

UNIVERSIDADE EVANGÉLICA DE GOIÁS - UNIEVANGÉLICA
ENGENHARIA DE SOFTWARE

THYERRE DOS SANTOS ESBALTAR

Desenvolvimento de uma aplicação que analisa sentimentos em frases do Twitter

Anápolis
Dezembro, 2021

UNIVERSIDADE EVANGÉLICA DE GOIÁS - UNIEVANGÉLICA
ENGENHARIA DE SOFTWARE

THYERRE DOS SANTOS ESBALTAR

Desenvolvimento de uma aplicação que analisa sentimentos em frases do
Twitter

Trabalho apresentado ao Curso de Engenharia de Software da Universidade Evangélica de Goiás – UniEVANGÉLICA, da cidade de Anápolis-GO como requisito parcial para obtenção do Grau de Bacharel em Engenharia de Software.

Orientador (a): Prof. Ms. William Pereira dos Santos Junior

Anápolis
Dezembro, 2021

UNIVERSIDADE EVANGÉLICA DE GOIÁS - UNIEVANGÉLICA
ENGENHARIA DE SOFTWARE

THYERRE DOS SANTOS ESBALTAR

Desenvolvimento de uma aplicação que analisa sentimentos em frases do
Twitter

Monografia apresentada para Trabalho de Conclusão de Curso de Engenharia de Software da Universidade Evangélica de Goiás - UniEVANGÉLICA, da cidade de Anápolis-GO como requisito parcial para obtenção do grau de Engenheiro(a) de Software.

Aprovado por:

**William Pereira dos Santos Junior, Mestre, UniEVANGÉLICA
(ORIENTADOR)**

**Natasha Sophie Pereira, Mestre, UniEVANGÉLICA
(AVALIADOR)**

**Pollyana dos Reis Pereira Fanstone, Mestre, UniEVANGÉLICA
(AVALIADOR)**

Anápolis, 06 de dezembro de 2021.

FICHA CATALOGRÁFICA

ESBALTAR, Thyerre dos Santos. **Desenvolvimento de uma aplicação que analisa sentimentos em frases do Twitter**. [Anápolis] 2021. (Universidade Evangélica de Goiás – UniEVANGÉLICA, Engenheiro(a) Software, 2021).

Monografia. Universidade Evangélica de Goiás, Curso de Engenharia de Software, da cidade de Anápolis-GO.

1. *Python; Twitter*; Análise de sentimentos.

REFERÊNCIA BIBLIOGRÁFICA

ESBALTAR, Thyerre dos Santos. **Desenvolvimento de uma aplicação que analisa sentimentos em frases do Twitter**. Anápolis, 2021. 34 p. Monografia - Curso de Engenharia de Software Universidade Evangélica de Goiás - UniEVANGÉLICA.

CESSÃO DE DIREITOS

NOMES DOS AUTORES: Thyerre dos Santos Esbaltar

TÍTULO DO TRABALHO: Desenvolvimento de uma aplicação que analisa sentimentos em frases do Twitter

GRAU/ANO: Graduação/2021

É concedida à Universidade Evangélica de Goiás - UniEVANGÉLICA, permissão para reproduzir cópias deste trabalho, emprestar ou vender tais cópias para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho pode ser reproduzida sem a autorização por escrito do autor.

DocuSigned by:

Thyerre dos Santos Esbaltar

838CDE63768A41C...

Thyerre dos Santos Esbaltar

Anápolis, 06 de dezembro de 2021

AGRADECIMENTOS

A Deus por ter proporcionado a oportunidade de ingressar nesse curso e permitir estar concluindo esse trabalho, sempre me dando forças, coragem e determinação para vencer os obstáculos encontrados.

Aos meus pais, avós e irmã que sempre me incentivaram e auxiliaram nos momentos difíceis.

Ao meu orientador Professor William, que orientou, ensinou e incentivou a aprimorar o meu trabalho.

E todos que de maneira direta ou indireta contribuíram para a conclusão deste trabalho, o meu sincero agradecimento.

RESUMO

É da natureza humana valorizar as opiniões e percebe-se que com o avanço da *internet* o compartilhamento dela tornou-se mais abrangente, sendo capaz de alcançar pessoas de todo o mundo. Essas opiniões que são compartilhadas na *internet* por meio de aplicações de redes sociais, demonstram-se bastante importantes, haja vista que as opiniões têm um amplo alcance e a possibilidade de impactar a marca do comércio eletrônico de forma positiva ou negativa. A fim de auxiliar no entendimento de como está a situação da marca do comércio eletrônico, faz-se necessário a utilização da análise de sentimentos, onde seria realizada uma investigação sobre opinião compartilhada e retornaria a classificação da mesma. Para a produção da aplicação foi analisada e escolhida a linguagem de programação *Python*, através da qual podemos fazer uso de diversas bibliotecas que têm como objetivo auxiliar no desenvolvimento do código, dentre essas bibliotecas podemos citar a Tweepy, preprocessor, Keras, Sklearn, Numpy, Pandas e NLTK. Este estudo busca apresentar a possibilidade de aplicação da análise de sentimentos utilizando o modelo de rede neural recorrente LSTM tendo como objetivo classificar as frases em “satisfeito”, “insatisfeito” e “neutro”. Analisou-se frases destinadas a somente um comércio eletrônico que foram compartilhadas por meio da rede social *Twitter*.

Palavras-chave: *Python*. *Twitter*. Análise de sentimentos.

LISTA DE ILUSTRAÇÕES

Figura 1 - Módulo de repetição de uma Rede Neural Recorrente.....	18
Figura 2 - Modelo de repetição LSTM.....	19
Figura 3 - Fluxograma das etapas do projeto.....	24
Figura 4 - Fluxograma dos passos realizados na segunda etapa.....	25
Figura 5 – Gráfico das emoções presentes nos tweets.....	27

LISTA DE TABELAS

Tabela 01 - Classificação da frase.....	24
---	----

LISTA DE ABREVIATURAS E SIGLAS

Siglas	Descrição
API	<i>Application Programming Interface</i>
LSTM	<i>Long Short Term Memory</i>
NLTK	<i>Natural Language Toolkit</i>
PLN	Processamento de Linguagem Natural
URL	<i>Uniform Resource Locator</i>

SUMÁRIO

INTRODUÇÃO.....	10
1. REFERENCIAL TEÓRICA.....	12
1.1 Redes sociais.....	12
1.2 Análise de sentimentos.....	14
1.3 Processamento de Linguagem Natural.....	14
1.4 Classificadores de sentimentos.....	15
1.4.1 Abordagem de aprendizado de máquina.....	15
1.5 Redes Neurais.....	16
1.6 Redes Neurais Recorrentes.....	17
1.7 Rede <i>Long Short Term Memory</i>	19
1.8 <i>Python</i>	20
1.9 Keras.....	21
1.10 NLTK.....	21
2. METODOLOGIA DA PESQUISA.....	23
3. RESULTADOS.....	27
4. CONCLUSÃO E CONSIDERAÇÕES FINAIS.....	28
REFERÊNCIAS.....	29

INTRODUÇÃO

Com a *Internet*, surgiu um meio no qual as pessoas são capazes de compartilhar opiniões, onde podem ser vistas em qualquer lugar do mundo, com isto cada vez mais pessoas começam a utilizar as redes sociais *on-line*. Segundo dados de 2012 aproximadamente 96% da população *on-line* utilizava redes sociais (MARQUES; PINTO; ALVAREZ, 2016). Com o avanço, a facilidade e a popularidade da *internet*, foram criadas diversas redes sociais, como *Twitter*, Facebook e Instagram.

A rede social *Twitter* é o local onde os usuários podem divulgar todos os tipos de informações em tempo real por meio de mensagens, no ano de 2021 o limite de caracteres que podem ser utilizados são de 280 (NASCIMENTO; OSIEK; XEXÉO, 2013).

O *Twitter* assim como Facebook e Instagram têm sido utilizados como fontes de dados para a análise de sentimentos, isso se deve pela grande quantidade de dados que essas redes sociais movimentam.

O *Twitter* tornou-se uma das maiores redes sociais do mundo, em junho de 2012 cerca de 41.200.000 pessoas produziram cerca de 2,3 bilhões de *tweets* expondo suas opiniões, considerando somente as cidades de Rio de Janeiro e São Paulo (SEMIOCAST, 2012).

De acordo com Antunes (2014), as redes sociais vêm ganhando destaque no setor privado, porque permitem uma melhor comunicação entre as empresas e os clientes possibilitando um melhor levantamento de informações referente a receptividade dos consumidores em relação aos produtos e serviços.

As opiniões são importantes para os humanos, pois muitas vezes com o auxílio delas decidem se irão comer em um determinado restaurante, se irão comprar determinado carro ou em qual comércio eletrônico farão as compras *on-line*. De acordo com Liu (2012), as opiniões são importantes para quase todas as atividades humanas, tendo em vista que são os grandes influenciadores de nosso comportamento. Segundo Becker (2014), empresas baseiam suas estratégias de negócios e *marketing* na opinião de seus clientes em relação aos seus produtos ou serviços.

Opiniões negativas são capazes de prejudicar a reputação de qualquer indivíduo. Para as empresas, esse tipo de opinião demonstra ser igualmente importante ou quiçá mais, haja vista que, tem potencial de influenciar negativamente os consumidores, podendo

ocasionar a perda de um potencial cliente. Segundo Evangelista e Padilha (2014), as empresas já têm a noção de que um comentário desfavorável pode atingir uma imensa quantidade de clientes de forma rápida e dessa maneira afetar a imagem da empresa no mercado.

De acordo com Gomes (2012), algumas empresas também já sabem como tirar vantagens dessas opiniões, compartilhadas em redes sociais, por meio da análise dos sentimentos que foram compartilhadas em diversos canais de comunicação. O resultado dessa análise de sentimento pode possibilitar à empresa um investimento de recursos de maneira eficiente e assertiva.

Levando em consideração a importância das opiniões, a grande utilização das redes sociais e o que isto proporciona para o comércio eletrônico em possibilidades de coleta de opinião sobre a sua marca, podemos levantar a seguinte questão: É possível automatizar a descoberta de opiniões dos clientes em relação a um comércio eletrônico e através de uma classificação apresentar o sentimento geral?

O primeiro passo para a produção deste trabalho será a pesquisa bibliográfica, a qual deverá proporcionar experiência teórica indispensável para a produção deste trabalho e possibilitará a escrita desta monografia.

O segundo passo consiste em identificar métodos e técnicas de mineração de dados direcionados para análise de sentimentos. No terceiro passo, buscaremos algoritmos de aprendizagem de máquina que possam ser utilizados para análise de sentimentos. Por fim iremos desenvolver uma aplicação que fará uma análise de sentimentos nos dados coletados da rede social *Twitter* com a função de demonstrar o ponto de vista dos clientes em relação ao comércio eletrônico.

Na seção 1 é apresentado o referencial teórico no qual foram abordados os seguintes tópicos: Redes Sociais, Análise de Sentimentos, Processamento de Linguagem Natural, Classificação de Sentimentos, Abordagem de aprendizado de máquina, Redes Neurais, Redes Neurais Recorrentes, Rede *Long Short Term Memory*, *Python*, *Keras* e *NLTK*.

Na seção 2 é exposto as etapas necessárias para o desenvolvimento da aplicação. Na seção 3 será apresentado os resultados alcançados e na seção 4 estão as considerações finais.

1. REFERENCIAL TEÓRICO

Neste tópico serão apresentados conceitos sobre redes sociais, análise de sentimentos, aprendizagem de máquina, processamento de linguagem natural. Também será apresentado algumas das principais tecnologias que serão utilizadas durante o desenvolvimento.

1.1 Redes sociais

As redes sociais são aplicações que se fazem amplamente utilizadas por pessoas com acesso à *internet*. Conforme dito por Marteleto (2001), as redes sociais representam “[...] um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados”. Para Souza e Quandt (2008), as redes sociais são estruturas complexas e dinâmicas constituídas por pessoas com princípios ou com um propósito em comum.

As redes sociais como o Instagram, *Twitter* e Facebook foram usados como fontes de dados para diversos trabalhos relacionados à análise de sentimento como na previsão de resultados eleitorais e qualificar campanhas publicitárias, na mensuração da satisfação dos consumidores e na aceitabilidade de certo produto no mercado, análise de sentimentos e na predição de crises em usuários acometidos por distúrbios maníaco-depressivo (LIMIRO; CORDEIRO; CASSIANO, 2018).

Segundo Recuero (2009), a rede social pode ser definida como um grupo de dois elementos, sendo atores que consistem em pessoas, grupos ou instituições e suas conexões que seriam os laços sociais ou interações.

De acordo com Kietzmann et al. (2011), as redes sociais têm as seguintes características:

- Identidade: representa o local onde os usuários revelam suas identidades no ambiente das redes sociais;
- Conversas: representam a extensão da comunicação entre os usuários na rede social;

- Compartilhamento: representa a possibilidade dos usuários de distribuírem e receberem conteúdo;
- Presença: trata-se dos outros usuários poderem notar que alguém se encontra *on-line* na rede social;
- Relações: representa os relacionamentos de um usuário para com outro usuário;
- Reputação: seria a oportunidade de um usuário checar a sua popularidade e a de outros usuários nas redes sociais, por meio de quantidade de seguidores, *likes*, postagem de fotos e vídeos;
- Grupos: representa a possibilidade da formação de comunidades e grupos nas redes sociais.

Essas características das redes sociais proporcionaram-na sua expansão, tornando-se um meio de compartilhamento de diversos tipos de opiniões e conteúdos, a formação de diversas comunidades e grupos com vários tipos de temas e objetivos.

As redes sociais produzem diariamente um enorme volume de dados, isto ocorre devido a sua grande quantidade de usuários e ao fato de ser uma aplicação a qual o compartilhamento de conteúdo é uma de suas principais funções. Conforme foi dito no Simpósio Brasileiro de Banco de Dados (2014), “A popularidade dessas plataformas pode ser evidenciada através da capacidade que possuem de produzir enormes volumes de conteúdo”.

O *Twitter* é uma rede social que comporta diferentes tipos de pessoas, ideias, pontos de vista e informações. O *Twitter* foi criado em março de 2006, sendo uma rede social gratuita que permite o compartilhamento de diversos conteúdos.

Barbosa et al. (2012) afirma que o padrão de interação do *Twitter* faz com que o usuário sinta a necessidade de compartilhar e expressar de maneira contínua sua opinião e sentimento. Por ser em tempo real, fácil e rápido publicar no *Twitter*, os usuários interagem mais com a rede social (ZHAO; ROSSON, 2009).

O *Twitter* é utilizado por seu usuário não apenas como uma ferramenta que tem o único propósito de divulgar informações pessoais, mas também o empregam para compartilhar opiniões e conhecimentos sobre eventos e fatos em geral (NAAMAN; BOASE, 2010). Conforme Java et al., (2007), o *Twitter* é uma ferramenta de grande

importância para dois tipos de usuário: os que procuram informação e os que querem compartilhar amplamente informações.

1.2 Análise de sentimentos

A utilização da análise de sentimentos, também conhecida como mineração de opinião, visa a captação da emoção predominante do seu público-alvo. A análise de sentimentos é uma ferramenta que se tornou valiosa para empresas, políticos, influenciadores e entre outros. “Empresas, eventos (e.g. Olimpíadas), personalidades, estão interessadas na compreensão de como são percebidas pelo público em geral em tempo real, e nas mais variadas mídias” (BECKER, 2013).

De acordo com o autor Liu (2012), a análise de sentimentos é uma área de estudo que tem como objetivo a análise das emoções das pessoas, avaliações, atitudes, emoções e sentimentos relacionados a entidades como organizações, serviços, produtos, indivíduos, eventos, questões, tópicos e seus atributos.

A análise de sentimentos tem como fundamental objetivo classificar, por intermédio da apuração das palavras e de maneira automática, os sentimentos que foram expressados no texto. De acordo com Medhat, Hassen e Korashy (2014), o propósito da análise de sentimentos é obter opiniões, identificar os sentimentos expressados nela e classificar sua polaridade.

A utilização da análise de sentimentos implementada para analisar frases que são postadas em redes sociais proporcionam uma forma automatizada de averiguar o ponto de vista do cliente para com a empresa. A utilização desses dados pode amparar as tomadas de decisões, fazendo com que essas sejam mais eficientes e detenham maiores chances de serem bem-sucedidas.

1.3 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) trata-se de um ramo da pesquisa da inteligência artificial que busca maneiras de como computadores poderiam ser utilizados para manipular e entender a linguagem natural. Conforme Allen (2003), o desafio básico

enfrentado em PLN refere-se à ambiguidade que está presente na linguagem humana. Brascher (2002) salienta motivos para tais ambiguidades na linguagem natural:

- Ambiguidade lexical: ocorre quando não se é possível interpretar o significado da unidade lexical. Exemplo da palavra “banco”, a palavra pode se referir ao banco de uma praça ou a instituição bancária.
- Ambiguidade semântica: ocorre quando se tem várias interpretações possíveis para o relacionamento dos termos na frase. Exemplo: "Um rio corre através de cada país europeu." A frase significa que apenas um rio corre através de cada país europeu, ou vários rios correm por entre diferentes países europeus.
- Ambiguidade pragmática: se relaciona ao que está ligado à situação do locutor no momento da enunciação. Exemplo na frase “Paulo vai à escola.” o sujeito seria um estudante ou está apenas indo à escola naquele momento?
- Ambiguidade predicativa: se dá por meio da interpretação das relações temáticas que encadeiam o predicado, participantes e argumentos. Exemplo na frase “A crítica deste autor.” o autor seria o objeto da crítica, ou seria o agente da crítica.
- Ambiguidade sintática: se dá no momento da estruturação das frases em constituintes hierarquizados. Exemplo a frase “A professora de dança espanhola.” a professora é espanhola ou da dança que é espanhola.

1.4 Classificadores de sentimentos

Há diversas formas de classificar sentimentos em textos. Segundo Maynard e Funk (2011), uma técnica que pode ser utilizada para a detecção de sentimentos é a abordagem de aprendizagem de máquina. Nos subtópicos abaixo discorreremos sobre a abordagem de aprendizagem de máquina e sobre um de seus classificadores.

1.4.1 Abordagem de aprendizado de máquina

Na abordagem de aprendizagem de máquina, estando no escopo da análise de sentimentos, é desenvolvido um modelo de classificação onde uma mensagem está associada a uma polaridade ou sentimento, na seguinte frase “Comprei uma TV e chegu

voando! Tenho muita simpatia por essa loja” o sentimento associado ao comércio eletrônico é de satisfação.

Os dados que utilizamos para a produção deste modelo são chamados de dados de treinamento e normalmente são formados por uma mensagem e cada uma está vinculada a uma classe, podendo ser um sentimento ou polaridade (*COMPUTER ON THE BEACH*, 2018).

Podemos dividir os métodos de classificação que utilizam a abordagem de aprendizagem de máquina, dentro do escopo da análise de sentimentos, em métodos de aprendizagem supervisionados e não supervisionados (MEDHAT; HASSAN; KORASHY, 2014).

A abordagem de aprendizagem supervisionada necessita de uma coleção prévia de treinamento, ou seja, depende de uma numerosa quantidade de documentos classificados. Existem diversas técnicas que podem ser utilizadas em aprendizagem supervisionada, no contexto da análise de sentimentos (MEDHAT; HASSAN; KORASHY, 2014). este trabalho será abordado apenas a técnica de Rede Neural.

Conforme citado no evento *Computer on the Beach* (2018), o método não supervisionado não necessita da coleção prévia de treinamento, o método utiliza uma coleção de palavras, dicionários ou estatística para a classificação. Este método tem como principal vantagem a independência do domínio, isto é, este modelo não tem como base um conjunto limitado de treinamento (SONI; PATEL, 2014).

1.5 Redes Neurais

As Redes Neurais, conhecidas também como Redes Neurais Artificiais, são uma estrutura de computação que procuram trabalhar de maneira semelhante à dos neurônios biológicos. Segundo Norvig e Russell (2014) a inspiração para a criação das Redes Neurais foi a aptidão do cérebro humano para processar informações.

Para TAFNER et al (1996) as Redes Neurais podem ser definidas como “uma rede massivamente paralela de elementos interconectados e suas organizações hierárquicas que

estão preparadas para iterar com objetos do mundo real do mesmo modo que um sistema nervoso biológico faz”.

As Redes Neurais são modelos matemáticos não-lineares que se baseiam no funcionamento de neurônios, as Redes Neurais podem ser utilizadas como um classificador (HAYKIN, 1994). Segundo Zhang, Patuwo e Hu (1997) as Redes Neurais apresentam boa capacidade de reconhecimento de padrões e de classificação.

Segundo Barone (1999), as Redes Neurais têm como principais características:

- a capacidade de adquirir conhecimento por meio de exemplos e conseguir generalizar este conhecimento de tal forma que possa reconhecer elementos similares, que não se encontram dentro do grupo de exemplos inicial;
- tem um bom desempenho em tarefas que são pouco ou mal definidas, onde haja falta do conhecimento explícito relacionado à como resolvê-las;
- robustez à presença de dados ausentes ou falsos;
- estando no escopo da classificação de padrões, a Rede Neural é capaz de fornecer informações relacionada a quais padrões se deve selecionar levando em conta a função do grau de confiança que foi apresentado;
- tolerância à falha.

Araújo, Vittorazzi e Junior (2018), afirmam que as Redes Neurais impulsionaram os desenvolvimentos de metodologias e estudos que possibilitaram superar certas adversidades impostas pela PLN.

1.6 Redes Neurais Recorrentes

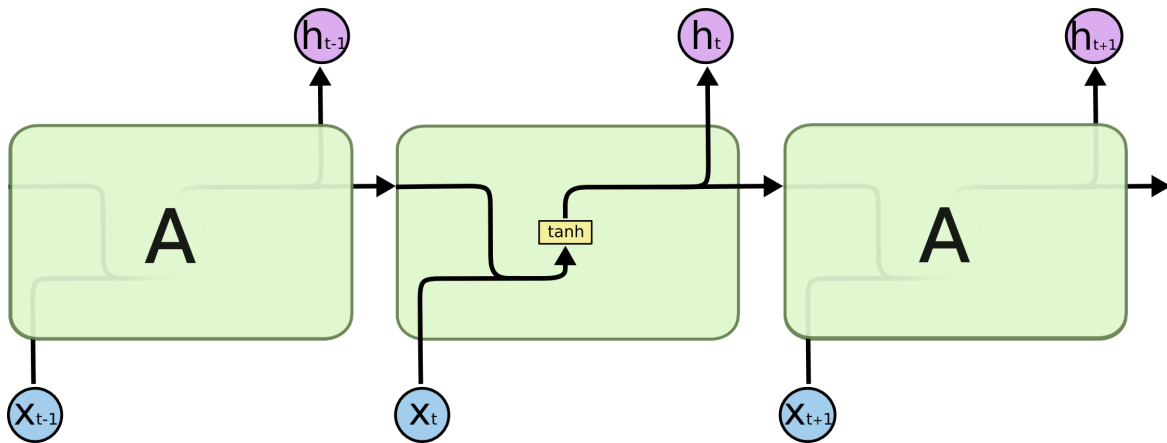
As Redes Neurais Recorrentes são um tipo de Rede Neural que diminui a complexidade de implementação e o tempo de treinamento (TOSCANO; LOBO, 2006). As Redes Neurais Recorrentes mostram-se ser algoritmos de redes neurais bastante poderosos no processamento de dados sequenciais como dados de séries temporais, som e linguagem natural (ACADEMY, 2019).

As Redes Neurais Recorrentes foram utilizadas para processamento de dados sequenciais de domínios distintos como na música, na captura em movimento e texto

(BOULANGER-LEWANDOWSKI et al., 2012; ECK; SCHMIDHUBER, 2002; SUTSKEVER et al., 2009; SUTSKEVER et al., 2011).

A Rede Neural Recorrente é estruturada de maneira que tenha um loop onde é passada a informação de uma camada a outra, desta maneira permite que a informação seja propagada através da rede (ARAÚJO; VITTORAZZI; JUNIOR, 2018). Na figura 1 podemos observar um módulo de repetição de uma Rede Neural Recorrente.

Figura 1 - Módulo de repetição de uma Rede Neural Recorrente



Fonte: (OLAH, 2015).

As Redes Neurais Recorrentes recebem a entrada de dados de duas fontes, sendo elas uma do presente quanto a outra é do passado. Por causa desta característica das Redes Neurais Recorrentes, é possibilitado o reconhecimento do contexto através da memória. De acordo com Zaremba, Sutskever e Vinyals (2015) às Redes Neurais Recorrentes são as que mais se assemelham à maneira como os humanos processam as informações.

Segundo Toscano e Cabral (2006) as Redes Neurais Recorrentes apresentam um aprendizado rápido precisando de poucas iterações de treinamento, oferecem melhor generalização e necessitam de menos parâmetros.

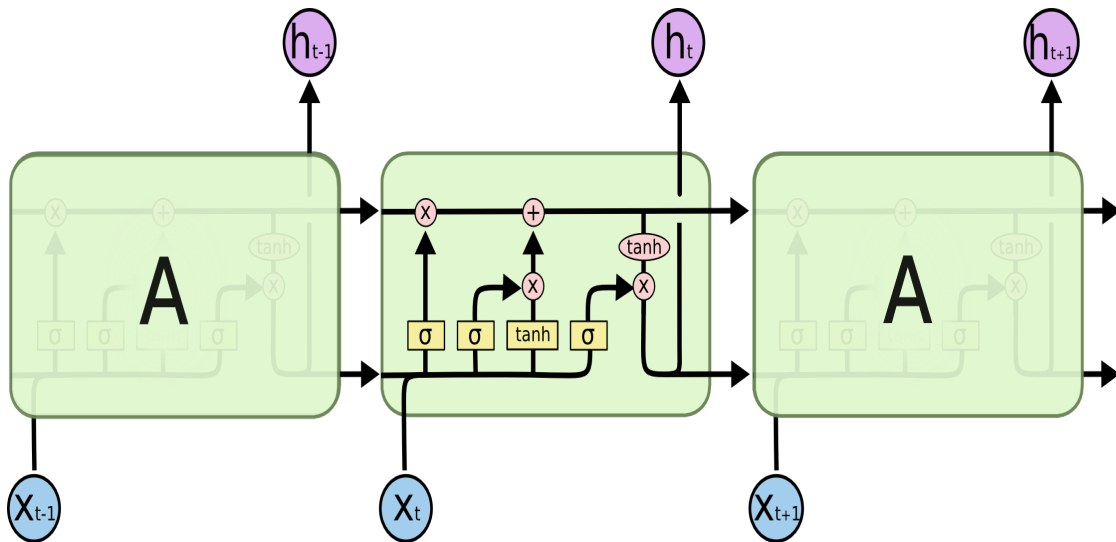
Esta Rede Neural possui um problema que é o gradiente da função de custo declina exponencialmente com o passar do tempo, este problema faz com que o modelo para de aprender. Esse problema é nomeado por Haykin (2008) como problema da dissipação do gradiente.

1.7 Rede *Long Short Term Memory*

O modelo LSTM é uma espécie de Rede Neural Recorrente que é capaz de adquirir conhecimento de dados do passado (KUNAS; HECK; PADOIN, 2021). De acordo com Hochreiter e Schmidhuber (1997) as LSTMs surgiram com o propósito de solucionar o problema da dissipação do gradiente. Para solucionar o problema a LSTM separou a memória da representação de saída, fazendo com que a unidade de memória atual seja dependente linearmente da unidade anterior (Ravuri; Stolcke, 2015).

Para Araújo, Vittorazzi e Junior (2018) as LSTMs não se diferem tanto das Redes Neurais Recorrentes tradicionais, quando observamos a sua estrutura, isso faz com que a maior divergência entre elas ocorra no número de camadas das redes neurais no modelo de repetição, onde as LSTMs dispõem de quatro camadas enquanto que as redes tradicionais possuem apenas uma. Podemos observar o modelo de repetição de uma LSTM na figura 2.

Figura 2 - Módulo de repetição LSTM



Fonte: (OLAH, 2015).

Temos o estado da célula sendo representado pela linha horizontal, por onde as informações fluirão. Na primeira etapa será decidido quais informações deverão ser descartadas do estado da célula, esta decisão é tomada pela camada de entrada ou camada sigmóide. Na camada de sigmóide são produzidos números entre zero e um, os valores

com zero deverão ser descartados e os com o valor um permanecem. Na segunda etapa será definido quais as informações devem ser guardadas no estado da célula, para que isto aconteça, a camada sigmóide determina quais os valores deverão ser atualizados, em seguida à camada tanh, tangente hiperbólica, produz um vetor que terá os possíveis candidatos de serem adicionados ao estado. Na terceira etapa é o que será enviado a saída.

O resultado terá como base o estado da célula no qual a camada sigmóide decidiu quais dados seriam produzidos, pondo a informação através de uma função tanh, define valores entre 1 e -1, e o multiplica pela saída da camada sigmóide, produzindo apenas o que precisamos (HOCHREITER; SCHMIDHUBER, 1997).

Vale ressaltar que nem todas as LSTMs seguem o modelo que foi abordado (OLAH, 2015).

1.8 Python

A linguagem de programação *Python* é uma linguagem de altíssimo nível, de tipagem dinâmica e forte, interativa e interpretada e orientada a objeto (BORGES, 2010). Segundo Borges (2014), possui uma sintaxe clara e concisa, favorecendo a legibilidade do código fonte e com isso faz com que o desenvolvimento utilizando esta linguagem seja mais produtivo. O *Python* foi criado por Guido van Rossum em 1991, tendo como origem de seu nome a série humorística britânica *Monty Python's Flying Circus* (SANTANA; GALESI, 2010). Atualmente tem conquistado bastante espaço entre as outras ferramentas de programação, por ter uma interface "amigável" de fácil aprendizagem e pela sua grande aplicabilidade.

De acordo com o TIOBE - *The Software Quality Company* (2021) o *Python* é a segunda linguagem mais popular dentre os desenvolvedores no mês de maio de 2021, além de ser uma linguagem popular é utilizada por grandes empresas de tecnologia, como a *Google, Microsoft, Instagram, Spotify*. *Python* é deveras popular na comunidade científica e bastante utilizado na análise de dados (CAELUM, 20-?). Bird, Klein e Loper (2009) afirmaram que *Python* é uma linguagem que dispõe de excelentes funcionalidades destinadas ao processo de dados linguísticos.

1.9 Keras

Keras é um *framework* de *deep learning* para o Python que dispõe de uma maneira cômoda de treinar e definir qualquer modelo de *deep learning*. Keras foi desenvolvido inicialmente para pesquisadores, tendo como objetivo viabilizar uma experimentação mais rápida. Grandes organizações fazem uso do Keras, podemos citar o Google, Netflix, CERN, Yelp, NASA, Uber (CHOLLET, 20--?).

1.10 NLTK

O pacote *Natural Language Toolkit* (NLTK) tem como função realizar o processamento de linguagem natural em *Python*, foi criado em 2001 como parte do curso de Linguística Computacional no Departamento de Ciência da Computação e informação da Universidade da Pensilvânia e desde então foi utilizado em cursos de diversas universidades (BIRD; KLEIN; LOPER, 2009).

De acordo com Garg (2016), o NLTK dispõe de diversas funções que são utilizadas no pré-processamento dos dados a fim de que se tornem próprios para a extração e a mineração de recursos.

Conforme Bird, Klein e Loper (2009), o NLTK tinha como foco atingir quatro objetivos principais quando foi projetado:

- Simplicidade: este objetivo busca fornecer uma estrutura que seja intuitiva, fornecendo ao usuário, de forma que não se torne entediante, conhecimento prático de PLN.
- Consistência: procura fornecer uma ferramenta que seja uniforme com interfaces e estruturas de dados consistentes e com métodos que tenham nomenclatura clara, ou seja, fácil de adivinhar a funcionalidade do método.
- Extensibilidade: busca fornecer uma estrutura em que novos módulos do *software* possam ser bem acomodados, onde haja a possibilidade de incluir em uma tarefa novas implementações e abordagens concorrentes.

- Modularidade: este objetivo procura fornecer componentes que são capazes de serem utilizados de forma independente, excluindo assim a necessidade de entender todo o *kit* de ferramentas.

2. METODOLOGIA DA PESQUISA

Tendo como origem a pesquisa bibliográfica para a composição deste trabalho, foi levantado, descoberto e analisado conhecimentos e informações antecipadas de extração de conhecimento sobre documentos textuais, buscando uma resposta ou uma contribuição na resolução do problema (VILAÇA, 2010).

Por meio de pesquisas realizadas, levantamos alguns trabalhos que possuem relação com a análise de sentimentos no *Twitter*, contendo problemas parecidos com esse estudo, porém a técnica que foi empregada no presente trabalho para analisar o sentimento será distinta dos trabalhos que mencionaremos.

Limiro, Cordeiro e Cassiano (2018) tiveram como objetivo analisar sentimentos em cima de *tweets*, no contexto da Copa do Mundo de Futebol. Neste trabalho foi utilizado a *API* de análise de sentimentos da *Microsoft Azure*, os sentimentos foram divididos em duas classes: 0, negativos, 1, positivos.

Evangelista e Padilha (2014) possuíam como objetivo desenvolver uma ferramenta que classificaria publicações *web* oriundas de redes sociais, sendo elas *Facebook* e *Twitter*. Nesse trabalho foram utilizados o algoritmo Naive Bayes e o recurso léxico SentiWordNet, foi trabalhado com três classes de sentimentos sendo elas positivo, negativo e neutro.

No trabalho de Gulli e Pal (2017) é apresentado diversos tipos de redes neurais, além de que, é mostrado a codificação dessas redes neurais em *Python* utilizando Keras, entre elas há um algoritmo de LSTM, utilizando Keras, que tem como objetivo analisar os sentimentos das frases e classificá-las, onde eram classificadas como positivo e negativo.

Para o início deste projeto fez-se necessário a identificação das técnicas que serão utilizadas para a análise dos sentimentos, junto com a identificação destas técnicas é necessário levantar o algoritmo que será utilizado para realizar a análise dos sentimentos.

A seguir temos como objetivo produzir a aplicação que deve coletar os dados e analisá-los. Na figura 3 apresentamos de forma superficial as etapas que iremos abordar no decorrer deste trabalho.

Figura 3 - Fluxograma das etapas do projeto

Fonte: Autor.

A primeira etapa é a da coleta, nesta etapa coletamos os dados que foram utilizados para a produção da base de dados, além de produzir um meio eficiente de coletar os dados do *Twitter*.

A base de dados foi desenvolvida utilizando apenas frases coletadas do *Twitter* e a base está salva no formato .csv, dentro deste arquivo há apenas as frases e as classificações. As frases foram classificadas como “neutro”, “insatisfeito” e “satisfeito”. A classificação “satisfeito” e “insatisfeitos” é atribuída às frases que apresentam opiniões subjetivas relacionadas ao comércio eletrônico, enquanto que, a classificação “neutro” é imposta as frases que expressem opiniões objetivas. A classificação das frases foi feita de maneira manual, ou seja, cada frase foi analisada e classificada por um ser humano, um exemplo de classificação das frases é mostrado na tabela 01.

Tabela 01 - Classificação da frase

Frase	Classe
‘tem o pior atendimento que eu ja vi, meu Deus do ceu que nervoso’	insatisfeito
‘Obrigado pela entrega rapida.’	satisfeito
‘Tem nenhum mouse top e teclado com desconto ai n?’	neutro

Fonte: Elaborado pelo Autor (2021).

Para extrair as frases do *Twitter* utilizamos a biblioteca *Tweepy* e com o auxílio dela tivemos acesso a API do *Twitter*, o código para executar essa tarefa foi desenvolvido na linguagem de programação *Python* e teve como objetivo coletar apenas frases do *Twitter* relacionadas ao comércio eletrônico.

Antes de coletar os dados fazendo uso do *Tweepy*, faz-se necessário estar cadastrado no Twitter e a conta está vinculada com o *Twitter Developer*. Esta vinculação é necessária pois carecemos de chaves, que são fornecidas pelo *Twitter Developer*, para que as requisições do acesso a API do Twitter sejam bem-sucedidas.

Na segunda etapa, desenvolvemos a aplicação que analisará os sentimentos. A aplicação foi desenvolvida na linguagem de programação *Python*, utilizou-se as seguintes bibliotecas para auxiliar no desenvolvimento: Keras, Sklearn, Numpy, Preprocessor, Pandas e NLTK. Na figura 4 abordamos os passos para a codificação.

Figura 4 - Fluxograma dos passos realizados na segunda etapa



Fonte: Autor.

O primeiro passo consiste em limpar as frases da base de dados, nesse passo buscamos retirar ou substituir das frases elementos que possam atrapalhar na classificação, como caracteres especiais e *URLs*, para isto iremos fazer uso da biblioteca Preprocessor. Nesta parte utilizamos também a biblioteca NLTK com o objetivo de retirar as *stopwords*. As *stopwords* são palavras irrelevantes para a análise de sentimentos como: as, os, e, de e com.

O segundo passo foi dividir a base de dados em base de treinamento e base de teste. A base de treinamento contém 80% das frases presentes na base de dados, enquanto que a base de teste tem 20%.

No terceiro passo construiu-se o modelo de rede neural LSTM na linguagem de programação *Python*, para isso foi utilizado o framework Keras. A criação deste modelo teve como base o algoritmo apresentado por Gulli e Pal (2017), porém houve modificações no algoritmo para melhor se encaixar no problema que buscamos solucionar.

Na terceira e última etapa iremos realizar a análise de sentimentos, para isto implementou-se a ideia de coleta de tweets que foi apresentada na primeira etapa, esses

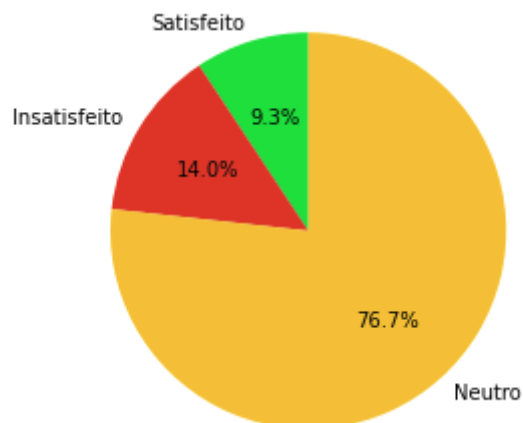
tweets que foram coletados neste momento não são salvos. Por meio do algoritmo que foi desenvolvido utilizando Keras as frases dos tweets que foram coletadas são classificadas como “satisfeito”, “insatisfeito” e “neutro”. Para apresentar o resultado da classificação utilizamos a biblioteca matplotlib, essa biblioteca foi utilizada com o objetivo de produzir um gráfico de pizza referente aos valores resultantes da classificação dos tweets.

3. RESULTADOS

Após a codificação da aplicação ser finalizada foi verificada a acurácia, precisão, do código. Tivemos como resultado uma acurácia de 80,58%, no código apresentado pelos autores Gulli e Pal (2017) conseguiu alcançar uma acurácia de 99%. No entanto, devemos levar em conta que a base de dados produzida por Gulli e Pal (2017) continha aproximadamente 7.000 frases, enquanto que, a base de dados que utilizamos continha apenas 1208 frases. A base de dados utilizada continha 455 frases classificadas como “neutro”, 399 classificadas como “insatisfeito” e 354 classificadas como “satisfeito”.

Tendo executado a aplicação, foram analisados *tweets* postados dos dias 10 de novembro de 2021 até o dia 11 de novembro de 2021 que estavam relacionados a um comércio eletrônico obtivemos como retorno o gráfico apresentado na figura 5.

Figura 5 – Gráfico das emoções presentes nos tweets



Fonte: Autor

Foram analisados 86 *tweets*, onde pode-se notar que a maioria deles expressava um sentimento neutro, ou seja, aproximadamente 76,7% das frases analisadas expressavam opiniões objetivas, enquanto que, aproximadamente 9,3% dos usuários que expressaram sua opinião estava satisfeito com o comércio eletrônico, cerca de 14,0% encontravam-se insatisfeitos com o comércio eletrônico.

4. CONCLUSÃO E CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo explicitar uma maneira de descobrir as opiniões dos clientes as quais foram expressas em uma rede social. Para realizar a análise de sentimentos das frases que eram compartilhadas no *Twitter* foram elaboradas e executadas várias etapas.

A fim de coletar as frases foi desenvolvido um script em Python que tinha como objetivo se comunicar com a API do Twitter, com este script foi facilitada a coleta de dados, a etapa de criação da base de dados mostrou-se ser a mais trabalhosa devido a classificação das frases ter sido feita de maneira manual.

O modelo de classificação que implementamos mostrou-se ser interessante devido a acurácia que foi apresentada, pois mesmo possuindo uma base de dados pequena se comparada à utilizada por Gulli e Pal (2017) não apresentou uma baixa acurácia.

Com a execução do código pode-se observar que a questão levantada: “É possível automatizar a descoberta de opiniões dos clientes em relação a um comércio eletrônico e através de uma classificação apresentar o sentimento geral?”, foi respondida, visto que após a execução de todas as etapas foi produzido uma aplicação que analisa e apresenta, através de gráfico, o sentimento geral dos clientes em relação ao comércio eletrônico.

Para a continuação deste trabalho aconselha-se algumas atividades: A utilização de uma base de dados mais robusta; A aplicação de outras técnicas de aprendizagem de máquina para análise de sentimentos; coletar a opinião dos clientes utilizando outros locais de compartilhamento de opinião.

REFERÊNCIAS

- ACADEMY, D. S. **Deep Learning Book**. 2019. Disponível em: Acesso em: 14 de setembro de 2021.
- ANTUNES, M. N.; SILVA, C. H.; GUIMARÃES, M. C. S.; RABAÇO, M. H. L. Monitoramento de informação em mídias sociais: o e-Monitor Dengue. **TransInformação**, v. 26, n. 1, p. 9-18, 2014.
- ALLEN, J. F. Natural language processing. John Wiley and Sons Ltd., 2003.
- ARAÚJO, Germano Renner de Oliveira; VITTORAZZI, William de Oliveira; JUNIOR, Florisvaldo Cardozo Bomfim. A aplicação de redes neurais artificiais recorrentes no processamento de linguagem natural. **ENTEC - Encontro de Tecnologia**, [s. l.], 2018.
- BARBOSA, Alexandre N. **Descoberta de Conhecimento Aplicado à Base de Dados Textual de Saúde**. São Leopoldo, 2012.
- BARONE, Dante Augusto Couto. **Projeto Revox**. [s. l.], 1999.
- BECKER, Karin; TUMITAN, Diego. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. [S.l.: s.n.], [s. l.], 20-?.
- BIRD, Stven; KLEIN, Ewan; LOPER, Edward. *Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit*. 1ª. ed. 2009. 479 p.
- BORGES, Luis E. **Python para desenvolvedores**. 2a. ed. Rio de Janeiro: Edição do Autor, 2010.
- BORGES, L. E. **Python para desenvolvedores**. Rio de Janeiro: Novatec, 2014, p. 27 - 29.
- BOULANGER-LEWANDOWSKI, N.; BENGIO, Y.; VINCENT, P. **Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription**. [S. l.: s. n.], 2012.
- BRASCHER, M. **A ambiqüidade na recuperação da informação**. IASI, 2002.
- CAELUM. **PYTHON E ORIENTAÇÃO A OBJETOS**. [S. l.], 20-?. Disponível em: https://www.academia.edu/257818/Metodologia_De_Análise_De_Redes_Sociais. Acesso em: 9 maio 2021.

CHOLLET, François. **Deep Learning with Python**. [S. l.: s. n.], 20--?. ISBN 9781617294433.

COMPUTER ON THE BEACH, 2018, Florianópolis. **Análise de Sentimento em Redes Sociais no Idioma Português com Base em Mensagens do Twitter** [...]. [S. l.: s. n.], 2018. Disponível em: <https://www.aclweb.org/anthology/J09-3003/>. Acesso em: 18 fev. 2021.

ECK, D.; SCHMIDHUBER, J. A first look at music composition using lstm recurrent neural networks. **Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale**, v. 103, 2002.

EVANGELISTA, T. ; PADILHA, T. P. P. . **Monitoramento de Posts Sobre Empresas de E-Commerce em Redes Sociais Utilizando Análise de Sentimentos**. III *Brazilian Workshop on Social Network Analysis and Mining* (BraSNAM 2014), 2014, Brasília.

GARG, Prateek. *Sentimental Analysis of Twitter Data using Python NLTK*. 2016. *Thesis (Master of Technology in Computer Science and Applications)* - Thapar University, [S. l.], 2016. Disponível em: <http://tudr.thapar.edu:8080/jspui/bitstream/10266/4273/4/4273.pdf>. Acesso em: 10 maio 2021.

GOMES, Helder Joaquim Carvalheira. **TEXT MINING: ANÁLISE DE SENTIMENTOS NA CLASSIFICAÇÃO DE NOTÍCIAS**. 2012. Trabalho de conclusão de curso (Mestre em Estatística e Gestão de informação, Especialização em Gestão do Conhecimento e Business Intelligence) - Instituto Superior de Estatística e Gestão de Informação Universidade Nova de Lisboa. Disponível em: <https://run.unl.pt/bitstream/10362/9182/1/TEGI0325.pdf>. Acesso em: 30 mar. 2021.

GULLI, Antonio; PAL, Sujit. **Deep Learning with Keras: Implement neural networks with Keras on Theano and TensorFlow**. [S. l.: s. n.], 2017. ISBN 978-1-78712-842-2.

HAYKIN, S. *Neural networks: a comprehensive foundation*. [S.l.]: Prentice Hall PTR, 1994.

HAYKIN, S. *Neural Networks and Learning Machines. Third edition. New York: Pearson Education*, 2008.

HOCHREITER, Sepp; SCHMIDHUBER, Jurgen. Long Short-Term Memory. **Neural Computation**, [S. l.], p. 1735-1780, 31 dez. 1997.

JAVA, Akshay et al. *Why We Twitter: Understanding Microblogging Usage and Communities*. **Joint 9th WEBKDD and 1st SNA-KDD Workshop**, [s. l.], 2007. Disponível

em:

https://www.researchgate.net/publication/234786863_Why_we_Twitter_Understanding_microblogging_usage_and_communities. Acesso em: 21 abr. 2021.

KIETZMANN, et. Al. ***Social media? Get serious! Understanding the functional building blocks of social media.*** Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0007681311000061> >. Acesso em 9 maio. 2021.

KÜNAS, Cristiano; HECK, Leandro; PADOIN, Edson. Desempenho de Modelos de Redes Neurais Recorrentes para Análise de Sentimentos. **ESCOLA REGIONAL DE ALTO DESEMPENHO DA REGIÃO SUL (ERAD-RS)**, 21. , 2021, Evento Online. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 33-36. ISSN 2595-4164.

LIMIRO, Renata Moreira; CORDEIRO, Douglas Farias; CASSIANO, Kátia Kelvis. Análise de Sentimentos em Tweets:: um estudo de caso da Copa do Mundo de Futebol. **Coloquio para apresentação de produções acadêmicas**, [s. l.], 2018.

LIU, B. ***Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies***, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.

MARQUES, Teresa P.; PINTO, Alexandra Marques; ALVEZ, Maria João. Estudo psicométrico da Escala de Avaliação dos Riscos e Oportunidades dos jovens utilizadores do Facebook. **Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica.**, [s. l.], 2016. Disponível em: <https://www.aidep.org/sites/default/files/articles/R41/Art12.pdf>. Acesso em: 28 mar. 2021.

MARTELETO, Regina Maria. Análise de redes sociais: aplicação nos estudos de transferência da informação. **Ciência da informação**, v. 30, n. 1, p. 71-81, 2001.

MAYNARD, D.; FUNK, A. ***Automatic detection of political opinions in tweets.*** In: SPRINGER. *The semantic web: ESWC 2011 workshops*. [S.l.], 2011. p. 88–99.

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

NAAMAN, Mor; BOASE, Jeffrey; LAI, Chih-hui. ***Is it Really About Me? Message Content in Social Awareness Streams.*** [S. l.: s. n.], 2010. Disponível em: https://www.academia.edu/5446826/Is_it_really_about_me_message_content_in_social_awareness_streams. Acesso em: 7 abr. 2021.

NASCIMENTO, Paula; OSIEK, Bruno Adam; XEXÉO, Geraldo. Análise de sentimento de tweets com foco em notícias. **Revista Eletrônica de Sistemas de Informação**, [s. l.], 2015. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/1525/pdf>. Acesso em: 7 abr. 2021.

NORVIG, P.; RUSSELL, S. **Inteligência Artificial: Tradução da 3ª Edição**. [S.l.]: Elsevier Brasil, 2014. v. 1.

OLAH, C. Understanding lstm networks. **GITHUB blog, posted on August**, v. 27, p. 2015, 2015.

RAVURI, S.; STOLCKE, A. Recurrent neural network and lstm models for lexical utterance classification. **Sixteenth Annual Conference of the International Speech Communication Association**. [S.l.: s.n.], 2015.

RECUERO, Raquel. **Redes sociais na internet**. Porto Alegre: Sulina, 2009.

SEMIICAST. (2012). *Twitter reaches half a billion accounts. More than 140 milions in the U. S.* Disponível em: http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US. Acesso em: 08 de março de 2012.

SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 2014, Curitiba. **Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais [...]**. [S. l.: s. n.], 2014. Disponível em: <https://www.inf.ufpr.br/sbbd-sbsc2014/sbbd/proceedings/artigos/pdfs/Topicos-em-Gerencia-mento-de-Dados-e-Informacoes-2014.pdf>. Acesso em: 9 maio 2021.

SONI, Vibha *et al.* *Unsupervised Opinion Mining From Text Reviews Using SentiWordNet*. **International Journal of Computer Trends and Technology**, [s. l.], 2014. Disponível em: <https://www.ijcttjournal.org/archives/ijctt-v11p150>. Acesso em: 20 abr. 2021.

SOUZA, Queila; QUANDT, Carlos. Metodologia de Análise de Redes Sociais. In: DUARTE, Fábio; QUANDT, Carlos; SOUZA, Queila. **O Tempo das Redes**. [S. l.: s. n.], 2008. Disponível em: https://www.academia.edu/257818/Metodologia_De_Análise_De_Red_Sociais. Acesso em: 25 mar. 2021.

SUTSKEVER, I.; HINTON, G. E.; TAYLOR, G. W. The recurrent temporal restricted boltzmann machine. **Advances in neural information processing systems**. [S.l.: s.n.], 2009. p. 1601–1608.

SUTSKEVER, I.; MARTENS, J.; HINTON, G. E. Generating text with recurrent neural networks. **Proceedings of the 28th International Conference on Machine Learning (ICML-11)**. [S.l.: s.n.], 2011. p. 1017–1024.

TAFNER M. A. **Reconhecimento de palavras isoladas usando redes neurais artificiais**. Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, Dissertação de Mestrado, 1996.

TIOBE - *THE SOFTWARE QUALITY COMPANY*. TIOBE *Index for May 2021*. [S. l.], 2021. Disponível em: <https://www.tiobe.com/tiobe-index/>. Acesso em: 11 maio 2021.

TOSCANO, W.; CABRAL, E. L. L. Contextualização de redes neurais recorrentes. **Exacta**, São Paulo, v. 4, n. especial, p. 65-66, 25 nov. 2006.

VILAÇA, M.L.C. **Pesquisa e Ensino: Considerações e Reflexões**. **Revista Escrita**, v. 1, 2010.

ZHAO, Dejin *et al.* *How and Why People Twitter: The Role that Micro-blogging Plays in Informal Communication at Work*. [S. l.: s. n.], 2009. Disponível em: <https://www.semanticscholar.org/paper/How-and-why-people-Twitter%3A-the-role-that-plays-in-Zhao-Rosson/c473cebc334ba5bc3b11beeaff6f9d772d0e9b89>. Acesso em: 23 abr. 2021.

ZAREMBA, Wojciech; SUTSKEVER, Ilya; VINYALS, Oriol. Recurrent Neural Network Regularization. [S.l.: s.n.], [s. l.], 2015.

ZHANG, Guaquiang; PATUWO, B. Eddy; HU, Michael Y. Forecasting with artificial neural networks:: The state of the art. **International Journal of Forecasting**, [S. l.], p. 35-62, 31 jul. 1997.