Mining. In: IX Simpósio de Excelência Em Gestão e Tecnologia, 2012, Resende-RJ. IX Simpósio de Excelência em Gestão e Tecnologia - Tema: Gestão, Inovação e Tecnologia para Sustentabilidade, 2012.

MIRANDA, M. D. ; Sassi, R. J. . *Opinion Mining* Aplicado ao *Email Marketing*: Criando Campanhas Baseadas na Opinião dos Consumidores de uma Empresa de Classificados Online. In: XXXII Encontro Nacional De Engenharia De Produção, 2012, Bento Gonçalves. XXXII Encontro Nacional De Engenharia De Produção Desenvolvimento Sustentável e Responsabilidade Social: As Contribuições da Engenharia de Produção, 2012.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BACCIANELLA, S., ESULI, A., & SEBASTIANI, F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. **In LREC**, v. 10, p. 2200-2204, 2010.
- BANERJEE, M., CAPOZZOLI, M., MCSWEENEY, L., SINHA, D. Beyond kappa: A review of interrater agreement measures. **Canadian Journal of Statistics**, n. 27, v.1, p. 3-23, 1999.
- BERNARDES, T. Caracterização do ambiente agrícola do Complexo Serra Negra por meio de Sensoriamento Remoto e Sistemas de Informação Geográfica. 2006. 119 p. Dissertação (Mestrado em Ciência do Solo) - Universidade Federal de Lavras, Lavras.2006.
- BERRY, L. L. Services Marketing is Different, **Bussines**, n. 3, v. 30, p. 24-28, 1980.
- BATTAGLIA, D.; BORCHARDT, M.; SELLITTO, M.; PEREIRA, G. Relative importance analysis and the performance of service recovery dimensions. **International Conference on Industrial Engineering and Operations Management- ICIEOM**, 2010, São Carlos. ICIEOM, 2010.
- BIRD, S.;LOPER, E.; KLEIN, E. **Natural Language Processing with Python**. O'Reilly Media Inc., 2009.
- BOOMS, B.H.; BITNER, M.J., Marketing Strategies and Organization Structures for Services Firms, **Marketing of Services**, Chicago, *American Marketing*, p. 47-51, 1981.
- BROOKE, J., TOFIOSKI, M., TABOADA, M. Cross-linguistic sentiment analysis: From English to Spanish. **In International Conference RANLP**, p. 50-54, 2009.
- CAMPBELL, J. B. **Introduction to remote sensing**. New York: The Guilford Press. 2002. 621 p.
- CAPELLI, P, Making the Most of On-Line Recruiting, **Harvard Business Review**, v. 79, n. 3, p. 139-146, 2001.
- CHAVES, M. S., DE FREITAS, L. A., SOUZA, M., & VIEIRA, R. Pirpo: An algorithm to deal with polarity in portuguese online reviews from the accommodation sector. **In Natural Language Processing and Information Systems**, p. 296-301. Springer Berlin Heidelberg, 2012.
- CHEN, H.; ZIMBRA D. AI and Opinion Mining. **Intelligent Systems, IEEE**, n. 3, v. 25, p. 74-80, 2010.
- ESULI, A., SEBASTIANI, F. Determining Term Subjectivity and Term Orientation for Opinion Mining, **In Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)**, v. 6, 2006.

- FLEISS, J. L., LEVIN, B., &PAIK, M. C. The Measurement of Interrater Agreement. **Statistical Methods for Rates and Proportions**, Third Edition, p. 598-626, Willey, 2003.
- GAMALLO, P., &GARCIA, M. Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets. **Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)**, Dublin, Irlanda, p. 171-175, 2014.
- GAMALLO, P.;GARCIA, M.; FERNANDEZ-LANZA, S. TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets. **In Workshop on Sentiment Analysis at SEPLN (TASS2013)**, pages 126–132, Madrid, Espanha, 2013.
- GAMON, M., Aue, A., Corston-Oliver, S., Ringger, E., Pulse: Mining customer opinions from freetext. **Advances in Intelligent Data Analysis VI**, p. 121-132, Springer Berlin Heidelberg, 2005.
- GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 4<sup>a</sup> ed.. São Paulo: Atlas, 2002.
- GOOGLE. Google Translate. Disponível em: <http://translate.google.com>. 2014a
- \_\_\_\_\_. Google Translate API. Disponível em: <https://developers.google.com/translate>. 2014b.
- JINDAL, N. LIU, B. Mining comparative sentences and relations, **American Association for Artificial Intelligence (AAAI)**, v. 22, p. 1331-1336, 2006.
- LANDIS, J., KOCH, G.: The measurement of observer agreement for categorical data. **Biometrics** 33, 159–174. 1977.
- LIU B.; HU, M. Mining opinion features in customer reviews. **In Proceedings of Nineteenth National Conference on Artificial Intelligence (AAAI-2004)**, pages 755-760, 2004.
- LIU, B. **Web Data Mining**, 2a. Ed., Springer-Verlag Berlin Heidelberg, 2011.
- \_\_\_\_\_. Sentiment Analysis and Opinion Mining, **Synthesis Lectures on Human Language Technologies**, n. 16, Morgan & Claypool, 2012.
- MIAO, Q., LI, Q., DAI, R. AMAZING: A sentiment mining and retrieval system. **Expert Systems with Applications**. v. 36, n. 3, p. 7192-7198, 2009.
- MILLER, G.A. WordNet: A Lexical Database for English. **Communications of the ACM**, v. 38, n. 11: 39-41, 1995.
- MIRANDA, M. D. ; SASSI, R. J. . Using Sentiment Analysis to Aid Customer Satisfaction Assessment in an Brazilian Online Job Search Company. **European Journal of Scientific Research**, v. 119, p. 234-243, 2014.

PANG, B.; LEE, L., Opinion Mining and Sentiment Analysis, **Foundations and Trends in Information Retrieval**, v. 2, nos. 1-2, p. 1-135, 2008.

PANG, B., LEE, L., VAITHYANATHAN, S. Thumbsup?: sentiment classification using machine learning techniques. In **Proceedings of the ACL-02 conference on Empirical methods in natural language processing**. v. 10, p. 79-86. Association for Computational Linguistics, 2002.

PARASSURAMAN, A.; ZEITHAML, V.A., BERRY, L.L. A Conceptual Model of Service Quality and its Implications for Future Research, **Journal of Marketing**, v. 49, p. 41-50, 1985.

PARASSURAMAN, A., ZEITHAML, V. A., MALHOTRA, A. ES-QUAL a multiple-item scale for assessing electronic service quality. **Journal of service research**, v. 7, n. 3, p. 213-233, 2005.

PARRY, E., TYSON, S., Ananalysis of the use and success of online recruitment in the UK. **Human Resource Management Journal**, v 18, n. 3, p. 257-274, 2008.

PINHO, J.A.G. Sociedade da informação, capitalismo e sociedade civil: reflexões sobre política, internet e democracia na realidade Brasileira. **Rev. adm. empres.**, São Paulo, v. 51, n. 1, fev. 2011.

PYTHON SOFTWARE FOUNDATION. *Python Programming Language*. Disponível em: <<http://www.python.org>>. Acesso em 01 mai. 2014.

REPUSTATE. Repustate: Sentiment analysis and social media analytics. Disponível em:<<https://www.repustate.com>>. Acesso em 01 mai. 2014a

\_\_\_\_\_. API documentation. Disponível em: <<https://www.repustate.com/docs/#api-2>>. Acesso em 01 mai. 2013b.

SAMOHYL, R. W. **Controle Estatístico de Qualidade**. Rio de Janeiro, Elsevier, p. 72, 2009.

SANTOS, J. E-service quality: a model of virtual service quality dimensions. **Managing service quality**, 13(3), 233-246, 2003.

SELENIUM. SeleniumHQ Browser Automation; Disponível em: <<http://www.seleniumhq.org>>, Acesso 01 mai. 2014.

SIEGEL S, CASTELLAN N. **Nonparametric Statistics for the Behavioral Sciences**, 2.ed. New York: McGraw-Hill, p 284-285. 1988.

SILVA, M.; CARVALHO,P.; SARMENTO, L. "Building a Sentiment Lexicon for Social Judgement Mining". In **Lecture Notes in Computer Science (LNCS)**, International Conference on Computational Processing of the Portuguese Language (PROPOR), Springer, p. 218-228, 2012.

- TONTINI, G.; SANT'ANA A., Interação de atributos atrativos e obrigatórios de um serviço na satisfação do cliente, **Produção**, v. 18, n. 1, p. 112-125, 2008.
- TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, v. 24, n. 3, p. 478-514, 2012.
- UPAH, G.D. Mass Marketing in Service Retailing: A Review and Synthesis of Major Methods, **Journal of Retailing**, 56, 59-76, 1980.
- VALARMATHI, B.; PALANISAMY, V. Opinion Mining Classification Using Key Word Summarization Based on Singular Value Decomposition, **International Journal on Computer Science and Engineering (IJCSE)**, n. 1, v. 3, p. 212-215, 2011.
- VAN RIEL, A. C., LILJANDER, V., & JURRIENS, P. Exploring consumer evaluations of e-services: a portal site. **International Journal of Service Industry Management**, 12(4), 359-377, 2001.
- YIN, R. K. **Estudo de Caso**: planejamento e métodos. Trad. Daniel Grassi e Cláudio Damacena. 2.ed. Porto Alegre: Bookman, 2006. p. 205.
- WAN, X. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. **In Proceedings of the Conference on Empirical Methods in Natural Language Processing**, p. 553-561. Association for Computational Linguistics, 2008.
- WAN, X. Co-training for cross-lingual sentiment classification. **In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP**, vol. 1, p. 235-243. Association for Computational Linguistics, 2009.
- WANG, W.; ZHOU, Y. E-Business Websites Evaluation Based on Opinion Mining, **International Conference on Electronic Commerce and Business Intelligence**, p. 87-90, 2009.
- ZEITHAML, V.A.; PARASURAMAN, A.; MALHOTRA, A. Service Quality Delivery Through Web Sites: A Critical Review of Extant Knowledge, **Journal of Academy of Marketing Science**, v. 30, n. 4, p. 362-375, 2002.

## APÊNDICE A: Código Python para Tradução dos Comentários

```

1. # _*_ coding: utf-8 _*
2.
3. from selenium import webdriver
4. import time
5. import mysql.connector
6.
7. browser = webdriver.Firefox()
8.
9. browser.get('http://translate.google.com/#pt/en/')
10. source = browser.find_element_by_id('source')
11. result_box = browser.find_element_by_id('result_box')
12. gt_clear = browser.find_element_by_id('gt-clear')
13. gt_submit = browser.find_element_by_id('gt-submit')
14.
15. conn1 = mysql.connector.connect(user='root',
16.                                     host='127.0.0.1',
17.                                     database='analise')
18. conn2 = mysql.connector.connect(user='root',
19.                                     host='127.0.0.1',
20.                                     database='analise')
21. cursor1 = conn1.cursor()
22. cursor2 = conn2.cursor()
23.
24. select = """SELECT id_comentario, comentario_usuario
25.             FROM analise.comentarios_traduzidos
26.             WHERE comentario_traduzido IS NULL"""
27. cursor1.execute(select)
28.
29. for idc, comentario in cursor1:
30.     source.send_keys(comentario)
31.     gt_submit.click()
32.     time.sleep(1.5)
33.     comentario_traduzido = result_box.text
34.     if comentario_traduzido == '':
35.         print 'Erro ao traduzir. Tentando novamente.'
36.         time.sleep(5.0)
37.         comentario_traduzido = result_box.text
38.     if comentario_traduzido == '':
39.         print 'Erro ao traduzir. Desistindo...'
40.     else:
41.         update = """UPDATE analise.comentarios_traduzidos
42.                     SET comentario_traduzido = %s
43.                     WHERE id_comentario = %s"""
44.
45.         cursor2.execute(update, (comentario_traduzido, idc,))
46.         cursor2.execute('COMMIT')
47.     gt_clear.click()
48.
49.
50. browser.quit()

```

## APÊNDICE B – Código Python para Geração dos Scores de Comentário com o Repustate

```

1. #_*_ coding: utf-8 _*_
2.
3. import mysql.connector
4. from selenium import webdriver
5. import requests
6. import time
7. import json
8.
9. # Necessario inicializar o browser antes
10. browser = webdriver.Firefox()
11. browser.get('https://www.repustate.com/api-demo/')
12. browser.find_element_by_xpath('//*[@id="id_api_call"]/option[2]').click()
13.
14.
15. class Score(object):
16.     def __init__(self, id_comentario, score):
17.         self.id = id_comentario
18.         self.score = score
19.
20.     def setId(self, id_comentario):
21.         self.id = id_comentario
22.
23.     def setScore(self, score):
24.         self.score = score
25.
26.     def getId(self):
27.         return self.id
28.
29.     def getScore(self):
30.         return self.score
31.
32.
33. def getScoreFromRepustate(comentario):
34.     input = browser.find_element_by_id('id_text')
35.     input.clear()
36.     time.sleep(0.2)
37.     input.send_keys(comentario)
38.     time.sleep(0.5)
39.     submit_btn = browser.find_element_by_id('submit')
40.     submit_btn.click()
41.     output = browser.find_element_by_xpath('//*[@id="wrap"]/div[2]/div/div[1]/form/div[2]/pre')
42.     retorno = json.loads(output.text)
43.     if retorno['status'] == 'OK':
44.         return retorno['score']
45.     else:
46.         raise Exception('Retorno do JSON nao foi OK')
47.
48.
49. def updateNewScore(score):
50.     conn = mysql.connector.connect(host='127.0.0.1', user='root', db='analise')
51.     cursor = conn.cursor()
52.     updateSql = """UPDATE scores_selenium
53.                 SET score = %s
54.                 WHERE id_comentario = %s"""
55.     cursor.execute(updateSql, (score.getScore(), score.getId(),))
56.     cursor.execute('COMMIT')
57.     conn.close()
58.
59. #####
60.

```

```
61. if __name__ == '__main__':
62.
63.     print "Iniciando processamento dos comentários"
64.     connRead = mysql.connector.connect(host='127.0.0.1',
65.                                         user='root',
66.                                         db='analise'
67.                                         )
68.
69.     cursorRead = connRead.cursor()
70.
71.     select = """SELECT id_comentario, comentario_traduzido
72.                 FROM scores_selenium
73.                 WHERE score is null"""
74.
75.     cursorRead.execute(select)
76.
77.     for idc, comment in cursorRead:
78.         try:
79.             score_value = getScoreFromRepustate(comment)
80.         except Exception, e:
81.             print 'Erro: ', e
82.
83.             score = Score(idc, score_value)
84.             updateNewScore(score)
85.             time.sleep(0.3)
```

## APÊNDICE C – Instruções aos Jurados para Classificação dos Comentários

Caro voluntário,

Preciso de vossa ajuda para validar pesquisa científica em andamento.

Estou desenvolvendo uma ferramenta capaz de processar um texto e classifica-la como expressando uma opinião positiva, negativa, ou neutra. Para validar a eficácia da ferramenta, preciso compará-la com a classificação feita por um ser humano.

Por isso preciso de vossa ajuda. Estou enviando junto com este e-mail, uma planilha contendo 150 comentários feitos por clientes da empresa onde trabalho. Os comentários foram escritos pelos clientes quanto estavam cancelando o serviço.

Nesta planilha, existem 3 colunas: id\_comentário, comentario e "Polaridade de opinião ou sentimento".

A primeira coluna, id\_comentario, não deve ser alterada, e serve apenas para controle.

A coluna comentario, contem o comentário a ser lido e analisado.

A coluna "Polaridade de opinião ou sentimento" é a que vocês devem preencher. Por favor, utilizem os seguintes valores (e somente esses):

1 - Caso o comentário seja positivo

2 - Caso o comentário seja negativo

3 - Caso o comentário seja neutro

Comentários neutros são os que não expressam opinião, ou que as opiniões positivas cancelam as negativas. Por exemplo:

"O serviço é ótimo" - É positivo. A coluna deve ser preenchida com o valor 1

"O serviço é bom, mas o atendimento foi péssimo". Uma opinião positiva e uma negativa. Resultado: neutro. A coluna deve ser preenchida com o valor 3

"O serviço é bom, o atendimento é bom, mas o site é feio". Dois positivos e um negativo. Resultado: positivo. Coluna preenchida com o valor 1.

"O site é ruim". Negativo. Coluna preenchida com o valor 2.

Desde já agradeço imensamente a contribuição com meu trabalho.

## APÊNDICE D – Exemplos de Comentários dos Clientes da Empresa

“Ótimos serviços, porém acho que poderia ser um pouco mais exclusivo, ter uma análise real do currículo com sugestões sobre vagas e perfil.”(**classificado como positivo por um jurado, e como neutro por outro**)

“agente precisa de informações se as empresas pelo menos olhou o currículo.”(**classificado como negativo por um jurado, e como neutro por outro**)

“Empresa séria, comprometida e interessada no bom atendimento de seus clientes.”(**Classificado como positivo por dois jurados**)

“Um jeito muito prático e dinâmico de se procurar um emprego, infelizmente me dispus por pouco tempo dos seus serviços, talvez por isso tenha sido pouco efetivo, mas gostei bastante da metodologia.”(**Classificado como positivo por um jurado, e como neutro por outro**)

“um bom site, não para o meio da veterinária, poucos anúncios, baixas remunerações.”  
(**Classificado como negativo por um jurado, e como neutro por outro**)

“otimotrabalhho mais seria bom se fosse mais viável para os clientes”  
(**Classificado como positivo por um jurado, e como neutro por outro**)

“Não existe minha área de atuação: Engenharia Biomédica que é o mais reconhecido no mundo.”(**Classificado como negativo por um jurado, e como neutro por outro**)

“Preenchimento do CV não permitiu efetuar o mesmo. Em épocas anteriores quando paguei e muito pelos serviços da Empresa fui atendido em minhas pretensões” (**Classificado como positivo por um jurado, e como negativo por outro**)

“Ótimo para trabalhos voltados a áreas mais corporativas.”(**Classificado como positivo por dois jurados**)

“não curti pois nem sabia que seria debitado da minha conta, pois dizi que era grátis”

“é muito ruim só gastei dinheiro porque arrumei um emprego mas não pela empresa eles não me ligaram nenhuma vez nem pra marcar entrevista”

“MUITO BEM INDICADA, MAS NÃO TIVE EXITO PELA MESMA.”

“me informaram que eu teria mais 7 dias gratuitos e mandaram falando que no dia 28 já viria o débito não estou com condições no momento!”

“muito bom fácil da gente se encontrar nas vagas ainda mas quem não sabe muito de computação foi fácil muito legal”

“A EMPRESA TEM TUDO PARA SER UMA EMPRESA BOA, MAIS FALTA AS VAGAS SEREM ENCAMINHADAS PARA O EMAIL.”