



**UNIVERSIDADE
ESTADUAL DE LONDRINA**

KRISTIANO PASINI DE OLIVEIRA

**UTILIZAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE
LINGUAGEM NATURAL E APRENDIZADO DE MÁQUINA
NA ANÁLISE DO USO DE BOTS NO INSTAGRAM**

LONDRINA - PR

2023

KRISTIANO PASINI DE OLIVEIRA

**UTILIZAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE
LINGUAGEM NATURAL E APRENDIZADO DE MÁQUINA
NA ANÁLISE DO USO DE BOTS NO INSTAGRAM**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação do Departamento de Computação da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Cinthyan Renata Sachs
Camerlengo de Barbosa

LONDRINA - PR

2023

A ficha catalográfica deverá ser solicitada na biblioteca central
(obrigatório apenas para Dissertação de Mestrado)

B222d Sobrenome, Nome do autor.

Gerência de redes – protocolo SNMP / nome e
sobrenome do aluno. – Londrina, 2010.

98 f. : il.

Orientador: nome sobrenome do orientador.

Dissertação (Mestrado em) – Universidade Estadual de Londrina, Centro de
Ciências Exatas, Programa de Pós-Graduação em xxxxx, 2010.

Inclui bibliografia.

1. Gerencia de redes – Teses. 2. SNMP. 3 NMS – Teses. 3. Assunto 3 – Teses. I.
Sobrenome, Nome do orientador. II. Universidade Estadual de Londrina. Centro de
Ciências Exatas. III. Título.

KRISTIANO PASINI DE OLIVEIRA

**UTILIZAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE
LINGUAGEM NATURAL E APRENDIZADO DE MÁQUINA
NA ANÁLISE DO USO DE BOTS NO INSTAGRAM**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação do Departamento de Computação da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA

Prof. Dr. Cinthyan Renata Sachs Camerlengo de
Barbosa

Universidade Estadual de Londrina

Prof(a). Dr(a). Segundo Membro da Banca
Universidade Segundo Membro da Banca

Prof. Dr. Terceiro Membro da Banca
Universidade Terceiro Membro da Banca

Prof. Dr. Quarto Membro da Banca
Universidade Quarto Membro da Banca

Londrina-PR, _____de _____de 2023.

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

AGRADECIMENTOS

Os agradecimentos principais são direcionados à Gerald Weber, Miguel Frasson, Leslie H. Watter, Bruno Parente Lima, Flávio de Vasconcellos Corrêa, Otavio Real Salvador, Renato Machnievszc1 e todos aqueles que contribuíram para que a produção de trabalhos acadêmicos conforme as normas ABNT com LATEX fosse possível.

Agradecimentos especiais são direcionados ao Centro de Pesquisa em Arquitetura da Informação² da Universidade de Brasília (CPAI), ao grupo de usuários latex-br³ e aos novos voluntários do grupo abnTEX²⁴ que contribuíram e que ainda contribuirão para a evolução do abnTEX².

*“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus: o
que é bom, o que Lhe é agradável, o que é
perfeito. (Bíblia Sagrada, Romanos 12, 2)*

OLIVEIRA, K. P. de. **UTILIZAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE LINGUAGEM NATURAL E APRENDIZADO DE MÁQUINA NA ANÁLISE DO USO DE BOTS NO INSTAGRAM**. 56p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina-PR, 2023.

RESUMO

As redes sociais são plataformas nas quais milhões de usuários se comunicam diariamente e esse número vem aumentando a cada ano. Os ganhos financeiros e ideológicos garantidos por perfis e opiniões com grande popularidade fomenta o desenvolvimento de diversas técnicas para alcançar tal projeção. Uma delas é o uso de ferramentas *bots*, programas que automatizam interações comuns de usuários a fim de gerar métricas artificiais que manipulem a organização natural da rede. Em virtude dos malefícios causados com essa prática, desde a dificuldade de manutenção dos próprios serviços até as possíveis influências negativas na política e na saúde pública, é mister a investigação de soluções que restrinjam tais ferramentas. Nesse sentido, o seguinte trabalho tem como proposta classificar um perfil da plataforma *Instagram* como bot ou não, com base em atividades na rede e sob a hipótese de que o comportamento de um usuário pode ser diferenciado de um programa por meio da análise de comentários e dados da plataforma. Para isso serão coletados do *Instagram* tanto métricas da conta como número de seguidos e seguidores quanto os comentários realizados em postagens. Sendo então aplicadas técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina para analisar se tais abordagens geram bons resultado para esse tipo de problema.

Palavras-chave: Aprendizado de Máquina, Redes Sociais, Processamento de Linguagem Natural, bots.

OLIVEIRA, K. P. de. **Use of Natural Language Processing and Machine Learning Techniques in the analysis of bots in Instagram** . 56p. Final Project (Bachelor of Science in Computer Science) – State University of Londrina, Londrina-PR, 2023.

ABSTRACT

Social networks are platforms where millions of people communicate on a daily basis, and that number grows more each year. Financial and ideological gains that profiles and opinions with high popularity gain encourages the development of many strategies get this status. One of those are called bots, which consists of computer programs responsible for simulating common user interactions in a automated way and by doing so generates artificial metrics that manipulate the natural organization of the network. Such practice generates a lot of issues: from the maintenance of the platform by the responsible company to even negative impacts in politics and public health. Because of that, it is critical to find ways to restrict this kind of tools. In this context, this paper proposes the use of Machine Learning algorithms that may help to identify the use of bots, under the premise that human users and automated behavior can be told apart based on comments and platform data analysis. Several data will be collected, both profile data, such as follower number and followed number and comment content across posts. That information will be analyzed with Natural Language Processing techniques and to train a Machine Learning model to verify if such approaches generates good results for this type of problem.

Keywords: Machine Learning, Social Media, Natural Language Processing, bots.

LISTA DE ILUSTRAÇÕES

Figura 1 - Exemplo de árvore de decisão para jogar tênis.....	30
Figura 2 - Exemplo de kNN com $k = 5$	31
Figura 3 - Aplicação do POS <i>Tagging</i> no <i>spaCy</i>	34

LISTA DE TABELAS

Nenhuma entrada de índice de ilustrações foi encontrada.

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
IA	Inteligência Artificial
PLN	Processamento de Linguagem Natural
POS	<i>Part of Speech</i>

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Objetivos	26
1.2	Organização do Trabalho	26
2	FUNDAMENTAÇÃO TEÓRICA.....	27
2.1	<i>Bots</i>	27
2.1.1	O uso de <i>bots</i> em redes sociais.....	27
2.2	Aprendizado de Máquina	28
2.2.1	Árvore de Decisão	29
2.2.2	<i>Random Forest</i>	30
2.2.3	<i>k-Nearest Neighbors</i>	31
2.2.4	Métricas.....	31
2.3	Processamento de Linguagem Natural.....	32
2.3.1	spaCy.....	33
2.4	Variáveis.....	34
2.4.1	Comentários em postagens.....	34
2.4.2	Dados da plataforma.....	35
3	MATERIAIS E MÉTODOS	37
3.1	Revisão Bibliográfica.....	37
3.2	Coleta dos dados e análise exploratória	39
3.3	Processamento dos comentários.....	40
3.4	Aplicação dos Algoritmos de AM.....	40
3.5	Métricas e análise dos resultados	40
4	CONSIDERAÇÕES FINAIS.....	42
	REFERÊNCIAS	43
	APÊNDICE A – NOME DO APÊNDICE	48
	ANEXO A – NOME DO ANEXO	52
	TRABALHOS PUBLICADOS PELO AUTOR	54

1 INTRODUÇÃO

Nos últimos anos as redes sociais vêm cada vez mais alcançando um maior número de usuários. Dentre as existentes pode-se explicitar as plataformas *Facebook*, *Twitter* e *Instagram* dada a grande popularidade dessas.

O aplicativo *Instagram*, o qual é desenvolvido e mantido pela *Meta Platforms*, é um aplicativo gratuito de compartilhamento de fotos e vídeos [1]. Conforme os relatórios financeiros disponibilizados pela empresa, os aplicativos do grupo família, que incluem *Facebook*, *Whatsapp* e *Instagram* ultrapassaram a margem dos 3 bilhões de usuários diários com um aumento de em média 5% ao ano [2].

Sendo assim, é clara a grande interação entre usuários que ocorre diariamente por meio da plataforma. Nesse contexto, devido à visibilidade tanto ideológica quanto monetária gerada por perfis que atraíam o interesse de uma grande quantidade de usuários, diversas abordagens são constantemente desenvolvidas a fim de obter maior sucesso e alcance. Entre elas podemos destacar o desenvolvimento de *bots*, programas capazes de gerar conteúdo de forma automatizada e interagir com outros usuários, simulando as atitudes de um humano [3].

Essa prática é contra as diretrizes da comunidade do *Instagram*, visto que esse conteúdo artificial cria uma experiência negativa e prejudica a capacidade das pessoas de interagirem de forma autêntica, além de também ameaçar a segurança, a estabilidade e a usabilidade dos serviços [4]. Nota-se também que tais atividades podem ser utilizadas de forma maliciosa em assuntos políticos e de saúde pública.

Ruediger [5] afirma que os *bots* sociais podem produzir opiniões artificiais, além de poder gerar dimensões irreais para essas ou determinadas figuras públicas, destacando também a existência de certa preocupação a respeito da propagação de notícias falsas ou de campanhas de poluição da rede, ofuscando debates com informações irrelevantes às discussões levantadas. Esse problema é ainda mais evidenciado pelo fato de tais robôs estarem envolvidos em um número notável de interações pela internet, visto que um estudo feito em 2018 pelo *Ghost Data* apontou que aproximadamente 95 milhões de contas do *Instagram* eram automatizadas e que em 2016, *bots* geraram mais tráfego na internet do que os próprios usuários comuns [6].

1.1 Objetivos

Tendo em vista esse cenário, o presente trabalho busca, por meio do uso de técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina, desenvolver um algoritmo capaz de analisar comentários e dados da plataforma *Instagram* e classificar um perfil como *bot* ou não. Tal objetivo se baseia na hipótese de que o comportamento de um humano difere de um *bot* e tais diferenças podem ser verificadas por meio da análise de métricas de perfis e atividade na plataforma.

1.2 Organização do Trabalho

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico do processo de verificação de *bots* em redes sociais explicitando o que são tais ferramentas e quais suas aplicações e seu uso nas redes. Também são descritos os métodos de Processamento de Linguagem Natural e Aprendizado de Máquina utilizados no trabalho e as variáveis coletadas da plataforma *Instagram*; a Seção 3 detalha quais foram os procedimentos metodológicos aplicados e é separada em cinco diferentes tópicos sequenciais, contendo a revisão bibliográfica, a coleta e análise exploratória dos dados, o processamento dos comentários, a aplicação dos algoritmos de Aprendizado de Máquina e a avaliação dos resultados obtidos por meio do cálculo de diferentes métricas, respectivamente. Por fim, a Seção 4 contém as considerações finais.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo descreve os conceitos necessários para o entendimento dos processos aplicados neste trabalho. A seção 2.1 aborda a definição de ferramentas *bot*, suas aplicações e sua presença em redes sociais. A seção 2.2 apresenta um detalhamento de Aprendizado de Máquina, os algoritmos utilizados e suas métricas de análise de resultados. O Processamento de Linguagem Natural se encontra na seção 2.3. Por fim, as variáveis retiradas da plataforma *Instagram* e seus significados estão na seção 2.4.

2.1 Bots

Conforme apresentado por Silva, Nunes e Santana [7], os *bots* se referem a programas que atuam como representantes de alguma entidade e que são capazes de interagir com pessoas de maneira semelhante a um humano, além de aprender e tomar decisões com base no ambiente em que estão inseridos. Tais atores não humanos estão cada vez mais inseridos em contextos informacionais que trazem algum tipo de repercussão social e que são relevantes na vida de pessoas. Como objeto de estudo os *bots* não seguem uma nomenclatura padronizada na literatura, sendo chamados dentre outros nomes, de agente de software, agente computacional, agentes inteligentes de informação, agente inteligentes, robôs, máquina ou *bot*, como diminutivo da palavra *robot* [7].

Diversas são as aplicações de programas de computador que buscam simular a presença humana em interações. Pode-se destacar os *chatbots* - sistemas computacionais que interagem com os usuários, em tempo real, por meio de diálogos em Linguagem Natural – utilizados em diversos setores devido à sua praticidade de uso e capacidade de manter um diálogo com o usuário [8]. De modo geral, aplicações desse tipo visam facilitar ou expandir a usabilidade de ferramentas computacionais e processos reais. No entanto, como apresenta Ruediger [5], o número crescente de *bots* em interações pela internet atuam na verdade com fins maliciosos.

2.1.1 O uso de *bots* em redes sociais

As redes sociais estão cada vez mais presentes na vida das pessoas. Durante a última década, essas têm ganhado espaço como uma ferramenta poderosa de comunicação

em tempo real [9]. São palco de diversas interações cotidianas e vêm se tornando parte integral não apenas da vida pessoal dos cidadãos, mas também de seus interesses públicos e políticos. Estudos realizados pela Pesquisa Brasileira de Mídia 2016 [5] mostram que 49% dos adultos brasileiros se informam pela internet e que tal fatia se apresenta em rápido crescimento.

Nesse contexto, a utilização maliciosa de *bots* nas redes sociais pode incluir a manipulação de debates políticos e a propagação de notícias falsas, gerando movimento e manipulando *trending topics* e o debate geral. Fora de seu impacto político, o uso organizado de ferramentas *bots* de forma mal-intencionada pode envolver também o compartilhamento de *links* maliciosos para o roubo de dados ou informações pessoais e até mesmo a manipulação do mercado de ações [5].

Outra situação conhecida, como apresenta Akyon e Kalfaoglu [6], é seu uso com o fim de aumentar inorganicamente a popularidade de figuras públicas, opiniões e produtos. Tal ato dificulta o processo de *marketing* das empresas inflando os preços dos serviços de anúncio em relação ao valor real, além de fazer com que os anúncios cheguem a audiências erradas. O sistema de recomendações funciona de forma inefetiva, dificultando o acesso a produtos e serviços de qualidade aos usuários.

No *Instagram*, o aumento inorgânico de popularidade envolve a automatização de atividades da plataforma, como seguir outros perfis, curtir postagens e comentários e utilização de *hashtags* específicas [6].

A aplicação de contas *bots* em tais atividades garante um baixo preço no mercado de venda de métricas de popularidade na rede por empresas privadas e facilita a existência de serviços como a venda de seguidores [6], contrários as políticas de uso da rede social [4].

2.2 Aprendizado de Máquina

O desenvolvimento de diferentes tecnologias computacionais capazes de solucionar problemas reais é um dos grandes focos atuais. A Inteligência Artificial (IA) é uma subárea da Ciência da Computação voltada para o estudo e desenvolvimento de sistemas que apresentem determinado grau de inteligência. A definição conceitual de inteligência é um dos pilares dos estudos em IA e envolve análises também nas áreas da filosofia, matemática, neurociência, psicologia etc. [10]. Para fins práticos podemos considerar de maneira

simplificada um agente como inteligente dado sua capacidade para aprender a executar uma tarefa simples ou resolver problemas complexos [11].

O Aprendizado de Máquina (AM) é uma das áreas da IA e tem como objetivo desenvolver programas capazes de tomar decisões baseado em experiências acumuladas por meio da solução bem sucedida de problemas anteriores. Sendo o aprendizado indutivo, caracterizado pelo reconhecimento de padrões e generalização de conceitos, a base por trás dos algoritmos de AM [12].

Os algoritmos de AM podem ser divididos em dois tipos os supervisionados e não-supervisionados [12].

Para os não-supervisionados o modelo analisa os exemplos fornecidos e busca de alguma maneira agrupar os dados formando diferentes agrupamentos, chamados de *clusters*. Posteriormente, é então necessária uma análise para determinar o que cada *cluster* representa naquele contexto [12].

Já para os supervisionados, utilizados neste trabalho, cada dado consiste em um vetor de atributos (*features*) e o rótulo da classe associada, sendo o objetivo do algoritmo classificar corretamente novos exemplos que não estejam rotulados em alguma das classes. Os objetos de estudo de AM são chamados de problemas de classificação quando envolvem rótulos com valores discretos e de regressão para os contínuos [12].

2.2.1 Árvore de Decisão

Árvore de Decisão é um modelo de abordagem em divisão e conquista para problemas de classificação ou regressão. Podem ser utilizadas para extrair padrões em grandes quantidades de dados e possuem um funcionamento intuitivo, justificando sua importância e popularidade na área de AM [13].

O modelo gerado pelo algoritmo pode ser representado em formato de árvore, sendo possível a observação visual a partir da raiz até os diferentes níveis da árvore e seus respectivos processos de classificação [11].

A formação da árvore se dá de acordo com critérios de particionamento, sendo os dois principais o *Ganho de Informação* e o *Índice Gini*. O primeiro busca maximizar a quantidade de informação obtida naquele nível da árvore ao definir o atributo de partição. Já o segundo é voltado para a minimização da impureza (quantidade de classes diferentes restantes) após um particionamento [13].

Na **Erro! Fonte de referência não encontrada.** está representado um exemplo de árvore de decisão para classificação binária do problema de escolha de jogar ou não uma partida de tênis em determinado dia. Os particionamentos foram feitos levando em consideração o maior Ganho de Informação e envolve perguntas sobre o clima no dia do jogo.

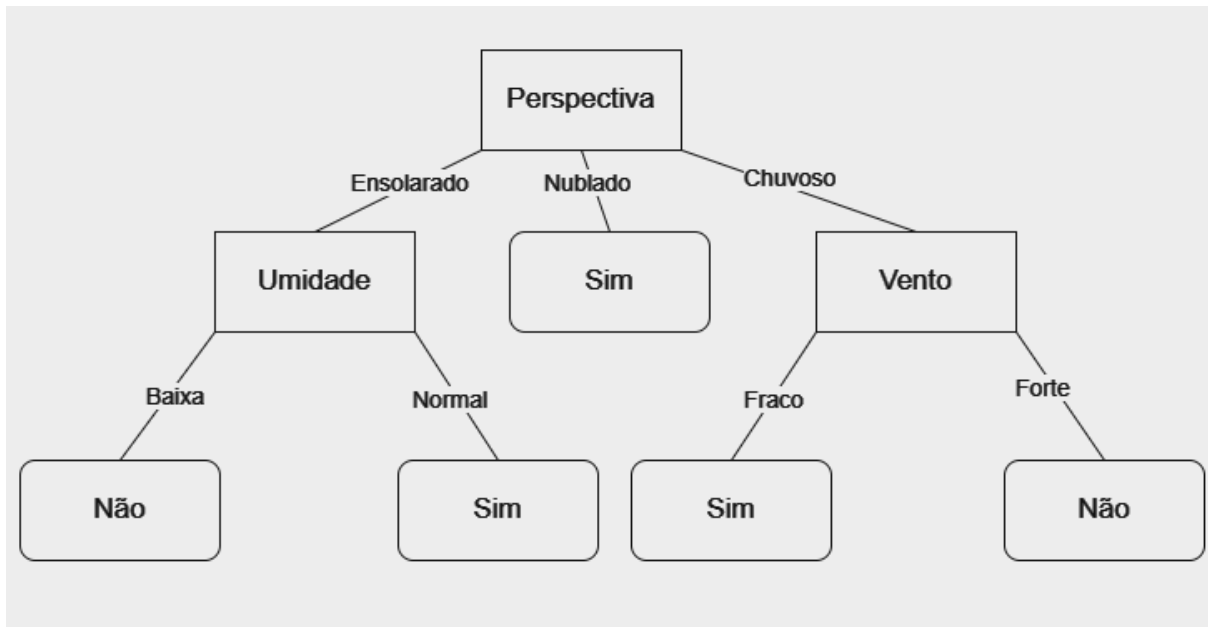


Figura 1 - Exemplo de árvore de decisão para jogar tênis.

2.2.2 *Random Forest*

O *Random Forest* é um algoritmo de AM do tipo *Ensemble* que se baseia na combinação de múltiplas árvores de decisão do tipo *Classification and Regression Tree* e busca garantir uma boa resolução genérica para problemas complexos por meio da agregação de ideias [11].

A criação de cada uma das árvores de decisão é feita por meio do processo de *bootstrap*, no qual os dados iniciais são divididos em subconjuntos de maneira aleatória por reamostragem com reposição e sem repetição e passados ao algoritmo [11]. Após a geração de um grande número de árvores é feita a votação da classe mais popular entre os resultados obtidos [14].

Dentre os algoritmos aplicados nas análises de *bots* em ambientes de rede social o *Random Forest* se destaca dado sua grande popularidade e bons resultados, conforme os estudos realizados por pesquisas na área [3].

2.2.3 *k*-Nearest Neighbors

O *k*-Nearest Neighbors (k-NN) é um algoritmo de AM supervisionado utilizados em problemas de classificação ou regressão. O processo de classificação do algoritmo é realizado por meio da verificação da classe de k-vizinhos em um ambiente n-dimensional, sendo “n” definido pela quantidade de *features* dos dados [15].

A verificação de proximidade entre os pontos é realizada pelo cálculo da distância desses em relação ao novo elemento analisado, sendo atribuído a esse a classe cuja maioria de seus vizinhos pertence [15].

A Figura 2 ilustra um exemplo apresentado por Artoni [11] demonstrando o processo de classificação de um novo elemento (representado por um losango dourado). Como três dos cinco vizinhos mais próximos pertencem a Classe 2, o novo dado será classificado também como Classe 2.

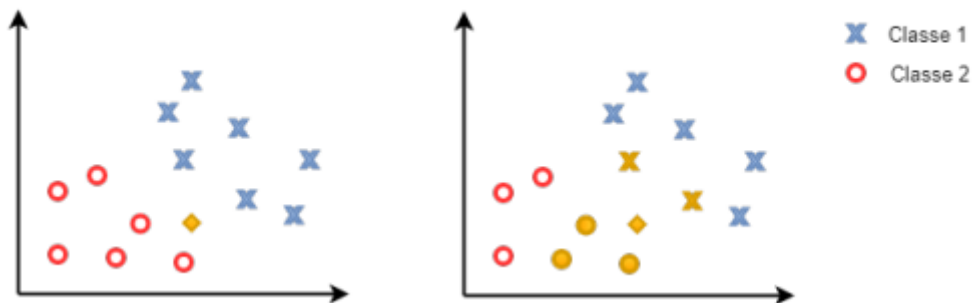


Figura 2 - Exemplo de kNN com $k = 5$ [11].

2.2.4 Métricas

O resultado obtido na aplicação de modelos de AM pode ser analisado com base em diferentes métricas estatísticas. Em geral, se deve buscar o cálculo de métricas que evidenciem diferentes aspectos dos resultados, dado que cada uma tem sua importância e quando combinadas, compensam pelas fraquezas das outras [16].

O *recall* consiste na fração de casos positivos preditos como positivos. A precisão é dada pela fração dos casos verdadeiramente positivos dentre todos os apontados como positivos pelo modelo. O *F-Measure* representa a média harmônica da precisão e do *recall*. Já a acurácia é a fração de casos que o modelo classificou corretamente [17].

Os cálculos das quatro métricas podem ser visualizados nas equações 1 a 4 [17]. Os termos VP (Verdadeiro Positivo), FP (Falso Positivo), FN (Falso Negativo) e VN (Verdadeiro Negativo) presentes nas equações se referem aos resultados considerados pelo modelo. Sendo eles os valores de fato positivos apontados como positivos, os valores negativos apontados como positivos, os valores positivos apontados como negativos e os valores de fato negativos apontados como negativos, respectivamente.

$$Precisão = \frac{VP}{VP + FP} \quad (1)$$

$$Recall = \frac{VP}{VP + FN} \quad (2)$$

$$F - measure = 2 * \frac{Recall \times Precisão}{Recall + Precisão} \quad (3)$$

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (4)$$

2.3 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é um campo de pesquisa voltado a propor sistemas para o processamento computacional da linguagem humana, sendo o termo “natural” utilizado para se referir as línguas faladas por humanos, diferenciando-as de outras linguagens como a matemática, a visual, as gestuais, de programação etc. Sob uma perspectiva da Ciência da Computação, o PLN é uma área da IA e da linguística computacional [18].

Como apresentam Caseli e Nunes [18], o PLN é dividido em duas grandes subáreas: a Interpretação (ou Compreensão) de Linguagem Natural – NLU (do inglês, *Natural Language Understanding*) e Geração de Linguagem Natural – NLG (do inglês, *Natural*

Language Generation). A NLU engloba tudo o que diz respeito à análise e ao processamento da língua e envolve duas etapas. A primeira é a segmentação e classificação dos componentes linguísticos. A segunda é a interpretação dos significados construídos. Já a NLG tem como objetivo a geração de linguagem natural, como, por exemplo, em aplicações de *chatbots* que retornem respostas ao usuário ou como destacam também os autores o próprio ChatGPT¹, exemplo atual de maior sucesso, capaz de gerar língua de forma tão fluente quanto muitos humanos [18].

A realização do processamento morfológico em PLN envolve a execução de algumas etapas conhecidas dentro do pré-processamento de dados. A segmentação do texto em sentenças (sentencição), a separação de palavras (tokenização), a tokenização em subpalavras (vetorização de *subtokens*) e a normalização de palavras (lematização e radicalização) [18].

A sentencição visa encontrar os limites de cada sentença por meio da avaliação de elementos textuais como, no caso do português, pontuações delimitadoras (“.”, “...”, “!”, “?” etc.) e suas possíveis ambiguidades na língua [18].

A tokenização realiza a separação do texto em unidades linguísticas mínimas (palavras, no caso do português) também por meio do uso de pontuações e espaços em branco, atentando-se a possíveis ambiguidades, como o caso de um número decimal (8,3) que deve ser considerado como um único token e não três [18].

A abordagem de tokenização em subpalavras busca codificar palavras mais complexas, representando-as como combinações de outras mais simples, como, por exemplo, representar “felizmente” como “feliz” + “mente” [18].

Por fim a normalização realiza a conversão das palavras em formas padronizadas. Alguns tipos de situações incluem as transformações de abreviaturas para suas formas reais, a substituição de caracteres maiúsculos e minúsculos, a detecção de radicais comuns das palavras (radicalização) e a verificação de verbos conjugados (lematização) [18].

2.3.1 spaCy

O *spaCy* é uma biblioteca de Processamento de Linguagem Natural para análise sintática que pode ser escrita nas linguagens Python e Cython, suporta a Língua

¹ <https://chat.openai.com/>

Portuguesa e possui código aberto, ou seja, disponibiliza de forma gratuita o *download* dos códigos fonte [8].

Dentre as funcionalidades oferecidas pela ferramenta *spaCy* estão incluídas as etapas de pré-processamento da língua portuguesa e a classificação dos *tokens* em relação a sua classe gramatical, como artigo, verbo, adjetivo, pronome, substantivo etc. Um exemplo de resultado retornado pela análise de uma frase pode ser observado na Figura 3. Tal processo recebe o nome de *Part of Speech (POS) Tagging* [18].

```
Não vão resolver meu problema?  
Não ADV advmod  
vão AUX aux  
resolver VERB ROOT  
meu DET det  
problema NOUN obj  
? PUNCT punct
```

Figura 3 - Aplicação do POS *Tagging* no *spaCy*.

2.4 Variáveis

Esta seção apresenta as variáveis escolhidas para os modelos de Aprendizado de Máquina supervisionados aplicados. As variáveis estão divididas em dois principais grupos:

1. Comentários em postagens: Grupo contendo os *tokens* e aspectos gramaticais dos comentários coletados;
2. Dados da plataforma: Grupo contendo as informações dos perfis disponibilizadas pela plataforma Instagram.

2.4.1 Comentários em postagens

Para os dados textuais coletados dos comentários foram escolhidas a análise tanto das próprias palavras, contendo a frequência de uso de *tokens* para cada tipo de usuário (*bot* ou não *bot*) como palavras, *emojis*, *hashtags*, siglas etc. quanto de aspectos morfológicos, analisados por meio da ferramenta *spaCy*.

Ao analisar dados textuais no Twitter, Miranda [19] aponta o maior uso de hashtags e emojis por perfis *bots*, sendo o último o mais significativo de acordo com os resultados encontrados. Dentre os trinta tokens mais utilizados por perfis classificados como *bot*, oito eram emojis, ao passo que esse valor caía para três para usuários reais.

A verificação da relevância das variáveis morfológicas seguiu a iniciativa apresentada por Miranda [19] em sua análise na plataforma do *Twitter*. São elas:

1. Número médio de verbos: dado pelo número de advérbios dividido pela quantidade total de *tokens* no comentário;
2. Número médio de adjetivos: dado pelo número de adjetivos dividido pela quantidade total de *tokens* no comentário;
3. Número médio de pronomes: dado pelo número de pronomes dividido pela quantidade total de *tokens* no comentário;
4. Número médio de hashtags: dado pelo número de hashtags dividido pela quantidade total de *tokens* no comentário;
5. Número médio de *emojis*: dado pelo número de *emojis* dividido pela quantidade total de *tokens* no comentário.

2.4.2 Dados da plataforma

Com base nos dados apresentados por Akyon e Kalfaoglu [6] e Alothali *et al.* [3] as informações escolhidas da plataforma *Instagram* para análise a nível de perfil foram:

1. Quantidade de dígitos numéricos no nome de usuário;
2. Se possui foto de perfil (sim ou não, campo binário);
3. Quantidade de caracteres na descrição do perfil (“*bio*”);
4. Número de seguidores;
5. Número de seguidos;
6. Número de postagens;
7. Ano de criação da conta;
8. Privacidade da conta (privada ou não privada, campo binário);
9. Número de curtidas do comentário coletado.

A configuração de perfil privado do *Instagram* realiza o ocultamento da visualização da listagem de identificação dos perfis de seguidores e seguidos e o conteúdo das postagens realizadas (porém, mantendo ainda visível o número de postagens existentes).

Portanto, tal configuração não acarreta a perda de alguma das informações para esse tipo de perfil.

Conforme Akyon e Kalfaoglu [6], a métrica de quantidade de dígitos numéricos no nome do usuário é justificada pelo fato de que mais de 50% das contas *bots* possuíam tal característica, ao passo que para usuários reais esse valor caía para 11% na pesquisa realizada pelos referidos autores.

Dados como a quantidade de dígitos no *username*, idade da conta e existência de foto podem indicar possíveis comportamentos de um baixo nível de detalhamento e baixa individualidade, diretamente relacionados ao processo de geração de contas falsas. Já o número de seguidores e seguidos, por exemplo, estão diretamente relacionados à busca de elevação da popularidade de outros usuários [6], e, portanto, podem ser boas métricas a nível de perfil.

3 MATERIAIS E MÉTODOS

Este capítulo descreve os materiais utilizados e a metodologia seguida ao longo deste trabalho. A Seção 3.1 contém a revisão bibliográfica, consistindo na busca de publicações científicas na área de detecção de *bots* e uso de PLN em redes sociais; a Seção 3.2 descreve o processo de coleta de dados na plataforma *Instagram*; a Seção 3.3 apresenta as etapas da aplicação de técnicas de PLN para o processamento dos comentários; a Seção 3.4 detalha o *script* desenvolvido para implementação dos modelos de AM; a Seção 3.5 contém o cálculo das métricas para cada algoritmo e os resultados obtidos.

3.1 Revisão Bibliográfica

Para a revisão bibliográfica foram buscadas publicações científicas que envolvem a aplicação de diferentes técnicas de Aprendizado de Máquina e de Processamento de Linguagem Natural para a detecção de *bots* em redes sociais. Tal pesquisa buscou permitir o entendimento sobre os conceitos e algoritmos utilizados com base em resultados obtidos por trabalhos já realizados.

Existem diversos estudos sobre a identificação de *bots* em redes sociais. O *Twitter*, por exemplo, por disponibilizar até o início de 2023 o acesso a sua API de forma gratuita, permitindo a coleta de um grande volume de dados para análise, garantiu a criação de múltiplos *datasets* para pesquisas na área.

Kudugunta e Ferrara [9] demonstram a utilização de uma rede neural profunda na detecção de perfis automatizados. Nesse trabalho foram coletados tanto os dados da plataforma quanto os conteúdos dos *tweets* na classificação dos perfis. Como resultado foi demonstrado uma acurácia de 96% ao analisar apenas mensagens avulsas e de 99% ao incluir análises em nível de perfil. Tais resultados indicam a eficiência de abordagens de Aprendizado de Máquina para a resolução desse problema.

Rodríguez-Ruiz *et al.* [20] discute as diferentes abordagens utilizadas na detecção de *bots* no *Twitter*, destacando o uso tanto de algoritmos supervisionados quanto os não supervisionados, sendo o primeiro o mais utilizado. O autor também pondera que, apesar da maior parte dos mecanismos propostos terem obtidos resultados satisfatórios, esses requerem dados rotulados de perfis automatizados para se obter padrões característicos. Além

disso, destaca a heterogeneidade desse tipo de ferramenta e a necessidade de métodos de detecção que não se baseiem unicamente em *bots* já existentes.

A detecção de *bots* também pode ser realizada sob uma perspectiva de classificação de classe única [20]. Tal metodologia busca analisar apenas dados de usuários legítimos com a finalidade de se encontrar comportamentos deviantes. Os resultados obtidos apontaram uma acurácia acima de 90%. Com tais resultados, Rodríguez-Ruiz *et al.* [20] conclui que classificadores desse tipo podem servir como uma forma inicial de detecção, apontando padrões de possíveis novos *bots*.

Miller *et al.* [16] explora dois algoritmos de aprendizado não supervisionado para a detecção de mensagens de spam. São eles o *Density-based clustering* e o *K-means based clustering*, apresentando métricas obtidas para cada um.

Alothali *et al.* [3] cita diferentes métodos de detecção de *bots* em redes sociais. Entre elas, pode-se destacar o *crowdsourcing* que consiste na utilização da inteligência humana para diferenciar contas falsas de usuários comuns. Tal abordagem envolve diferentes critérios, de acordo com as métricas escolhidas para justificar a classificação dada.

Um exemplo de ferramenta disponível *online* para a verificação de perfis em tempo real é a PEGABOT², feita pelo Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio) em parceria com o Instituto Equidade e Tecnologia com o objetivo de classificar perfis do *Twitter* com base na probabilidade de serem *bots*. No entanto, com a mudança dos planos gratuitos da API do *Twitter* permitindo apenas a postagem e não mais a leitura de conteúdo, a plataforma teve seus serviços interrompidos conforme comunicado no próprio perfil da ferramenta.

Miranda [19] realizou a aplicação de Processamento de Linguagem Natural e Aprendizado de Máquina em dados coletados da atividade de perfis no *Twitter*, desenvolvendo uma análise da relevância de variáveis morfológicas no processo de detecção de *bots*. Utilizou de diversas ferramentas como o próprio PEGABOT, o algoritmo de Aprendizado de Máquina chamado *Naive Bayes* e algumas bibliotecas para Processamento de Linguagem Natural como o pacote spaCy³ e o LeIA⁴ (*Léxico para Inferência Adaptada*).

O uso de PLN como ferramenta para a solução de problemas em redes sociais pode envolver também outras temáticas. Mioni [21] aplicou PLN para detectar

² <https://pegabot.com.br/>

³ <https://spacy.io/>

⁴ <https://github.com/rafjaa/>

comportamento tóxico no *Twitter*, descrevendo cada uma das etapas seguidas na análise dos *tweets*. A base de dados utilizada foi um *corpus* contendo 450 tweets capturados enquanto um episódio do programa *MasterChef* Brasil era exibido. Dentre os resultados do trabalho o estudo de Processamento da Língua Portuguesa em conteúdos inseridos por usuários em plataformas como redes sociais e o processo de análise de dados reais da *web* contribuíram com a coleta e o processamento dos comentários.

No caso do *Instagram*, Akyon e Kalfaoglu [6] apresentam algumas diretrizes para a análise de perfis na perspectiva de verificação de contas falsas. Características como a ausência de foto de perfil, nomes de usuários incomuns e confusos, com número elevado de números, por exemplo, são grandes indicativos para *bots*. Alto número de seguidos e baixo de seguidores também se enquadram no grupo, visto que o principal objetivo de tais programas é a manipulação de métricas da plataforma.

A plataforma *Instagram* foi escolhida como fonte de dados para a análise deste trabalho, apesar de impor mais barreiras ao acesso de sua API. Tal escolha se deu com base nas mudanças de acesso a API do *Twitter*, menor foco em exploração de múltiplas atividades de um único perfil e a maior familiaridade do autor com a primeira.

3.2 Coleta dos dados e análise exploratória

A coleta dos dados será realizada por meio da recuperação manual de perfis e comentários em postagens de diferentes temáticas. O acesso será feito por meio da ferramenta de exploração do *Instagram*, que apresenta ao usuário uma listagem contínua de *posts*. Será feita a coleta apenas nas postagens com popularidade considerável (que possuam mais de 100 comentários), sendo elas tanto sobre assuntos pessoais quanto de negócio. *Hashtags* e conteúdo de comentário que referenciem usuários da plataforma serão tratados a fim de manter a privacidade. Serão coletados cerca de 300 comentários e perfis para o modelo.

A rotulação dos perfis como bots será feita pela atribuição de um grau de automatização por meio da apresentação dos dados coletados a diferentes equipes que avaliarão o perfil, a fim de obter uma pontuação final após a união das diferentes perspectivas, sendo descritos quais os critérios utilizados em cada avaliação. Tal abordagem se baseia na possibilidade de aplicação da análise humana para esse problema, pois como é apresentado por Alothali *et al.* [3], essa metodologia apresenta boa acurácia dos resultados, apesar de

demandar um maior tempo. Nesse contexto, os autores apresentam como exemplo uma competição de verificação de *bots* no *Twitter* organizada pela DARPA em 2015, na qual times submetiam seus chutes na verificação de perfis envolvidos em discussões sobre as vacinações.

Na competição, o conjunto de contas analisadas era formado por 7.038 perfis do *Twitter*, contendo usuários reais e *bots*. Esses últimos eram configurados para influenciar o debate sobre determinado tópico, sendo o trabalho dos participantes detectar os *bots* programados estritamente para apoiar discussões sobre a importância das vacinações [22].

Cada conta possuía um bloco de *tweets* associado, totalizando 4.095.083 *tweets* entre todos os perfis. Os critérios analisados pelas equipes envolveram a sintaxe e semântica dos *tweets*, as variáveis temporais (mudança de sentimento ao longo do tempo, frequência de *tweets* por dia etc.) e as métricas de perfil como existência de foto, número de seguidores e seguidos, número de menções etc. Três times dentre os seis participantes tiveram pontuação máxima em relação à habilidade de verificação de contas automatizadas [22].

3.3 Processamento dos comentários

O processamento dos dados textuais será realizado com a biblioteca spaCy. Serão analisados aspectos como termos populares, quantidade de palavras, uso de verbos, adjetivos, pronomes e demais classes apontadas pelo spaCy.

3.4 Aplicação dos Algoritmos de AM

Será realizada a utilização de dois métodos supervisionados, dado que o uso de vários tipos de algoritmos busca explorar diferentes aspectos do *dataset* a fim de encontrar uma boa forma de detecção de *bots* [6].

Os algoritmos a serem aplicados serão o Random Forest e o k-NN.

3.5 Métricas e análise dos resultados

Serão calculadas métricas estatísticas para cada um dos dois algoritmos, com o intuito de verificar se os dados utilizados obtiveram resultados satisfatórios. Com base

nas variáveis escolhidas anteriormente serão desenvolvidos gráficos para explicitar quais foram as diferenças encontradas entre os perfis *bots* e não *bots*.

4 CONSIDERAÇÕES FINAIS

REFERÊNCIAS

- [1] INSTAGRAM. Sobre o Instagram. *Instagram*, 2023. Disponível em: <<https://help.instagram.com/424737657584573>>. Acesso em: 08 Julho 2023.
- [2] PLATFORMS, M. Meta Reports First Quarter 2023 Results. *Meta Investor Relations*. Disponível em: <<https://investor.fb.com/investor-news/press-release-details/2023/Meta-Reports-First-Quarter-2023-Results/default.aspx>>. Acesso em: 08 Julho 2023.
- [3] ALOTHALI, E. et al. Detecting Social Bots on Twitter: A Literature Review. *International Conference on Innovations in Information Technology (IIT)*, 13., Al Ain, United Arab Emirates: IEEE. 2018. p. 175-180.
- [4] META. Spam. *Transparency Center*. Disponível em: <<https://transparency.fb.com/pt-br/policies/community-standards/spam/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2Fspam>>. Acesso em: 08 Julho 2023.
- [5] RUEDIGER, M. A. *Robôs, redes sociais e política no brasil: estudo sobre interferências ilegítimas no debate público na web, riscos à democracia e processo eleitoral de 2018*. Rio de Janeiro: FGV DAPP, 2017.
- [6] AKYON, F. C.; ESAT KALFAOGLU, M. Instagram Fake and Automated Account Detection. *Innovations in Intelligent Systems and Applications Conference (ASYU)*, 10., Turquia: IEEE. 2019. p. 1-7.
- [7] SILVA, C. T.; NUNES, A. A.; SANTANA, C. A. Bots como objeto de estudo da Ciência da Informação. *PontodeAcesso: Revista do Instituto de Ciência da Informação da UFBA*, Salvador, v. 15, n. 1/2, p. 138-153, Dezembro 2021.
- [8] FARIA, C. R. E. *Ferramenta Carolina para identificação de pragas e doenças na cultura da soja utilizando processamento de linguagem natural*. Dissertação de Mestrado, Departamento de Computação, Universidade Estadual de Londrina. Londrina, 163p. 2021.
- [9] SNEHA KUDUGUNTA, E. F. Deep neural networks for bot detection. *Information Sciences*, v. 467, p. 312-322, Outubro 2018.
- [10] RUSSEL, S.; NORVIG, P. *Artificial Intelligence: A modern Approach*. 3°. ed. Upper Saddle River, NJ: Prentice Hall, 2010.

- [11] ARTONI, A. A. *Aplicação de Aprendizado de Máquina no auxílio ao diagnóstico do transtorno do espectro autista*. Dissertação de Mestrado, Departamento de Computação, Universidade Estadual de Londrina. Londrina, 70p. 2020.
- [12] MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre Aprendizado de Máquina. In: REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. 1ª. ed. Barueri: Manole, v. I, 2003. Cap. 4, p. 89-114.
- [13] MYLES, A. J. et al. An introduction to decision tree modeling. *Journal of Chemometrics*, v. 18, n. 6, p. 275–285, Junho 2004.
- [14] BREIMAN, L. Random Forests. *Machine Learning*, California, v. 45, p. 5-32, Outubro 2001.
- [15] CUNNINGHAM, P.; DELANY, S. J. k-Nearest Neighbour Classifiers - A Tutorial. *ACM Computing Surveys*, v. 54, n. 6, p. 1-25, Julho 2021.
- [16] MILLER, Z. et al. Twitter spammer detection using data stream clustering. *Information Sciences*, v. 260, p. 64-73, Março 2014.
- [17] ERICKSON, B. J.; KITAMURA, F. Magician's Corner: 9. Performance Metrics for Machine Learning Models. *Radiology Artificial Intelligence*, v. 3, n. 3, p. e200126, Maio 2021.
- [18] CASELI, H. D. M.; NUNES, M. D. G. V. *Processamento de Linguagem Natural Conceitos, Técnicas e Aplicações em Português*. 1ª. ed. São Carlos: BPLN, 2023.
- [19] MIRANDA, G. A. *Detecção de Bots no Twitter Através de Técnicas de Processamento de Linguagem Natural*. TCC (Graduação) - Curso de Estatística, Instituto de Matemática e Estatística, Universidade Federal Fluminense. Niterói, 48p. 2022.
- [20] RODRÍGUEZ-RUIZ, J. et al. A one-class classification approach for bot detection on Twitter. *Computers & Security*, v. 91, p. 101715, Abril 2020.
- [21] MIONI, J. L. V. M. *Processamento da língua portuguesa na detecção de toxicidade na rede social twitter*. Dissertação de Mestrado, Departamento de Computação, Universidade Estadual de Londrina. Londrina, 84p. 2023.
- [22] SUBRAHMANYAN, V. S. et al. The DARPA Twitter Bot Challenge. *In Computer*, v. 49, n. 6, p. 38-46, Junho 2016.

Apêndices

Anexos

TRABALHOS PUBLICADOS PELO AUTOR

Trabalhos publicados pelo autor durante o programa (obrigatório somente para teses de doutorado e dissertações de mestrado no template DC/UEL).

1. Jose da silva, autor2 da silva, orientador da silva, **Titulo do artigo**, local onde foi publicado, mes/ano, editora, numero de pagina, isbn, (Qualis CC 2012, xx)
2. Jose da silva, autor2 da silva, orientador da silva, **Titulo do artigo**, local onde foi publicado, mes/ano, editora, numero de pagina, isbn, (Qualis CC 2012, xx)
3. Jose da silva, autor2 da silva, orientador da silva, **Titulo do artigo**, local onde foi publicado, mes/ano, editora, numero de pagina, isbn, (Qualis CC 2012, xx)
4. Jose da silva, autor2 da silva, orientador da silva, **Titulo do artigo**, local onde foi publicado, mes/ano, editora, numero de pagina, isbn, (Qualis CC 2012, xx)