

## **Análise de Sentimento para *Reviews* Apresentados em Vídeos: Modelo de Redes Neurais Treinado em Base de *Reviews* Escritos**

### **Sentiment Analysis for *Reviews* Featured in Videos: Neural Network Model Trained on Written *Reviews* Database**

DOI:10.38152/bjtv4n1-001

Recebimento dos originais:08/12/2020

Aceitação para publicação: 22/01/2021

#### **João Paulo Vieira Costa**

Mestrando em Computação Aplicada

Instituição: Universidade de Brasília

Endereço: ICC Centro -Módulo 14, Subsolo CSS-361 -Campus Darcy Ribeiro.  
Brasília -DF

E-mail: jpcosta1990@gmail.com

#### **Romulo Baldez de Barros**

Secretaria de Economia do Distrito Federal

Especialista em Business Analytics e Big Data

Endereço: SBN Qd 02 Ed. Vale do Rio Doce 5º andar, Brasília-DF

E-mail: romulobaldez1964@gmail.com

#### **Caio César Silva Dantas**

Especialista em Business Analytics e Big Data

Instituição: Banco do Brasil

Endereço: Av. do Sol, QD 09 Casa Acácias 13, Jardim Botânico, Brasília - DF

E-mail: caioaires@gmail.com

#### **Raquel Cristina de Sousa**

Especialista em Business Analytics e Big Data

Instituição: Banco do Brasil

Endereço: SAUN QD 5 LT B, Asa Norte, Brasília-DF

E-mail: raquel.sousa@gmail.com

#### **Cristiano Gonçalves Nascimento Gouveia**

Mestrando em Tecnologia Ambiental e Recursos Hídricos

Instituição: Universidade de Brasília

Endereço: Anexo SG-12, Térreo Campus Universitário Darcy Ribeiro. Brasília -DF

E-mail: gouveia.crs@gmail.com

**Gustavo Correa Mirapalheta**

Doutor em Administração de Empresas

Instituição: Fundação Getulio Vargas

Endereço: Rua Itapeva 474 9º andar Bela Vista, São Paulo - SP

E-mail: gustavo.mirapalheta@fgv.br

**RESUMO**

A análise de sentimentos tem sido apresentada como uma técnica importante para avaliação de *reviews* em grandes volumes de dados disponibilizados nas redes sociais. Este trabalho apresenta uma abordagem de busca dos textos a partir de transcrições de vídeos disponibilizados sobre *reviews* de *smartphones* e aplica a análise de sentimentos neste contexto, a partir do treinamento de modelo de Redes Neurais Convolutivas em uma base de dados com 10.063.255 *reviews* escritos de um site de compras online e avaliado o seu desempenho em classificações dos trechos de *reviews* classificados manualmente pelos pesquisados. Como resultado, conclui-se que o aproveitamento do padrão textual observado em uma base de dados de *reviews* escritos pode trazer resultados positivos mesmo aplicado em um padrão diferente e que não esteja disponível ou classificada para uma tarefa de aprendizado supervisionado.

**Palavras-chave:** Análise de Sentimentos, Text Mining, Redes Neurais.

**ABSTRACT**

Sentiment analysis has been presented as an important technique for evaluating reviews on large volumes of data available on social networks. This work presents a text search approach from video transcripts made available on smartphone reviews and applies sentiment analysis in this context from Convolutional Neural Network model training in a database with 10,063,255 reviews written from an online shopping site and evaluated its performance in ratings of the reviews excerpts manually classified by respondents. As a result, it is concluded that taking advantage of the textual pattern observed in a written review database can bring positive results even applied to a different pattern and that it is not available or classified for a supervised learning task.

**Keywords:** Sentiment Analysis, Text Mining; Neural Networks.

**1 INTRODUÇÃO**

Um dos caminhos para se obter informações sobre os mais variados assuntos é a partir das redes sociais. A todo o momento milhares de novos usuários interagem e trocam informações de diversos tipos de contexto através de websites, blogs, vídeos, imagens, redes sociais, etc. (CAETANO *et al.*, 2019). Como uma base para o seu processo de decisão de compra, essa oferta de opiniões é crescente e diversificada.

Em relação à busca de informações sobre produtos, uma das bases de consulta disponíveis para os consumidores são os *reviews* disponibilizados nas mais variadas plataformas para distribuição deste tipo de conteúdo. A informação pode ser transmitida

em vários formatos, e um deles é a partir de vídeos explicativos de usuários ou especialistas sobre as impressões sobre os produtos adquiridos.

A análise de sentimentos tem sido apresentada como uma técnica importante para avaliação de *reviews* em grandes volumes de dados disponibilizados nas redes sociais (CHAWLA *et al.*, 2018). Geralmente, essas análises estão voltadas para análises de comentários escritos deixados por usuários de redes sociais como Twitter ou mesmo nos comentários do YouTube (AFONSO, 2017) e tem por objetivo condensar uma grande quantidade de informações em uma classificação do contexto do texto (positivo, negativo). O interesse por resumir informações disponibilizadas em grandes bases de textos também é útil para vídeos, considerando a grande quantidade de vídeos disponibilizados nas redes sociais sobre estes temas. Este trabalho apresenta uma abordagem de busca dos textos a partir de transcrições de vídeos disponibilizados sobre *reviews* de *smartphones* e aplica a análise de sentimentos neste contexto.

A construção de um modelo de classificação baseado em textos, geralmente, utiliza como insumo uma base de dados já classificada para treinamento do modelo desenvolvido. No entanto, uma base de dados de trechos vídeos com *reviews* sobre *smartphones* classificados quanto aos comentários positivos e negativos não está disponível para uso. Apesar disso, há bases de *reviews* escritos com comentários e classificações (*ratings*) atribuídos pelos próprios usuários sobre produtos adquiridos em sites de compras *online*.

Neste contexto, este trabalho busca responder a seguinte pergunta de pesquisa: É possível desenvolver modelo de análise de sentimentos para classificação de textos falados em vídeos sobre *smartphones* utilizando *reviews* escritos de uma base de site de compras online?

## 2 REFERENCIAL TEÓRICO

Aprendizado de máquina é um subgrupo da área de inteligência artificial que automatiza a construção de modelos analíticos a partir de dados (BUSSON *et al.*, 2018). Pode ser definido também como o campo de estudos que dá aos computadores a habilidade de aprender sem serem explicitamente programados (SAMUEL, 1959 *apud* BUSSON *et al.*, 2018). Pode-se dizer que o aprendizado de máquina é constituído de algoritmos que têm um modelo interno, feito de forma automática a partir de amostras de entrada e, a partir daí, fazem previsões que, ao contrário de seguir instruções programadas, são guiadas pelos dados.

O aprendizado de máquina se divide em: supervisionado, não supervisionado e por reforço.

- **Aprendizado supervisionado:** aqui as amostras fornecidas são pré-classificadas para processamento. Neste caso, a máquina deve aprender uma função que mapeia entrada para um tipo de saída (BUSSON et al., 2018). Esse tipo de aprendizado é usado para classificação e regressão.
- **Aprendizado não supervisionado:** aqui os dados são fornecidos sem classificação. Neste caso a máquina deve encontrar alguma estrutura (padrão) nos dados. Esse tipo de aprendizado é usado na clusterização, pois não há um conhecimento prévio de como os dados estão agrupados. Cabendo ao computador definir a quantidade dos grupos a partir das técnicas utilizadas (distância euclidiana, p. ex.)
- **Aprendizado por reforço;** quando a máquina tem de cumprir um objetivo dado e recebe prêmios ou punições durante o processo.

## 2.1 REDES NEURAIAS

As Redes Neurais são modelos utilizados no aprendizado de máquina para tarefas de aprendizado supervisionado também utilizadas em tarefas associadas a mineração de textos (HOLE, 2018). Dentre as diferentes estruturas de Redes Neurais, destacam-se as Redes Convolutivas, que foram as primeiras redes neurais a terem sucesso em aplicações comerciais relevantes e continuam tendo um papel importante nestes tipos de problema (GOODFELLOW *et al.*, 2016).

## 2.2 ANÁLISE DE SENTIMENTOS

Análise de sentimentos é o processo de identificação das opiniões e sentimentos em um conjunto de textos (corpus) (BRITO, 2017). Tem como principal objetivo definir técnicas automáticas capazes de extrair informações subjetivas em linguagem natural, como opiniões e sentimentos (BENEVENUTO et al.). A análise de sentimentos tende a simplificar e reduzir uma opinião complexa em uma polaridade (RODRÍGUEZ-OLIVEIROS et al., 2017). Atualmente tem um amplo campo de aplicação. Sendo utilizada para aferir a avaliação de um produto, pesquisa eleitoral, opinião sobre um filme e até prever o comportamento da Bolsa de Valores. No caso dos produtos se usa atualmente, a análise baseada em comentários às publicações no Twitter e a vídeos publicados no YouTube. Neste trabalho, a análise de sentimento foi feita diretamente no

vídeo do YouTube que produz a *review* sobre um produto, no nosso caso aparelhos celulares. Pois, por se tratar de uma opinião dada por um especialista, o sentimento expresso por ele em seus comentários sobre o produto muito provavelmente exercerá influência sobre aqueles que pretendem adquirir o mais recente lançamento de celular.

Na análise de sentimentos, a polaridade representa o grau de positividade ou negatividade do texto (BENEVENUTO et al.). É utilizado para a análise de sentimentos, pois classifica o texto de duas formas: apresentando o resultado de forma binária (positivo ou negativo); ou de forma ternária (positivo, negativo e neutro). Os sentimentos positivos são expressos pelas avaliações favoráveis ao celular, ao contrário dos sentimentos negativos, que denotam desaprovação ao produto como um todo, ou à alguns aspectos dele. As frases classificadas como neutras não expressam opinião, apenas descrevem ou explicam algo sobre o aparelho. Por exemplo: a resolução da tela do celular é muito boa (positivo); a câmera do celular não atendeu às expectativas (negativo); o celular XYZ é o mais recente lançamento da empresa (neutro).

### 3 METODOLOGIA

#### 3.1 DADOS UTILIZADOS

O maior problema na disponibilidade de dados para realização deste trabalho está na inexistência de dados classificados para modelos de aprendizado supervisionado em *reviews* de YouTube. Neste caso, foi necessário acessar mais de uma fonte de dados com o objetivo de identificar padrões que possam ser aproveitados na análise de sentimentos dos *reviews*.

Nesse caso, além dos dados para desenvolvimento do modelo, faz-se necessária a obtenção de dados do YouTube tanto para aplicação do modelo quanto para avaliação do seu desempenho neste contexto. Dessa forma, há duas bases de dados principais acessadas:

- *Reviews* Escritos: base dados com 10.063.255 *reviews* de telefones celulares e acessórios da Amazon, coletados de 1996 a 2018; e
- Transcrições do YouTube: Legendas extraídas de vídeos do YouTube com *reviews* sobre smartphones.

### 3.1.1 Reviews Escritos

Os dados foram extraídos da internet<sup>1</sup>, armazenados em formato json, com os *reviews* escritos, a classificação dada de 1 a 5 pelo usuário e outras informações que não foram utilizadas neste trabalho. O principal desafio em relação a estes dados está em seu volume, tendo sido necessários fracionamento do arquivo para possibilitar sua importação, conforme detalhado nos códigos incluídos nos Anexos deste trabalho.

Após importação, foram feitos procedimentos para retirada de *stopwords* descritos nos códigos em anexo, válidas para todos os modelos desenvolvidos. Além desses tratamentos, outros procedimentos foram feitos para aplicação do modelo, no entanto, estas serão detalhadas nos tópicos de cada um deles.

No escopo do trabalho foram consideradas apenas classificações de uma ou duas estrelas (negativos) ou cinco estrelas (positivos). Classificações com três ou quatro estrelas foram retiradas devido às possíveis ambiguidades existentes nessas classificações. Além deste filtro, foram retirados textos nulos, restando uma base disponível para modelagem com 7.677.226 *reviews*.

### 3.1.2 Transcrições YouTube

Dado o objetivo deste trabalho de analisar os sentimentos em *reviews* de YouTube, a coleta das transcrições dos vídeos é uma etapa necessária. Neste trabalho, essa etapa foi feita a partir das legendas, em língua inglesa, associadas aos vídeos.

Foi criada uma base de transcrições de vídeos do YouTube, referentes a reviews de celulares, a partir da seleção de 7 vídeos de 4 modelos diferentes (Samsung A51, Samsung A71, Xiaomi mi 10 e Samsung M40), apresentados em 3 canais. Foram selecionados canais relevantes, especializados no assunto e que apresentavam, de maneira clara e objetiva, os principais pontos dos celulares. A Tabela 1 apresenta informações sobre os canais selecionados.

**Tabela 1** - Informações dos Canais do YouTube, obtidas em 19/07/2020

Canal	Número de Inscritos	Número de Visualizações	Número de Vídeos
Geekyranjit	3 milhões	644.621.792	2.817
Android Authority	3,36 milhões	749.530.451	3.805
C4ETech	1,77 milhões	374.457.417	2.676

<sup>1</sup> Jianmo Ni, Jiacheng Li, Julian McAuley  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2019  
<http://deepyeti.ucsd.edu/jianmo/amazon/index.html>

Na primeira etapa será feito um treinamento de um modelo de classificação dos textos a partir de uma base de textos da Amazon que foi utilizada para classificar uma base de dados das transcrições de vídeos do YouTube.

Na segunda etapa, foram classificados, manualmente, os vídeos dos celulares do YouTube da seguinte maneira: através de um código em Python, utilizou-se uma que dá acesso as transcrições dos vídeos do YouTube. Neste caso essa *API* permite baixar as transcrições das *closed captions* dos vídeos das *reviews* dos produtos objeto desse trabalho, no caso de objetos celulares. Esse conjunto de transcrições de um vídeo é transformado em uma base de dados.

### 3.2 FERRAMENTAS PARA MINERAÇÃO DE DADOS

Para aplicação do modelo de redes neurais, foi utilizado o Keras, que é um framework de alto nível para Python, capaz de rodar sobre o Tensorflow, a sua maior vantagem de uso é o tempo poupado pelos API's de alto nível, que são poderosos mesmo sendo fáceis de usar, e que permitem uma prototipagem rápida (SARKAR et al., 2018). Permite também que se use os constructos oferecidos pelo Tensorflow de maneira mais intuitiva e fácil de usar (SARKAR et al., 2018). Algumas de suas principais vantagens são: prototipagem fácil e rápida (como dito acima), suporte a redes convolutivas e recorrentes, incluindo combinação de ambas, suporte a esquemas de conectividade arbitrária (incluindo treino de N para N) e rodar na CPU ou GPU.

## 4 RESULTADOS

### 4.1 ANÁLISE EXPLORATÓRIA

Previamente ao desenvolvimento do modelo, a análise exploratória de dados é uma boa prática para se encontrar padrões nos vídeos que possam ser aproveitados ou mesmo para uma melhor compreensão sobre os dados trabalhados. A Figura 1 serve para ilustrar o objetivo deste trabalho. Um bom modelo deve identificar nas palavras e na sua organização entre os textos os fatores associados a comentários negativos e não negativos.

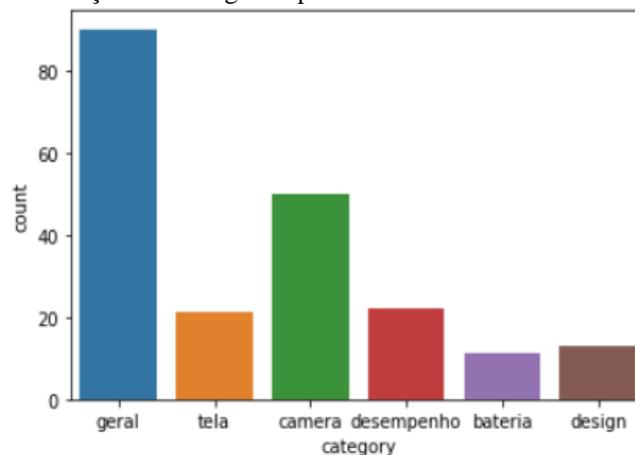
**Figura 1** - Nuvens de Palavras dos Trechos Negativos e Não Negativos nos Vídeos Analisados



Percebe-se que, nos dados analisados, a simples frequência das palavras não aparenta conferir um significado claro aos textos. A associação das palavras na formação da frase deve indicar os sentimentos a elas associadas, de maneira similar à compreensão humana.

Além da classificação dos dados como negativos ou não negativos, cada trecho foi classificado quanto ao tema tratado. Sendo classificado nas categorias Geral, Tela, Câmera, Desempenho, Bateria e *Design*. A Figura 2 mostra a distribuição do tempo dos vídeos analisados dedicados a cada uma dessas categorias.

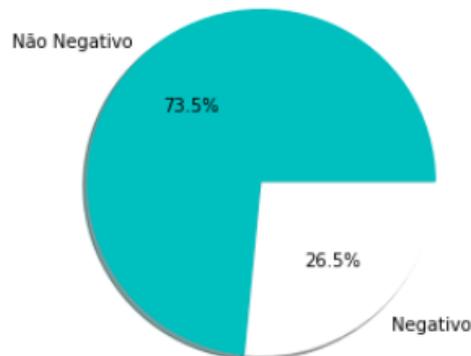
**Figura 2** - Distribuição das Categorias por Minutos Falados nos *Reviews* do YouTube



A maior parte do tempo nos vídeos analisados foi destinado a comentários gerais ou mesmo sem relação com os aparelhos. Todo o conteúdo apresentado nas frases ditas nos vídeos tem como objetivo prever os comentários negativos feitos em relação aos aparelhos naquele vídeo.

Em nenhum dos modelos desenvolvidos estes dados foram usados para modelagem devido à indisponibilidade de *reviews* classificados do YouTube. A pressuposição principal para desenvolvimento do modelo se refere à relação da disposição das palavras em *reviews* falados, nas legendas do YouTube, ou *reviews* escritos, em avaliações de usuários da Amazon sobre celulares e acessórios. A Figura 3 apresenta proporção *reviews* negativos e não negativos na base de modelagem.

**Figura 3** - Distribuição das Classificações dos *Reviews* da Amazon (7.677.226 *reviews*)



Destaca-se o desbalanceamento da base, com número mais elevado de comentários não negativos que negativos. Esse comportamento deve ser considerado no momento da avaliação da qualidade dos modelos ajustados, pois medidas baseadas em matrizes de confusão podem apresentar boas taxas de acerto, sem fazer, de fato, uma boa distinção dos comentários.

#### 4.2 MODELAGEM PREDITIVA

Esta seção descreve as etapas realizadas para criação dos modelos preditivos para classificação dos *reviews* do YouTube. Foi desenvolvido modelo de Redes Neurais e seus resultados foram comparados com o modelo Vader, modelo pré-treinado do Orange como *Benchmark*.

Como mencionado, o desbalanceamento da amostra de modelagem ou de aplicação do modelo pode tornar a medida de acurácia problemática. Um modelo que

classificasse todos os vídeos da base de modelagem como Não Negativo apresentaria uma taxa de acerto de 73% na base de *reviews* da Amazon.

Por este motivo, além da medida de acurácia (taxa de acerto), a *Area Under the Receiver Operating Characteristics* (AUROC) foi levada em consideração na avaliação dos modelos, pois esta medida mensura o desempenho dos modelos a partir da relação entre falsos positivos e falsos negativos, o que pode distinguir uma boa acurácia devido ao desbalanceamento de uma boa classificação feita pelo modelo.

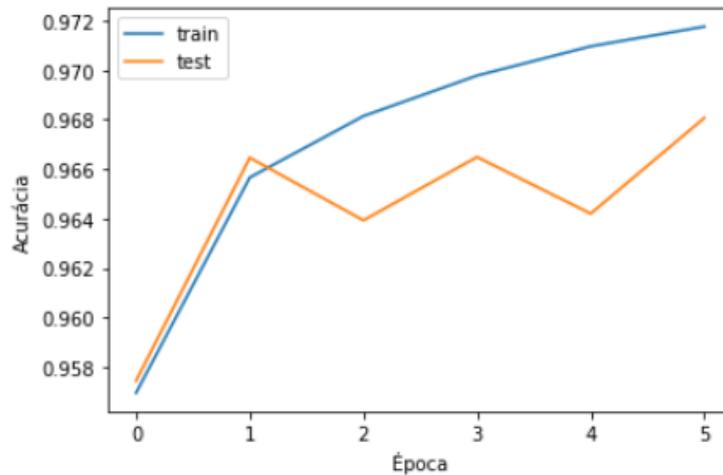
#### 4.2.1 Redes Neurais

Ao se desenvolver um modelo de redes neurais aplicado a análise de sentimentos, as variáveis explicativas, nas quais são realizados os cálculos dos parâmetros são os textos a partir dos quais são previstas as probabilidades de determinado sentimento.

Nesse sentido, uma etapa necessária para a aplicação do modelo é a transformação dos textos com palavras em vetores a serem utilizados no modelo. Esta transformação foi feita a partir dos vetores GloVe (Pennington *et al*, 2014), modelo já popularizado para tarefas de mineração de texto. É comum em modelos de redes neurais em textos a limitação de vocabulário para aquele conjunto que de fato confere contexto e significado às frases. Neste trabalho, o vocabulário foi limitado a 7.000 palavras, principalmente por limitação de processamento nos equipamentos disponíveis. Em todos os casos a base de dados foi dividida em 80% para e 20% para teste.

Foram testadas diferentes configurações de redes neurais, sendo o melhor resultado obtido pela rede neural convolutiva com função de ativação ReLU e sigmoide. A Figura 6 mostra a história do treinamento do modelo, realizada em seis épocas na amostra de treino e nos *splittings* de validação (20%)

**Figura 6** - *History* da Rede Neural Convolutiva



Como pode ser observado, tanto na amostra de treino como na amostra de teste, o modelo apresentou boa acurácia, sempre acima de 95% e crescente nas duas curvas. Há certo descolamento nas últimas épocas, o que pode indicar fragilidade de generalização da amostra de treino para a amostra de teste, o que pode ser objeto de melhorias do modelo testando-se outras configurações e partição das bases. O modelo foi aplicado a 20% da base obtendo resultados semelhantes, com 96,82% de acurácia e custo de 9,12%, sendo que na amostra de treino os resultados foram 97,18% de acurácia e 8,08% de custo, não havendo perdas significativas na generalização do modelo obtido.

#### 4.2.2 Comparação na base do Youtube

Diante dos resultados obtidos em cada um dos modelos, pretende-se avaliar qual deles apresentou os melhores resultados para ser utilizado na classificação dos vídeos. Além dos dois modelos desenvolvidos, foram incluídos os resultados do Vader aplicado a partir do Orange.

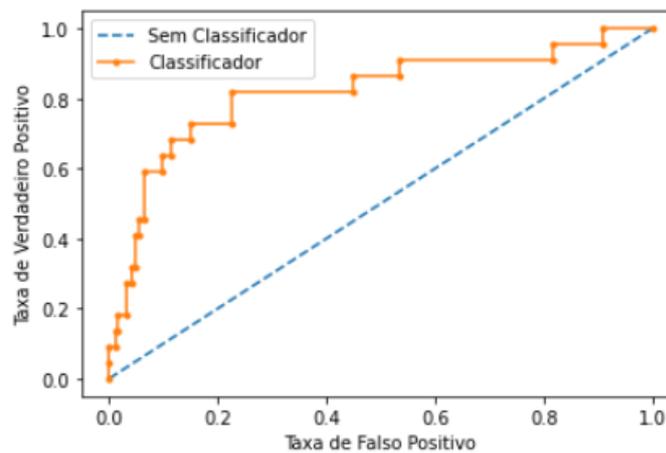
Destaca-se que nenhum dos dados foi utilizado para treino, ou mesmo teste dos modelos, sendo testada uma única vez para analisar a capacidade dos modelos de transferir o conhecimento adquirido em um novo contexto. A Tabela 2 descreve os resultados dos modelos desenvolvidos para este trabalho, sendo apresentadas as medidas de acurácia e AUROC.

**Tabela 2** – Medidas de Desempenho dos Modelos

Modelo	Acurácia	AUROC
Benchmark <sup>2</sup>	0,8164	0,8092
Rede Neural	0,8647	0,8179

Já o modelo de Redes Neurais aplicado, apresentou um bom resultado na acurácia, acrescido de um desempenho considerável na AUROC. Este resultado indica que o modelo treinado com 7.677.226 de *reviews* conseguiu identificar um padrão de forma replicável em outra fonte de informações, em um contexto de *reviews* falados, ao invés de escritos. A Figura 7 apresenta a curva ROC calculada para as classificações feitas pela rede neural na base classificada de *reviews* do YouTube.

**Figura 7** - Curva ROC da Rede Neural na Base do YouTube

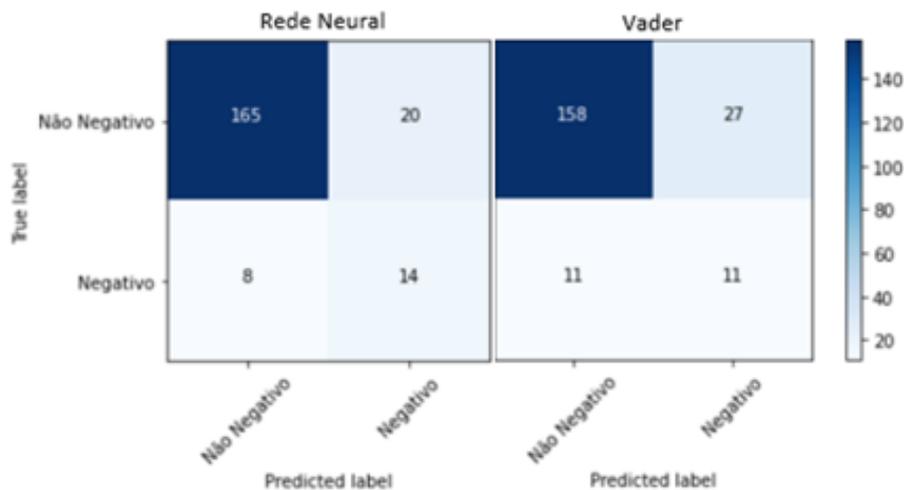


Vale destacar que o desempenho do modelo pré-treinado do Orange, utilizado como *benchmark*, apresentou bons resultados. No entanto, o modelo de redes neurais desenvolvido apresentou melhor desempenho, ao aproveitar a organização das palavras para formação de sentimentos nos *reviews* da Amazon sobre celulares.

A diferença entre o desempenho dos modelos pode ser considerada sutil nos indicadores da Tabela 1, mas as matrizes de confusão, apresentadas na Figura 8, apresentam um ganho pela utilização do modelo de Redes Neurais.

<sup>2</sup> A análise de sentimento feita após a etapa do pré-processamento, tem como objetivo classificar os trechos como positivos ou negativos. Utilizou-se para classificação o léxico VADER (*Valence Aware Dictionary for sEntiment Reasoning*), utilizado principalmente para análise de sentimentos de mensagens postadas em mídias sociais, prescindindo de treinamento. O modelo Vader calcula um indicador que varia de -1 a 1, os valores previstos como negativos foram os com este indicador menor do que 0.

**Figura 8** - Matrizes de Confusão do Modelo Proposto x Benchmark



O modelo de Redes Neurais mostrou maior quantidade de classificações nos verdadeiros positivos e nos verdadeiros negativos, indicando uma melhor adequação aos dados analisados. Dessa forma, a partir do entendimento de que este foi o melhor modelo, as análises para as discussões posteriores deste trabalho serão feitas utilizando-o.

A Tabela 3 mostra o tempo total gasto com informações negativas sobre os aparelhos analisados nos vídeos de cada um dos canais e sua comparação com o tempo previsto pelo modelo.

**Tabela 3** - Minutos Negativos Esperados vs Observados

Aparelho	Canal	Minutos Negativos		Minutos Totais
		Real	Previsto	
Samsung M40	Geekyranjit	5,0	5,5	16,0
Sansung A51	Android Authorit	5,5	5,5	17,0
	C4ETech	0,5	1,5	13,5
Sansung A71	C4ETech	0,0	0,5	14,0
	Geekyranjit	0,0	1,0	15,5
Xiaomi mi 10	Android Authorit	0,0	0,5	10,5
	C4ETech	0,0	2,5	17,0
<b>Total</b>		<b>11</b>	<b>17</b>	<b>103,5</b>

Percebe-se que a estimativa dos minutos feita pelo modelo se aproxima do ocorrido nos vídeos. Apesar disso, nota-se que, nos vídeos em que se observou pouco ou nenhum tempo com comentários negativos, houve maior distanciamento nas estimativas do modelo.

## 5 CONSIDERAÇÕES FINAIS

Tendo em vista a relevância do conhecimento gerado nas redes sociais e da variedade das informações nos tempos atuais, o aprimoramento das análises para dar valor a essas informações deve ocorrer constantemente. Especificamente quanto ao YouTube, a quantidade de informação diária gerada o transforma em um repositório de informação muito rico. Para muitos temas, o YouTube é uma das principais plataformas utilizadas para transmissão de conhecimento, a exemplo dos *reviews* sobre telefones celulares.

Além de várias outras informações disponíveis nesta plataforma, os textos transcritos ganham relevância tendo em vista o avanço das técnicas para análise de textos e aquisição de conhecimento a partir deles. Em linhas gerais, este trabalho apresentou uma abordagem para transferência de conhecimento adquirido em bases de *reviews* escritos para análise de sentimentos em transcrições de vídeos do YouTube.

Um dos pontos interessantes em se trabalhar com textos de *reviews* falados é que, adicionalmente aos escritos, pode-se atribuir tempo aos textos analisados. Cada trecho analisado foi segmentado em espaços de 30 segundos, possibilitando uma estimativa de minutos dedicados a desaprovação de um determinado produto. Cada visualização recebida por qualquer um dos vídeos, resulta em uma percepção negativa no espectador. Portanto, essa percepção pode se destacar como uma ferramenta de acompanhamento da aceitação de produtos pelas empresas.

O trabalho foi focado na apresentação do modelo e na capacidade de transferência de aprendizado do modelo em *reviews* da Amazon para *reviews* do YouTube. A relevância dos vídeos analisado e a seleção automática de vídeos em volume elevado, que não possa ser classificado manualmente, para posterior análise de sentimentos necessita de APIs ou procedimentos específicos para obtenção dos IDs ou dos *links* dos vídeos, o que seria uma melhoria para trabalhos futuros.

A partir da extração de análise de volume expressivo de vídeos em canais relevantes e de considerável número de visualizações, um estudo futuro possível a partir da aplicação proposta neste trabalho é a avaliação da reputação de aparelhos lançados ou mesmo de marcas de celulares, assim como apresentado por (VIDYA *et al.*, 2015).

Por fim, conclui-se que o aproveitamento do padrão textual observado em uma base de dados de *reviews* escritos pode trazer resultados positivos mesmo aplicado em um padrão diferente e que não esteja disponível ou classificada para uma tarefa de aprendizado supervisionado.

## REFERÊNCIAS

AFONSO, A. R., A Referenciação em Textos do YouTube: Um Estudo com Vistas à Análise de Sentimentos. in **Liinc em Revista**, Rio de Janeiro, v.13, n.2, p. 357-370, novembro 2017. <http://www.ibict.br/liinc> <http://dx.doi.org/10.18617/liinc.v13i2.3933>.

AFONSO, A. R., DUQUE, C.G., Análise de Sentimentos em Comentários de Vídeos do YouTube Utilizando Aprendizagem de Máquinas Supervisionada, in **Revista Ibict, Ciência da Informação**, Brasília, DF, v.48 n.3, p.21-33, set./dez. 2019. <http://revista.ibict.br/ciinf/article/view/4315>.nn

BARION, E. C. N., LAGO, D., Mineração de Textos, in **REVISTA DE CIÊNCIAS EXATAS E TECNOLOGIA**, vol. III, nº 3, Ano 2008 p.p 123-140, Faculdade Anhaguera. <https://revista.pgskroton.com/index.php/rcext/article/view/2372/2276>.

BECKER, K., TUMITAN, D., Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. Trabalho Final submetido para aprovação no **Programa de Pós-Graduação em Ciência da Computação - Instituto de Informática Universidade Federal do Rio Grande do Sul (UFRGS)**. [http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd\\_versaosubmetida.pdf](http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd_versaosubmetida.pdf).

BENEVENUTO, F., RIBEIRO F., ARAÚJO M., Métodos para Análise de Sentimentos em Mídias Sociais. Artigo apresentado na página do **Departamento de Ciência da Computação da UFMG**. <https://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf>

BRITO, E. M.N., Mineração de Textos: Detecção Automática de Sentimentos em Comentários nas Mídias Sociais, Dissertação apresentada ao Programa de Mestrado em Sistemas de Informação e Gestão do Conhecimento da Universidade Fundação Mineira de Educação e Cultura — **FUMEC**, <http://www.fumec.br/revistas/sigc/article/view/4951>.

BUSSON, A. J. G, FIGUEIREDO, L. C., dos SANTOS, G. P., DAMASCENO, A.L.B., COLCHER, S. E MILIDIÚ, R. L. Desenvolvendo Modelos de Deep Learning para Aplicações Multimídia no Tensorflow, cap. 3. IN: **Anais do XXIV Simpósio Brasileiro de Sistemas Multimídia e Web: Minicursos**, PUC-RIO, outubro de 2018 ([https://www.researchgate.net/publication/328341764\\_Desenvolvendo\\_Modelos\\_de\\_Deep\\_Learning\\_para\\_Aplicacoes\\_Multimidia\\_no\\_Tensorflow](https://www.researchgate.net/publication/328341764_Desenvolvendo_Modelos_de_Deep_Learning_para_Aplicacoes_Multimidia_no_Tensorflow))

CAETANO, M.F, GONÇALVES, R. S.S. TALOU, A.F., Uso de Mineração de Dados e Inteligência Artificial para Classificar Opiniões nas Redes Sociais, in **CADERNO DE ESTUDOS TECNOLÓGICOS** Faculdade de Tecnologia de Bauru (FATEC), v. 6, n. 1 (2019). <http://www.fatecbauru.edu.br/ojs/index.php/CET/article/view/399>.

CHAWLA, S., DUBEY, G., RANA, A., Product opinion mining using sentiment analysis on smartphone reviews. 2017 6th **International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions**, ICRITO 2017, 2018-January, 377–383. <https://doi.org/10.1109/ICRITO.2017.8342455>

CUIZON, J.C., LOPEZ, J., JONES, D.R., Text Mining Customer Reviews for Aspect-Based Restaurant Rating, **International Journal of Computer Science & Information Technology (IJCSIT)** Vol 10, No 6, December 2018. [https://www.researchgate.net/publication/330212354\\_Text\\_Mining\\_Customer\\_Reviews\\_For\\_Aspect-based\\_Restaurant\\_Rating](https://www.researchgate.net/publication/330212354_Text_Mining_Customer_Reviews_For_Aspect-based_Restaurant_Rating).

DOMINIK MAHR; SUSAN STEAD; GABY ODEKERKEN-SCHRÖDER. Making sense of customer service experiences: a text mining review. **Journal of Services Marketing**, [s. l.], v. 33, n. 1, p. 88–103, 2019. DOI 10.1108/JSM-10-2018-0295. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsemr&AN=edsemr.10.1108.JSM.10.2018.0295&lang=pt-br&site=eds-live>.

GONÇALVES, E.C, Mineração de Texto: Conceitos e Aplicações Práticas, in **SQL Magazine**, v. 105, 2012, p. 31-44. [https://www.researchgate.net/publication/317912973\\_Minerao\\_de\\_texto\\_-\\_Conceitos\\_e\\_aplicacoes\\_praticas](https://www.researchgate.net/publication/317912973_Minerao_de_texto_-_Conceitos_e_aplicacoes_praticas).

GOODFELLOW, I., BENGIO, Y., COURVILLE, A., **Deep Learning**, The MIT Press, London, England, 2016, p. 360.

HOLE, V., CHAVAN, M., GAVHANE, T., YADAV, S., An Approach for Sentiment Analysis using Neural Network, **International Journal of Trend in Scientific Research and Development (IJTSRD)**, Volume -2, Issue - 4. <https://www.ijtsrd.com/engineering/computer-engineering/14452/an-approach-for-sentiment-analysis-using-neural-network/varsha-hole>.

Jianmo Ni, Jiacheng Li, Julian McAuley **Empirical Methods in Natural Language Processing (EMNLP)**, 2019 <http://deepyeti.ucsd.edu/jianmo/amazon/index.html>.

JOSHI, K., BHARATHI, H.N., RAO, J., Stock Trend Prediction Using News Sentiment Analysis **International Journal of Computer Science & Information Technology (IJCSIT)** Vol 8, No 3, June 2016. [https://www.researchgate.net/publication/304995235\\_Stock\\_Trend\\_Prediction\\_Using\\_News\\_Sentiment\\_Analysis](https://www.researchgate.net/publication/304995235_Stock_Trend_Prediction_Using_News_Sentiment_Analysis).

KAKKAD, J., MAKWANA, S., SHAH, R., CHACHRA, S., Real Time Predictive Analysis of Indian Stock Market using Machine Learning and Natural Language Processing, IRJCS:: **International Research Journal of Computer Science, Volume VI**,184-188. [https://www.academia.edu/39011906/REAL\\_TIME\\_PREDICTIVE\\_ANALYSIS\\_OF\\_INDIAN\\_STOCK\\_MARKET\\_USING\\_MACHINE\\_LEARNING\\_AND\\_NATURAL\\_LANGUAGE\\_PROCESSING](https://www.academia.edu/39011906/REAL_TIME_PREDICTIVE_ANALYSIS_OF_INDIAN_STOCK_MARKET_USING_MACHINE_LEARNING_AND_NATURAL_LANGUAGE_PROCESSING)

LIANG, T. P., LI, X., YANG, C. T., WANG, M., What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. **International Journal of Electronic Commerce**, 20(2), 236–260, 2015. <https://doi.org/10.1080/10864415.2016.1087823>

MARASANAPALLE, J., VIGNESH T.S, SRINIVASAN P.K., SAHA, A., Business Intelligence From Twitter For The Television Media: A Case Study, [https://www.researchgate.net/publication/251993435\\_Business\\_intelligence\\_from\\_Twitter\\_for\\_the\\_television\\_media\\_A\\_case\\_study](https://www.researchgate.net/publication/251993435_Business_intelligence_from_Twitter_for_the_television_media_A_case_study).

MATOS, F.F., SOUZA, R.R., REIS, Z.S.N., Análise De Dados de Saúde: Mineração de Texto com A Utilização do Orange Canvas para Exploração da Informação, **XX ENANCIB de 21 a 25 de outubro de 2019**. <https://conferencias.ufsc.br/index.php/enancib/2019/paper/viewFile/639/591>.

MORAIS, E. A. M. M., AMBRÓSIO, A.P.L., Mineração de Textos, in **Relatório Técnico** Dezembro - 2007 UNIVERSIDADE FEDERAL DE GOIÁS (UFG) [http://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_005-07.pdf](http://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf).

NIE, B., SUN S., Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research, **Applied Sciences**, April 2017. [https://www.researchgate.net/publication/316986983\\_Using\\_Text\\_Mining\\_Techniques\\_to\\_Identify\\_Research\\_Trends\\_A\\_Case\\_Study\\_of\\_Design\\_Research](https://www.researchgate.net/publication/316986983_Using_Text_Mining_Techniques_to_Identify_Research_Trends_A_Case_Study_of_Design_Research).

PARLAR, T., OZEL, S.A., An Investigation of Term Weighting and Feature Selection Methods for Sentiment Analysis, **Majlesi Journal of Electrical Engineering Vol. 12**, No. 2, June 2018. [https://www.researchgate.net/publication/326069724\\_An\\_investigation\\_of\\_term\\_weighting\\_and\\_feature\\_selection\\_methods\\_for\\_sentiment\\_analysis](https://www.researchgate.net/publication/326069724_An_investigation_of_term_weighting_and_feature_selection_methods_for_sentiment_analysis).

PENNINGTON, J., SOCHER, R., and CHRISTOPHER D. Manning. 2014. GloVe: Global Vectors for Word Representation.

POPESCU, A. M., ETZIONI, O., Extracting Product Features and Opinions from Reviews. [http://turing.cs.washington.edu/papers/emnlp05\\_opine.pdf](http://turing.cs.washington.edu/papers/emnlp05_opine.pdf).

SALLOUM, S.A., MOSTAFA, A., MONEM, A.A., SHAALAN, K., Using Text Mining Techniques for Extracting Information from Research Articles, Chapter in **Studies in Computational Intelligence - January 2018**. [https://www.researchgate.net/publication/321150349\\_Using\\_Text\\_Mining\\_Techniques\\_for\\_Extracting\\_Information\\_from\\_Research\\_Articles](https://www.researchgate.net/publication/321150349_Using_Text_Mining_Techniques_for_Extracting_Information_from_Research_Articles).

SARKAR, D., BALI, R., SHARMA T., Practical Machine Learning with Python. 2018. <https://doi.org/10.1007/978-1-4842-3207-1>

SITUMEANG, F.; DE BOER, N.; ZHANG, A. Looking beyond the stars: A description of text mining technique to extract latent dimensions from online product reviews. **International Journal of Market Research**, [s. l.], v. 62, n. 2, p. 195–215, 2020. DOI 10.1177/1470785319863619.

SULOVA, S., TODORANOVA, L., PENCHEV, B., NACHEVA, R., Using text mining to classify research papers, **17th International Multidisciplinary Scientific GeoConference SGEM 2017, Paper - July 2017**. [https://www.researchgate.net/publication/319688772\\_Using\\_text\\_mining\\_to\\_classify\\_research\\_papers?enrichId=rgreq-78580fa8ed417decd7cf0d61bfbb685d-](https://www.researchgate.net/publication/319688772_Using_text_mining_to_classify_research_papers?enrichId=rgreq-78580fa8ed417decd7cf0d61bfbb685d-)

XXX&enrichSource=Y292ZXJQYWdlOzMxOTY4ODc3MjtBUzo1MzgyMjA3MjAwNzA2NTZAMTUwNTMzMzIyNzg2MQ%3D%3D&el=1\_x\_2&\_esc=publicationCover Pdf.

VIDYA, N. A, FANANY, M. I., BUDI, I., Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers, **Procedia Computer Science**, v. 72, 2015, 519-526. doi: 10.1016/j.procs.2015.12.159.

YOGAPREETHI, N., MAHESWARI, S., A Review on Text Mining in Data Mining, artigo publicacao em **International Journal on Soft Computing (IJSC) Vol.7**, No. 2/3, August 2016, [https://www.researchgate.net/publication/307902719\\_A\\_Review\\_on\\_Text\\_Mining\\_in\\_Data\\_Mining](https://www.researchgate.net/publication/307902719_A_Review_on_Text_Mining_in_Data_Mining).