UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

Lucas Lima de Oliveira

UTILIZAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA PREVER A POPULARIDADE DE TUÍTES

Santa Maria, RS
2018

Lucas Lima de Oliveira

UTILIZAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA PREVER A POPULARIDADE DE TUÍTES

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Sistemas de Informação**.

ORIENTADOR: Prof. Sérgio Luís Sardi Mergen

Santa Maria, RS
2018

Ficha catalográfica elaborada através do Programa de Geração Automática
da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

de Tal, Fulano
TÍTULO DO TRABALHO / Fulano de Tal.-2015.
50 f.; 30cm

Orientador: João da Silva
Coorientadora: Maria da Costa
Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Naturais e Exatas, Programa de Pós-Graduação em Meteorologia, RS, 2015

1. Teste 1 2. Teste 2 3. Teste 3 I. da Silva, João II. da Costa, Maria III. Título.

End. Eletr.: loliveira@inf.ufsm.com.br

Lucas Lima de Oliveira

UTILIZAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA PREVER A POPULARIDADE DE TUÍTES

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Sistemas de Informação**.

Aprovado em 25 de dezembro de 2018:

Sérgio Luís Sardi Mergen, Dr. (UFSM)
(Presidente/Orientador)

Banca Um, Dr. (AAAA)

Banca Dois, Dr. (BBBB)

Santa Maria, RS
2018

DEDICATÓRIA

Ao Rei da Espanha!

AGRADECIMENTOS

A mim!

O livro é uma criatura frágil, ele sofre o desgaste do tempo, ele teme os roedores, os elementos e mãos desajeitadas. Então o livreiro proteje os livros não apenas da humanidade, mas também da natureza e devota sua vida a uma guerra contra as forças do esquecimento.

(Umberto Eco)

RESUMO

UTILIZAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA PREVER A POPULARIDADE DE TUÍTES

AUTOR: Lucas Lima de Oliveira
ORIENTADOR: Sérgio Luís Sardi Mergen

Resumo aqui.

Palavras-chave: Palavra Chave 1. Palavra 2. Palavra 3. (...)

ABSTRACT

USE OF MACHINE LEARNING ALGORITHMS TO PREDICT TWEETS POPULARITY

AUTHOR: Lucas Lima de Oliveira
ADVISOR: Sérgio Luís Sardi Mergen

Abstract here.

Keywords: Keyword 1. Keyword 2. Keyword 3. (...)

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

API	*Application Programming Interface*
$HTTP$	*Hypertext Transfer Protocol*
$JSON$	*JavaScript Object Notation*

SUMÁRIO

1 INTRODUÇÃO

Com a grande popularização dos chamados influenciadores digitais, é notável o crescimento das mídias sociais como meios de comunicação e divulgação de conteúdos. Neste cenário, onde o número de seguidores determina a sua influência, torna-se muito importante que essas personalidades compreendam seu público, pois conteúdos direcionados refletem diretamente no alcance das publicações. Dentre as redes sociais mais utilizadas atualmente, o Twitter é um meio de veiculação de mensagens que se destaca, por sua simplicidade e objetividade. Mesmo não tendo o mesmo destaque de outras plataformas, como o Facebook ou o Instagram, o Twitter ainda conta com cerca de 335 milhões de usuários ativos, segundo Statista [1], e em média são publicados 500 milhões de tuítes por dia, segundo *Internet Live Stats* [2], fazendo dessa rede uma fonte de dados muito poderosa.

Tendo em vista o interesse dos usuários em aumentar o alcance de suas postagens, poder identificar os fatores que têm influência sobre sua popularidade é uma grande vantagem ao tentar aumentar o engajamento por parte de seus seguidores. Ser capaz de prever/estimar a popularidade que um tuíte poderá obter baseando-se nas características presentes no corpo de sua mensagem, pode trazer muitos benefícios para usuários com relativa influência nessa rede social.

Dentro deste contexto, o objetivo deste trabalho é monitorar e extrair tuítes de determinadas contas do Twitter, a fim de elaborar um modelo, utilizando algoritmos de aprendizado de máquina, para realizar a predição e classificação da popularidade de tuítes com base em suas características. Devido ao grande volume de dados, faz-se necessário automatizar o processo de análise e classificação dos dados, para isso, serão estudados e testados algoritmos já consolidados, como Naive Bayes, J48 e LTM (OLIVEIRA; MERGEN, 2018).

[1] *Statista:* `https://www.statista.com/topics/737/twitter/`
[2] *Internet Live Stats:* `http://www.internetlivestats.com/twitter-statistics/`

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados, os conceitos relacionados ao aprendizado de máquina, definindo as diferenças entre o aprendizado supervisionado e o não supervisionado. Na sequência, são apresentados alguns dos principais algoritmos deste segmento, os quais também foram utilizados na realização de experimentos no decorrer deste trabalho.

2.1 APRENDIZADO DE MÁQUINA

Entende-se como sistemas inteligentes, aqueles que são capazes de processar dados de entrada e ajustar padrões internos a fim de otimizar seus resultados de saída, de acordo com os objetivos esperados para aquele algoritmo. Dentro deste contexto, o aprendizado de máquina foca no treinamento desses algoritmos para melhorar seu desempenho. Esse processo está ligado com a redução de dimensionalidade, classificação e associação dos dados e a previsão de comportamentos.

Algoritmos para o aprendizado de máquina dividem-se em dois segmentos, aqueles que necessitam de uma supervisão para melhorar seus resultados e aqueles fazem esse processo de maneira independente. Nesta seção serão apresentados esses dois tipos de algoritmos, especificando suas características e diferenças.

2.1.1 Supervisionado

A aprendizagem supervisionada realiza o treinamento dos algoritmos com dados para os quais suas respostas já sejam conhecidas. Ou seja, dependem sempre da entrada de um padrão de valores e da comparação das respostas do sistema com aquelas consideradas corretas. Conforme o algoritmo é treinado seus padrões vão sendo ajustados a fim de diminuir o erro e otimizar os resultados. Um típico problema que utiliza a aprendizagem supervisionada é a classificação de dados, o algoritmo recebe as entradas já classificadas para realizar o treinamento e, a cada iteração, ajusta seus parâmetros a fim de obter a melhor saída, podendo ser minimizar o erro, maximizar a precisão ou a acurácia. Frequentemente, após a etapa de treinamento, é realizada uma etapa de validação, passando ao algoritmo entradas sem classificação, dessa forma seu desempenho pode ser realmente avaliado e, se necessário, o treinamento pode ser realizado novamente com novos ajustes em seus parâmetros.

2.1.2 Não Supervisionado

No caso dos algoritmos de aprendizado não supervisionado, ao contrário do outro grupo, recebem os dados sem nenhum classificação prévia. Dessa forma, conforme os dados vão sendo recebidos, o próprio algoritmo é responsável por identificar as relações e padrões presentes nos dados de entrada, o que por si só pode ser considerado um objetivo a ser alcançado. A aprendizagem não supervisionada não prevê soluções específicas para realizar o treinamento e validação dos resultados.

2.2 ALGORITMOS DE APRENDIZADO DE MÁQUINA SUPERVISIONADO

2.2.1 Naive Bayes

2.2.2 Árvores de Decisão

2.2.3 Algoritmos Genéticos

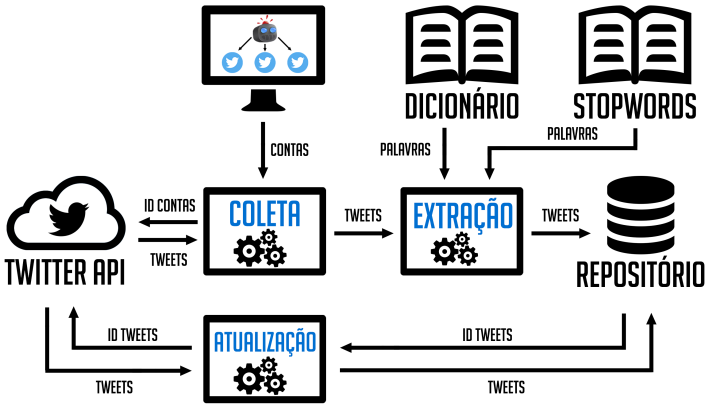
3 TRABALHOS RELACIONADOS

4 PROPOSTA

4.1 DEFINIÇÃO DOS ATRIBUTOS

4.2 ARQUITETURA DE EXTRAÇÃO DE TUÍTES

Figura 4.1 – Arquitetura adotada para extração de tuítes



Fonte: Autor.

5 EXPERIMENTOS

6 CONCLUSÕES

REFERÊNCIAS BIBLIOGRÁFICAS

OLIVEIRA, L. L. de; MERGEN, S. L. S. Análise da popularidade de tuítes com base em características extraídas de seu conteúdo. **Escola Regional de Banco de Dados (ERBD)**, v. 14, n. 1/2018, 2018. ISSN 2595-413X. Disponível em: <http://portaldeconteudo.sbc.org.br/index.php/erbd/article/view/2834>.

APÊNDICE A – DEMONSTRAÇÃO DE ALGO

Algo como apêndice.

ANEXO A – ALGO INTERESSANTE QUE ALGUÉM FEZ

Algo como anexo.