

# Análise Comparativa de Abordagens de Análise de Sentimento Utilizando *Tweets* em Língua Portuguesa

Caio H. P. de Souza, Wladimir C. Brandão<sup>1</sup>

<sup>1</sup>Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte – Minas Gerais – Brasil

cpsouza2013@gmail.com, wladimir@pucminas.br

**Abstract.** *Social networks are responsible for a massive amount of data, nowadays highly valued by organizations for its potential to generate significant relevant information for business. Thus, extract relevant information from data circulating on social networks and analyze those information to generate value becomes crucial for organizational survival. However, there are numerous approaches to extraction and analysis of data from social networks, especially for sentiment analysis. Thus, it becomes important for organizations to know the potential of application and limitations of these approaches to their business. This paper presents a comparative analysis of three different approaches of sentiment analysis widely used by organizations in general. Significant gains were achieved for each context studied on this experiment, highlighting the most relevant approach.*

**Resumo.** *Redes sociais movimentam volumes massivos de dados, hoje muito valorizados por organizações por terem o potencial de gerar informação de significativa relevância para os negócios. Dessa forma, extrair informação relevante a partir de dados que circulam nas redes sociais e analisar tal informação para geração de valor se torna crucial para a sobrevivência organizacional. No entanto, existem inúmeras abordagens para extração e análise de dados provenientes de redes sociais, em especial para análise de sentimentos. Dessa forma, se torna fundamental para as organizações conhecerem os potenciais de aplicação e as limitações dessas abordagens para seus negócios. O presente trabalho apresenta uma análise comparativa entre três diferentes abordagens para análise de sentimento amplamente utilizadas pelas organizações em geral. Foram alcançados ganhos significativos para cada contexto trabalhado nestes experimentos, ressaltando-se a abordagem com maior relevância.*

## 1. Introdução

Os avanços recentes no campo das tecnologias de informação e comunicação (TICs) possibilitaram que as pessoas expressassem suas opiniões por meio das redes sociais. Consumidores passaram a usar intensamente as redes sociais para declarar suas frustrações e satisfações sobre suas experiências de consumo. Consequentemente, as redes sociais se consolidaram como fonte primária de dados sobre experiências pessoais extremamente úteis para que organizações possam descobrir conhecimento relevante sobre seu mercado consumidor, a fim de melhorar seu desempenho nos negócios. No entanto, diante do volume massivo de dados presentes nas redes sociais, a escolha de abordagens efetivas

para mineração de opiniões e análise de sentimentos de consumidores se torna um grande desafio, mas fundamental para a melhoria do desempenho organizacional.

O presente artigo apresenta uma comparação entre três abordagens para análise de sentimentos utilizadas por organizações, visando estabelecer um paralelo entre as abordagens em diferentes cenários de uso. Para tanto, foram coletadas mensagens de diferentes contextos do *Twitter*<sup>1</sup>, tais como mensagens sobre organizações, locais e pessoas. Cada uma das abordagens foi utilizada para classificar as mensagens em positivas, expressando sentimentos positivos relacionados à satisfação, negativas, expressando sentimentos negativos relacionados à frustração, e neutras, não expressando sentimento algum. Em seguida, métricas presentes na literatura foram utilizadas para comparação da efetividade das abordagens [Sokolova and Lapalme 2009].

A análise comparativa possibilitou demonstrar que cada abordagem se comporta de maneira distinta nos diferentes contextos propostos no experimento, identificando que determinada abordagem alcançou maiores ganhos para determinado contexto. Visou-se oferecer insumos para organizações que desejam obter conhecimento sobre como seu mercado consumidor reage ao seu produto, proporcionando às organizações a tomada de decisão sobre qual abordagem utilizar. Foram identificadas as vantagens, desvantagens e limitações das abordagens disponíveis, possibilitando maiores ganhos de acordo com o contexto ao iniciar um processo de análise de sentimento dos dados referentes a sua companhia.

O conteúdo deste artigo está estruturado da seguinte maneira: a Seção 2 apresenta o referencial teórico, incluindo uma breve descrição das técnicas de inteligência de negócio e de análise de conteúdo. A Seção 3 descreve os trabalhos relacionados, em especial aqueles referenciados na literatura, que propõem abordagens para descoberta de conhecimento em textos. A Seção 4, apresenta o fluxo de operação e os principais componentes das abordagens propostas. A Seção 5 descreve os resultados oriundos da aplicação da abordagem proposta no contexto da PUC Minas. Finalmente, a Seção 6 apresenta a conclusão, bem como direções para trabalhos futuros.

## **2. Referencial Teórico**

Nesta seção são apresentados conceitos e técnicas referenciados na literatura técnico-científica, fundamentais para o entendimento do trabalho, incluindo coleta de dados na Web, redes sociais e análise de sentimentos.

### **2.1. Redes Sociais**

Segundo [Wasserman and Faust 1994], por definição, uma rede social é um conjunto de atores que pode possuir relacionamentos uns com os outros. São exemplos de redes sociais os grupos familiares, grupos de amigos de faculdade, de academia, de trabalho ou até mesmo encontros casuais, imprevistos. Eles podem ser vistos e caracterizados como a criação de um relacionamento entre dois indivíduos (nós), ligando assim as redes já existentes de ambos. Tal relacionamento pode nunca mais ser nutrido ou, como em alguns casos, vir a se tornar algo mais forte do que todos os relacionamentos já existentes.

Rede Social online é um tipo de mídia online que permite que usuários ao redor do mundo se conectem, troquem experiências e compartilhem conteúdo de forma instantânea

---

<sup>1</sup>Twitter é uma rede social que permite aos usuários enviar e ler mensagens curtas denominadas *tweets*.

por meio da Internet. As redes sociais são fruto do processo de socialização da informação nos últimos anos representado pela extensão do diálogo e do modo como as informações passaram a ser organizadas através da Web [França et al. 2014].

## **2.2. Coleta de Dados de Redes Sociais**

Atualmente, as principais redes sociais online (RSO) provêem interfaces ou serviços para a captura parcial ou total de seus dados. Nesta seção, serão expostos os principais desafios e recursos para se trabalhar com as principais RSO existentes atualmente.

Existem formas diferentes de coleta de dados das redes sociais online. A primeira forma consiste em determinar termos e coletar por citações destes termos no passado. Desta forma, existe a possibilidade de restrições na obtenção de dados antigos, pois normalmente há um período de tempo viável para a coleta dos dados. A segunda baseia-se em um conceito de *streaming*, onde a aplicação criada funciona como um “ouvinte” da rede e captura os dados à medida que estes surgem.

Algumas redes sociais disponibilizam uma *API (Application Programming Interface)* para que estudos ou coletas de dados sejam realizadas de maneira simplificada, o que garante que os dados que ali circulam e são coletados estão vinculados a uma conta de desenvolvedor e sob um conjunto de termos de responsabilidade da rede social.

## **2.3. Análise de Sentimentos**

Segundo [Timmermann et al. 2013], a Análise de Sentimentos visa identificar o sentimento que os usuários apresentam a respeito de alguma entidade de interesse (um produto específico, uma empresa, um lugar, uma pessoa, dentre outros) baseado no conteúdo disponível na Web. O objetivo principal é permitir que um usuário obtenha um relatório contendo o que as pessoas andam dizendo sobre algum item selecionado sem precisar encontrar e ler todas as opiniões e notícias a respeito. Cada vez mais pessoas e principalmente empresas estão interessadas em observar as opiniões de um grupo de pessoas sobre temas que lhe interessam. Um caso comum é uma empresa que se interessa em medir a aceitação de um novo produto, monitorando as opiniões de um grupo em relação ao produto. A análise de sentimento é um tipo de mineração de dados que possibilita tal demanda.

Existem abordagens de análise de sentimento que interpretam os dados de diferentes maneiras. O dicionário léxico trata-se de um conjunto das palavras de que dispõe de *scores* para mensurar os pesos para cada termo, chegando a um valor entre 0 e 1 para a sentença analisada. Existem basicamente duas abordagens para o aprendizado de máquina, que é a supervisionada e a não supervisionada. A primeira é caracterizada pelo fato de que existe a figura de um supervisor ou professor que ensina ao algoritmo o que cada registro significa.

## **3. Trabalhos Relacionados**

Nesta seção são apresentados trabalhos relacionados na literatura que propõem abordagens de análise de sentimento em textos, embasando a natureza deste experimento.

[Araújo et al. 2013] realizaram um estudo comparativo entre oito abordagens de análise de sentimentos, cada qual com sua característica relevante, tornando capaz a

análise do sentimento nos *tweets*. Os autores utilizaram duas bases de dados com características distintas, uma base histórica e outra base rotulada da Web, definindo métricas comparativas entre as abordagens e observado suas vantagens e desvantagens. Adicionalmente, os autores propuseram uma nova abordagem que realiza a combinação de características de várias abordagens de análise de sentimento buscando obter maior precisão nos resultados.

Parte das abordagens para análise de sentimentos em textos estão disponíveis apenas para o idioma inglês. Com o crescimento do conteúdo da internet disponibilizado em outros idiomas, alguns trabalhos têm sido propostos com o objetivo de explorar a análise de textos contendo opiniões ou sentimentos, independente do idioma no qual estão escritos, numa perspectiva multilíngue. Este experimento visou realizar um paralelo entre as abordagens com base em diferentes idiomas como alemão, português, espanhol. Como os resultados dos algoritmos de análise de sentimento são limitados pela interpretação apenas da língua inglesa, o estudo foi viabilizado por meio de uma aplicação desenvolvida para interagir com o *Google Translate*<sup>2</sup> capacitando a conversão dos dados.

Conforme [Martins et al. 2015] até onde foi investigado, constatou-se que se carece de abordagens eficientes para análise de sentimento para o idioma português brasileiro. Além disso, as abordagens existentes para o idioma inglês produzem resultados inferiores quando aplicados a conteúdos na língua portuguesa brasileira. Sendo assim, a abordagem aqui proposta visa trazer benefícios com a utilização combinada de algumas abordagens e suas adaptações para conteúdos em português brasileiro. Trata-se de experimentos voltados para interpretar qual o resultado obtido entre análises realizadas na língua portuguesa brasileira ou convertidas para a língua inglesa, possibilitando a utilização de ferramentas com um dicionário vasto de termos.

Para a realização deste trabalho, foram realizadas várias pesquisas relacionadas ao tema, principalmente quanto a necessidade de tratar estes dados na língua portuguesa brasileira e em diferentes contextos. Os estudos da academia que demonstram, pelo grande volume de dados, a necessidade de entender melhor como as abordagens de análise de sentimento funcionam embasaram este trabalho.

## 4. Análise Comparativa

### 4.1. Abordagens Comparadas

A primeira das abordagens comparadas é a denominada *SentiWordNet*<sup>3</sup>. Tal abordagem realiza a mineração de opinião em textos a partir de um dicionário de palavras extraído da *WordNet*<sup>4</sup>, um banco de dados léxico contendo milhares de termos no idioma inglês [Esuli and Sebastiani 2006].

O *WordNet* agrupa adjetivos, verbos e outras classes gramaticais em conjuntos chamados *synsets*. A abordagem *SentiWordNet* associa a cada *synset* um peso que indica a intensidade do sentimento que aquele *synset* remete: positivo, negativo e neutro.

---

<sup>2</sup>*Google Translate* é um serviço de tradução multilíngue de máquina fornecida pelo Google para traduzir o textos.

<sup>3</sup>*SentiWordNet* é um recurso léxico para a mineração de opinião em texto em inglês. Atribuindo a cada *synset* de *WordNet* três contagens de sentimento: positividade, negatividade, objetividade.

<sup>4</sup>*WordNet* é um banco de dados léxico para o idioma Inglês

Cada peso tem seus valores variando entre 0 e 1 e é obtido utilizando-se um método de aprendizagem de máquina semi-supervisionado. Para melhor entender o funcionamento da abordagem, suponha que um dado *synset*  $s = [bad, wicked, terrible]$  tenha sido extraído de um *tweet*. O resultado obtido pela abordagem é 0,000 para positividade, 0,850 para negatividade e 0,150 para objetividade. A avaliação do *SentiWordNet* é feita utilizando um dicionário léxico rotulado, que apresenta informações relevantes. Segundo [Lucas Ventura de Souza 2011], a abordagem *SentiWordNet* apresenta desempenho positivo em tarefas de mineração de opinião.

O *Sentiment140*<sup>5</sup> é também um dicionário léxico de palavras associadas a sentimentos positivos e negativos [Mohammad et al. 2013]. A abordagem foi criada com uma base de dados que consistia de cerca de 1,6 milhões de *tweets* rotulados como positivo ou negativo. Esta ferramenta possui um serviço de requisições web com grande dicionário para as línguas inglesa e espanhola mas para este experimento, foi implementado uma aplicação capaz de interagir com a plataforma do *Google Translate* viabilizando a utilização desta abordagem que possui capacidade apenas de manipular dados em inglês e espanhol.

Segundo [Chikersal et al. 2015] a ferramenta *Sentiment140* utiliza um classificador de aprendizagem de máquina baseada na presença de textos em inglês e de *emojicons*. A ferramenta usa da palavra e parte do discurso acoplado com um máximo de aprendizagem automática baseada em entropia do classificador ou pontuador da sentença, sendo algumas das ferramentas autônomas única. *Sentiment140* fornece um valor sentimento em uma escala de 0 ( negativo ), 2 ( neutro ) a 4 ( positiva ). Para obter melhores valores de comparação são convertidos em obter três categorias de sentimento positivo, negativo e neutra.

O *Tableau*<sup>6</sup> é uma ferramenta capaz de explorar informações de um grande conjunto de dados, definindo a maneira na qual estarão organizados e trazendo resultados junto com o experimento da análise. Em contrapartida a abordagem não disponibiliza um dicionário léxico de termos para a análise de sentimento. Sendo assim, foram definidos um conjunto de termos capazes de expressar sentimento para os contextos dos termos que foram analisados no experimento. Assim, para obter maior ganho foram contemplados erros de grafia, conforme visualizados em recorrência na base de dados.

Como a abordagem *Tableau* não possui um dicionário léxico para a realização da análise de sentimento, foi definido um dicionário de termos para demonstrar quais os termos expressam sentimento positivo, negativo e neutro.

## 4.2. Configuração Experimental

### 4.2.1. Fluxo de Atividades

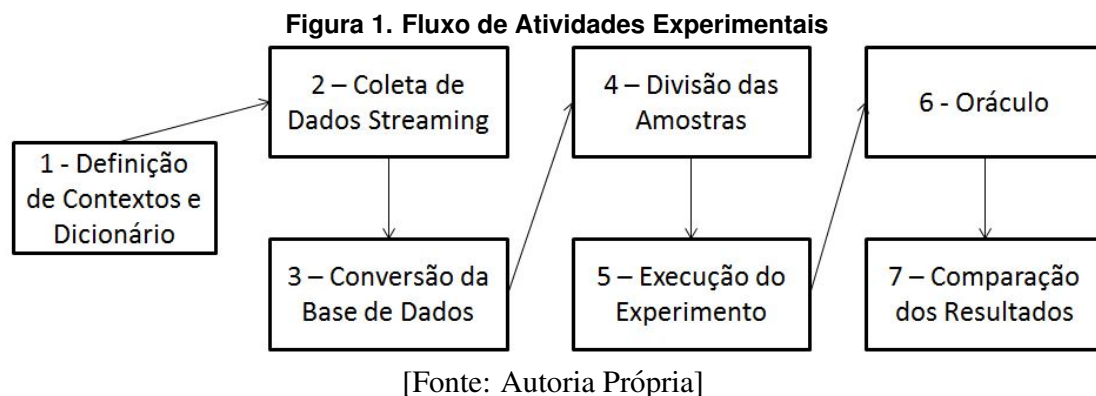
Nesta seção são apresentadas as configurações experimentais realizadas para a concepção deste estudo. A Figura 1 apresenta o fluxo de atividades experimentais.

Primeiramente, para este experimento, foram definidos os termos coletados e classificados devidamente de acordo com cada contexto pertencente. A Tabela 1 apresenta os

---

<sup>5</sup>Ferramenta de análise de sentimento baseada no *Twitter*.

<sup>6</sup>Tableau é um software de visualização de dados que lhe permite ver e compreender dados.



termos utilizados para coleta por categoria.

**Tabela 1. Termos Coletados para o Experimento por Contexto.**

Organizações	"BNDES"	"FIFA"
Pessoas	"Faustão"	"Neymar"
Eventos	"Dia das Mães"	"Campeonato Mineiro"
Produtos	"Mobi"	"iPhone"

As abordagens submetidas a este experimento são *SentiWordNet*, *Sentiment140* e *Tableau*. Destas, apenas a abordagem *Tableau* não possui um dicionário léxico previamente definido para a realização da análise de sentimento. Visando viabilizar a utilização de uma abordagem capaz de interpretar textos no idioma português brasileiro, neste experimento foi definido um dicionário próprio com mensagens positivas, expressando sentimentos positivos relacionados à satisfação, negativas, expressando sentimentos negativos relacionados à frustração, e neutras, não expressando sentimento algum. A Tabela 2 apresenta o conjunto de termos utilizados na abordagem *Tableau* como dicionário léxico.

**Tabela 2. Dicionário de Termos Utilizados na Abordagem *Tableau***

Positivos	Negativos	Neutros	Positivos	Negativos
Feliz	Chateado	Informativo	Eu quero	Não quero
Adorei	Detestei	Notícia	Te amo	Odeio
Amei	Odiei	Olhem	Feliz	Pior
Curti	Ruim	Encontrei	Bom	Não Gostei
Melhor	Péssimo	Achei	Joga Muito	Perna de Pau
ótimo	Horrível	Vemos	Ainda Vou	Nunca Irei
Me da	Feio	Lance	Bonito	Pecado
Quero	Insuportável	Situação	Lindo	Não Curti

A coleta de dados foi realizada a partir de um *Crawler* que monitora todos os dados que circulam na rede social *Twitter*, por meio de uma *API* de *streaming*. Assim, foi desenvolvida uma ferramenta utilizando o *framework* da Microsoft<sup>7</sup> .NET versão 4.5. O mesmo consiste em uma aplicação *console* que foi parametrizada com termos da Tabela 1 que monitorou a rede buscando citações para os termos descritos na Tabela 1, utilizados contextos que possuem grande volume de dados no *Twitter*.

<sup>7</sup>Microsoft é uma empresa transnacional americana de softwares de computador

Foram coletados cerca de 2.000 *tweets* sobre cada termo da Tabela 1, que foram armazenados em arquivos de texto (.txt). A partir dos dados coletados foi selecionada uma amostra aleatória de 100 *tweets*. Apenas a abordagem *Tableau* é capaz de interpretar dados no idioma português brasileiro, assim foi desenvolvida uma segunda aplicação *console*, que interage junto ao *Google Translate* realizando a conversão de toda a base de dados para o idioma inglês. Foram realizados pequenos ajustes de palavras na base de dados pois as mesmas possuíam apóstrofe que eram interpretadas em baixo nível de programação.

Ao consolidar uma base de dados capaz de ser aplicada ao cenário de todas as abordagens, foram definidas amostras selecionadas aleatoriamente para a submissão ao experimento. A partir de cada amostra de 100 *tweets* dividiu-se a amostra total de dados em  $k$  subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir disto, um subconjunto foi utilizado para o experimento. Perante as amostras definidas e submetidas a interpretação das abordagens, aplicou-se um gabarito ou oráculo manualmente em todas os subconjuntos para identificar a assertividade das ocorrências. Com estes insumos foram realizadas comparações de acordo com as métricas definidas na seção 4.

#### 4.2.2. Métricas de Avaliação

Segundo [Ribeiro et al. 2015] um aspecto chave na avaliação das abordagens para a análise de sentimentos diz respeito às métricas utilizadas. Neste contexto, três métricas principais são comumente empregadas para validar a eficiência de um método: acurácia, precisão e revocação.

Cada letra na Tabela 3 representa o número de instâncias de texto cuja classe correta é  $X$  e cuja predição é a classe  $Y$ , onde  $X:Y$  positivo; neutro; negativo.

**Tabela 3. Matriz Confusão**

	Positiva Máquina	Neutra Máquina	Negativa Máquina
Positiva Humano	A	B	C
Neutra Humana	D	E	F
Negativa Humana	G	H	I

A revocação ( $R$ ) da classe  $X$  é a taxa de número de elementos corretamente classificados pelo total de elementos na classe  $X$ . Já a precisão ( $P$ ) de uma classe  $X$  é taxa de número de elementos classificados corretamente pelo total de elementos classificados como sendo da  $X$ .

Considera-se igualmente importante a correta classificação de cada sentença, independente da classe, ou seja, ela mede basicamente a capacidade da abordagem em prever uma entrada corretamente. Como forma de permitir uma comparação global entre as abordagens foi utilizado um critério de comparação simples mas que permite ter uma ideia interessante da performance. A métrica é basicamente o *rank* médio em que um método ficou em cada conjunto. Por exemplo, se um método ficou em primeiro lugar, ou seja no *rank* 1, em todos os conjunto, seu *rank* médio será, obviamente 1. Para realizar este cálculo bastou somar o *rank* do conjunto em cada conjunto e dividir pela quantidade de conjuntos utilizados no experimento.

A validação cruzada dividiu o conjunto de dados em 5 partes de igual tamanho, sendo utilizadas para validação da amostra, extraindo a média e posteriormente extraímos o valor de desvio-padrão dos dados. [PERISSINOTTO and DE MOURA 2007]

Segundo [Refaeilzadeh et al. 2009], a validação cruzada é um análise estatística capaz de avaliar e comparar algoritmos de aprendizagem por dados dividindo-se em dois segmentos: um usado para aprender ou treinar um modelo e outro usado para validar o modelo. A forma básica de validação cruzada é *k-fold cross-validation*, em K vezes de validação cruzada, os dados são primeiro divididos em K igual (ou quase iguais) segmentos de tamanho ou dobrados. Dentro do experimento a amostra foi dividida em 5 amostras menores para aplicarmos a análise estatística da validação cruzada.

Ao utilizarmos a validação cruzada extraímos informações como desvio-padrão das amostras e média que posteriormente foram utilizados como insumo para o intervalo de confiança obtendo um experimento com dados confiáveis. Além disto, após o tratamento dos dados coletados no *Twitter*, traduzidos para inglês ou organizados em português, o experimento foi iniciado.

A partir dos dados extraídos da validação cruzada como média e desvio-padrão, utilizaremos para extrair a informação relevante para analisarmos os resultados, o intervalo de confiança *T-Student*, este será calculado com 95% para demonstrar os ganhos obtidos. [Altman 1990]

## 5. Resultados das Comparações

Para alcançar o objetivo de identificar vantagens, desvantagens e possíveis limitações das abordagens na detecção de polaridade, apresenta-se os resultados das comparações feitas sobre eles. Os experimentos foram divididos por contexto, assim foi possível analisar de maneira específica cada conjunto de dados e seus respectivos resultados.

### 5.1. Análise para o Contexto de Produtos

Os valores obtidos para o experimento direcionado ao contexto de mercado de Produtos é descrito nesta subseção. Na Figura 2(a) exploramos a Acurácia e na Figura 2(b) exploramos a Precisão obtidas para o termo iPhone destacado na Tabela 1.

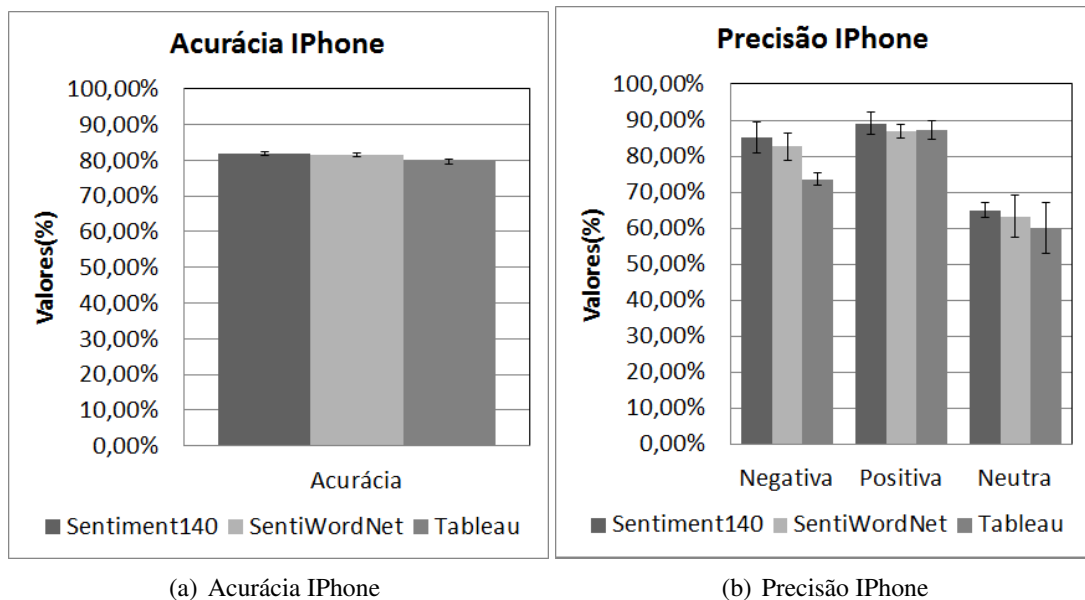
Observamos na Figura 2(a) que a abordagem *Sentiment140* obteve ganhos de 0,44% em relação a abordagem *SentiWordNet*, que posteriormente obteve ganhos de 1,76% sobre a abordagem *Tableau*. Assim, podemos constatar que estatisticamente não houveram ganhos significativos em relação a acurácia das abordagens analisadas neste contexto.

Quanto a precisão nota-se na Figura 2(b) que a abordagem *Sentiment140* obteve ganhos superiores nas três classificações de dados presentes no experimento mas sem valores estatisticamente expressivos para constataremos um destaque para alguma abordagem. Podemos observar que para a classificação negativa a abordagem *Sentiment140* é superior a abordagem *Tableau*.

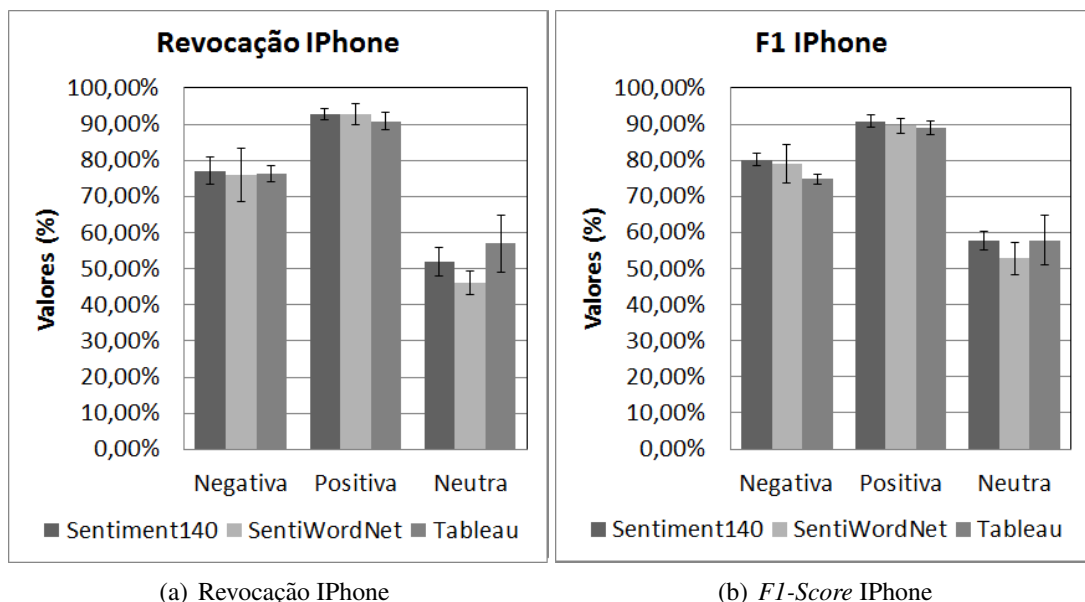
Na Figura 3(a) exploramos a Revocação e na Figura 3(b) exploramos o *F1-Score* no experimento realizado para o termo iPhone.

Observamos nas Figura 3(a) e Figura 3(b) que as abordagens obtiveram ganhos com valores aproximados em relação a Revocação para os classificadores negativos, posi-





**Figura 2. Análise dos Dados para *Tweets* Vinculados ao Termo iPhone**



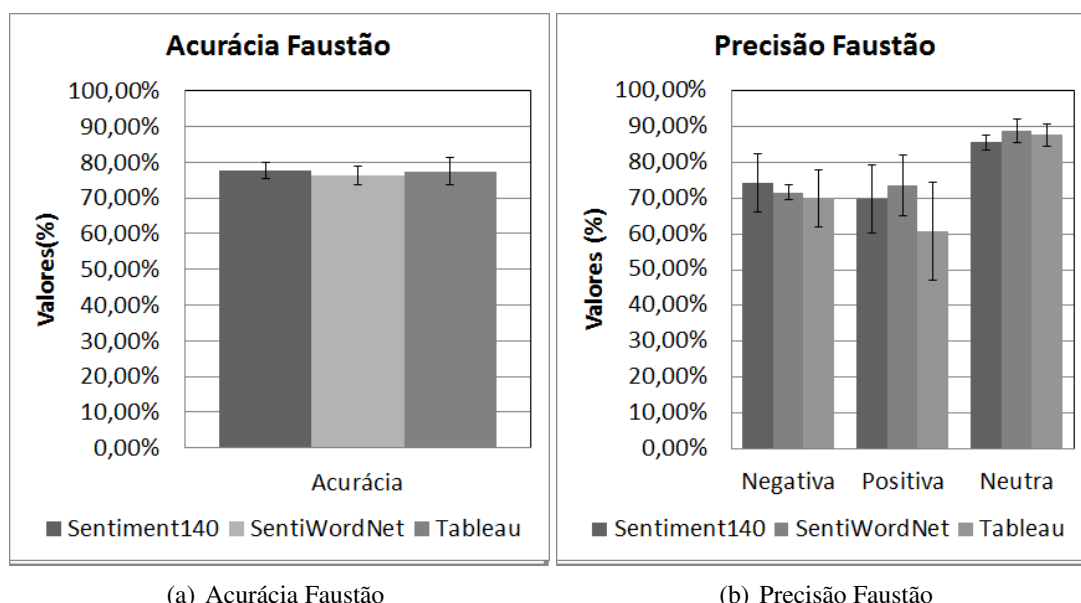
**Figura 3. Análise dos Dados para *Tweets* Vinculados ao Termo iPhone**

tivos e neutros. Levamos em consideração a abordagem *Sentiment140* que obteve ganhos superiores nos classificadores negativos, positivos e neutros mas estatisticamente não obtivemos valores significativos.

## 5.2. Análise para o Contexto de Pessoas

Os valores obtidos para o experimento direcionado ao contexto de mercado de Produtos é descrito nesta subseção. Na Figura 4(a) exploramos a Acurácia e na Figura 4(b) exploramos a Precisão obtidas para o termo Faustão na Tabela 1.

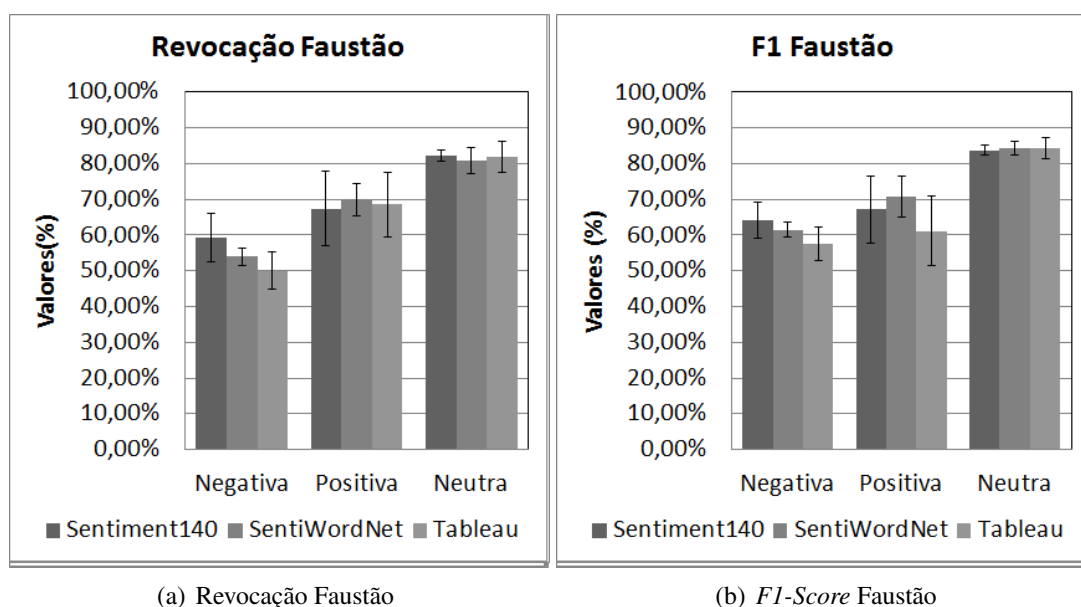
Na Figura 4(a) demonstramos os valores obtidos para o experimento com a pes-



**Figura 4. Análise dos Dados para Tweets Vinculados ao Termo Faustão**

soa Faustão observamos que os ganhos para a Acurácia não foram estatisticamente significativos pois ambas as abordagens chegaram a valores semelhantes sem um valor considerável com o intervalo de confiança. Ao analisarmos a precisão que o experimento alcançou perante ao termo Faustão, os ganhos alcançados foram semelhantes em ambas as classificações constatando valores estatisticamente insignificantes.

Na Figura 5(a) exploramos a Revocação e na Figura 5(b) exploramos o *F1-Score* no experimento realizado para o termo Faustão.



**Figura 5. Análise dos Dados para Tweets Vinculados o Termo Faustão**

Notamos observando a Figura 5(a) que os ganhos em relação a Revocação e na

Figura 5(b) os ganhos relativos ao *F1-Score* obtidos foram semelhantes em todas as classificações mas com uma pequena superioridade para a abordagem positiva e neutra nos dados já normalizados.

## 6. Conclusão

O experimento de coleta de dados e comparação entre abordagens visou abranger um processo de uma abordagem de análise de sentimento realizada por instituições ou pessoas, desde a sua coleta de dados até resultados gráficos. Buscando interpretar vantagens, desvantagens e possíveis limitações das três abordagens utilizadas de análise de sentimento sendo *Sentiment140*, *SentiWordNet* e *Tableau*. A partir da divisão por contextos cada abordagem reagiu a um contexto mercado de maneira diferente. Notamos que ao analisar os sentimentos relevantes, extraímos dados importantes para analisar a acurácia, predição, revocação e *F1-Score* das abordagens.

Destacando os valores alcançados quanto a Acurácia que em 98% dos casos no experimento não obtiveram ganhos estatisticamente significativos, observando também os valores de intervalo de confiança que acompanhavam a diferença dos ganhos obtidos. Vale ressaltar que este experimento define mais métricas para alcançarmos ganhos diferenciados e estatisticamente significativos para a academia, alavancando os ganhos alcançados pela abordagem *Sentiment140* para os contexto de Produtos e Pessoas que expressam sentimentos fortes de extremidade entre as classificações de positiva ou negativa. Destacamos os ganhos obtidos para a abordagem *Tableau* referente sentimentos simples sem ocorrências de ironia, que cruzem inúmeras informações dos *Tweets*, também observamos que para o contexto de Eventos a abordagem *SentiWordNet* obteve ganhos para contextos de organizações no qual o grande volume da dos foi para neutralidade em manchetes e notícias relacionadas as organizações presentes no experimento.

Notamos que devido a divisão por contextos, as amostras por si só demonstraram que devemos analisar o grande conjunto de dados e seu maior número de ocorrências como um volume massivo de dados neutros ou um conjunto de dados que expressem fortes sentimentos. Observamos que abordagens se comportam de maneira diferente em um conjunto de dados diferente, no qual foram refletidos em cada contexto com sua devida particularidade. Este experimento também pode ser expandido para mais abordagens de análise de sentimento e consequentemente evoluirmos a quantidade de métricas para alcançarmos sempre ganhos relevantes em estudos desta natureza.

## Referências

- Altman, D. G. (1990). *Practical statistics for medical research*. CRC press.
- Araújo, M., Gonçalves, P., Benevenuto, F., and Cha, M. (2013). Métodos para análise de sentimentos no twitter. In *Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)*.
- Chikersal, P., Poria, S., and Cambria, E. (2015). Sentu: sentiment analysis of tweets by combining a rule-based classifier with supervised learning. In *Proceedings of the International Workshop on Semantic Evaluation, SemEval*, pages 647–651.
- Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422. Citeseer.

- França, T. C., de Faria, F. F., Rangel, F. M., de Farias, C. M., and Oliveira, J. (2014). Big social data: Princípios sobre coleta, tratamento e análise de dados sociais. *XXIX Simpósio Brasileiro de Banco de Dados – SBBD '14. Curitiba – PR.*
- Lucas Ventura de Souza, R. B. C. P. (2011). Análise de sentimentos no twitter utilizando sentiwordnet.
- Martins, R., Pereira, A., and Benevenuto, F. (2015). Uma abordagem para análise de sentimentos de aplicações da web em lingua portuguesa. *Brazilian Symposium on Multimedia and the Web.*
- Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242.*
- PERISSINOTTO, M. and DE MOURA, D. J. (2007). Determinação do conforto térmico de vacas leiteiras utilizando a mineração de dados/evaluation of thermal comfort in dairy cattle using data mining. *Revista Brasileira de Engenharia de Biosistemas*, 1(2):117–126.
- Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. In *Encyclopedia of database systems*, pages 532–538. Springer.
- Ribeiro, F., Araújo, M., and Benevenuto, F. (2015). Métodos para análise de sentimentos em mídias sociais. *Brazilian Symposium on Multimedia and the Web. (Webmedia).*
- Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.
- Timmermann, N., Rodrigues, C. A. S., Vieira, L. L., and Malagoli, L. (2013). Mineração de opinião / análise de sentimentos.
- Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.