



**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

2024

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.

Orientadora: Prof. Dra. Cinthyan Renata Sachs
Camerlengo de Barbosa

LONDRINA - PR

2024

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

Profa. Dra. Cinthyan Renata Sachs
Camerlengo de Barbosa
Universidade Estadual de Londrina

Prof. Dr. Vitor Valério de Souza Campos
Universidade Estadual de Londrina

Profa. Carolinne Roque e Faria
Centro Universitário das Faculdades Integradas
de Ourinhos

Londrina-PR, 07 de maio de 2024.

AGRADECIMENTOS

Agradeço a minha orientadora Profa. Dra. Cinthyan Renata Sachs Camerlengo de Barbosa pelo auxílio no desenvolvimento deste trabalho e aos demais professores do Departamento de Computação pelo apoio durante a graduação.

Também agradeço meus colegas de curso, com destaque ao Guilherme e ao Pedro, por me ajudarem na coleta de dados do trabalho.

Por fim, gostaria de agradecer minha família que sempre me deu apoio e incentivo.

OLIVEIRA, K. P. de. **UTILIZAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE LINGUAGEM NATURAL E APRENDIZADO DE MÁQUINA NA ANÁLISE DO USO DE BOTS NO INSTAGRAM**. 50p. 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 teve como proposta treinar um modelo de Aprendizado de Máquina a fim de 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. Foi feita uma Revisão Sistemática da Literatura para verificar estudos na área e posteriormente realizou-se a coleta de perfis do *Instagram*, envolvendo dados de métricas da conta como número de seguidos e seguidores e também os comentários realizados em postagens. O *Random Forest* foi o modelo de Aprendizado de Máquina com melhor resultado, obtendo 100% de acurácia na classificação dos perfis *bots* do conjunto de dados coletado. Na análise dos dados foi observado um maior uso de pontuações pelos usuários reais e de substantivos para os *bots*. Informações como a quantidade de seguidores e seguidos, privacidade do perfil e idade da conta também se apresentaram como características diferenciadoras. Com os resultados obtidos foi possível complementar o estudo da detecção de *bots* no *Instagram*, com foco em perfis em Língua Portuguesa.

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** . 50p. 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 daily, and that number grows more each year. The financial and ideological gains achieved by profiles and opinions with high popularity encourage the development of a variety of strategies to reach this level of visibility. One of them is the use of bot tools, which are programs that automate common user interactions in order to generate artificial metrics that manipulate the natural organization of the network. Due to the harm caused by this practice, both in terms of the difficulty of maintaining the services and the possible negative influences on politics and public health, it is necessary to examine solutions that restrict these tools. In this context, this project proposes the use of Machine Learning algorithms to classify an Instagram profile as bot or not, under the premise that human users and automated behavior can be separated from that of a program by analyzing comments and data from the platform. A Systematic Literature Review was carried out to verify existing studies and then data was collected from Instagram, including profile metrics, such as follower number and followed number, and comment content across posts. Random Forest achieved the best result, with a 100% accuracy in classifying the bot profiles of the dataset. Real users showed a higher usage of punctuation while bots comments had more nouns. Information such as follower and the following number, profile privacy and account age also presented as good differentiators for profiles. As a result of this research, it was possible to complement the study of bot detection on Instagram, focusing on profiles in Portuguese.

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

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de Árvore de Decisão para jogar tênis [14].	16
Figura 2 – Exemplo de kNN com $k = 5$ [11].	17
Figura 3 – Aplicação do POS Tagging no <i>spaCy</i>	20
Figura 4 – Exemplo da ferramenta de explorar e <i>stories</i> patrocinado.	30
Figura 5 – Percentual de classes gramaticais nos comentários.	34
Figura 6 – Palavras mais escritas por <i>bots</i>	36
Figura 7 – Palavras mais escritas por usuários reais.	36
Figura 8 – Média da quantidade de curtidas nos comentários para cada perfil.	37
Figura 9 – Gráfico do ano de criação das contas <i>bots</i>	38
Figura 10 – Gráfico do ano de criação das contas de usuários reais.	38
Figura 11 – Gráfico contendo a quantidade de perfis <i>bots</i> privados e não privados.	39
Figura 12 – Gráfico contendo a quantidade de perfis reais privados e não privados.	39

LISTA DE TABELAS

Tabela 1 – Artigos selecionados na Revisão Sistemática da Literatura	24
Tabela 2 – Métricas estatísticas de avaliação do Primeiro Modelo.....	41
Tabela 3 – Métricas estatísticas de avaliação do Segundo Modelo.....	42

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
IA	Inteligência Artificial
PLN	Processamento de Linguagem Natural
POS	<i>Part of Speech</i>
RSL	Revisão Sistemática da Literatura

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivos.....	12
1.2	Organização do Trabalho.....	12
2	FUNDAMENTAÇÃO TEÓRICA.....	13
2.1	<i>Bots</i>	13
2.1.1	<i>Bots</i> em redes sociais.....	14
2.2	Aprendizado de Máquina.....	14
2.2.1	Árvore de Decisão	15
2.2.2	<i>Random Forest</i>	16
2.2.3	<i>k-Nearest Neighbors</i>	17
2.2.4	Métricas	17
2.3	Processamento de Linguagem Natural	18
2.3.1	<i>spaCy</i>	20
2.4	Variáveis para os modelos de AM.....	20
2.4.1	Comentários em postagens	20
2.4.2	Dados da plataforma.....	21
3	MATERIAIS E MÉTODOS	23
3.1	Revisão Sistemática da Literatura	23
3.2	Coleta dos dados	30
3.2.1	Classificação dos perfis	31
3.3	Processamento dos comentários	32
3.4	Algoritmos de Aprendizado de Máquina.....	33
4	RESULTADOS	34
4.1	Análise morfológica	34
4.2	Influência das variáveis	36
4.3	Modelos de Aprendizado de Máquina.....	41
5	CONSIDERAÇÕES FINAIS.....	43
	REFERÊNCIAS	45
	APÊNDICE A – CORPUS COMENTÁRIOS	48

1 INTRODUÇÃO

Nos últimos anos as redes sociais têm cada vez mais alcançado um maior número de usuários. Dentre as existentes podem-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 três 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 problema, 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: o Capítulo 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*; o Capítulo 3 detalha quais foram os procedimentos metodológicos aplicados e esses são separados em quatro diferentes tópicos sequenciais, contendo a Revisão Sistemática da Literatura, a coleta e análise exploratória dos dados, o processamento dos comentários e o desenvolvimento dos algoritmos de Aprendizado de Máquina. O Capítulo 4 apresenta os resultados obtidos na avaliação dos dados e no cálculo das métricas de performance dos modelos. Por fim, o Capítulo 5 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 *chatterbots* - 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 atua na verdade com fins maliciosos.

2.1.1 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 a base por trás dos algoritmos de AM, caracterizado pelo reconhecimento de padrões e generalização de conceitos [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. Pode ser utilizada 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 Figura 1 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.

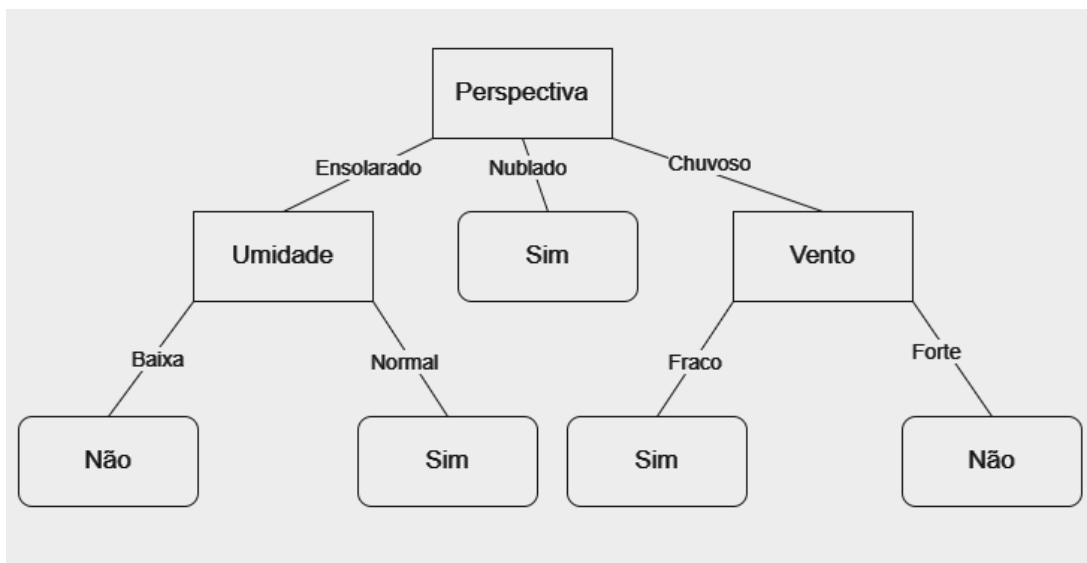


Figura 1 – Exemplo de *Árvore de Decisão* para jogar tênis [14].

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 [15].

Dentre os algoritmos aplicados nas análises de *bots* em ambientes de rede social o *Random Forest* se destaca dada 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 utilizado 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 [16].

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 [16].

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 à Classe 2, o novo dado será classificado também como Classe 2.

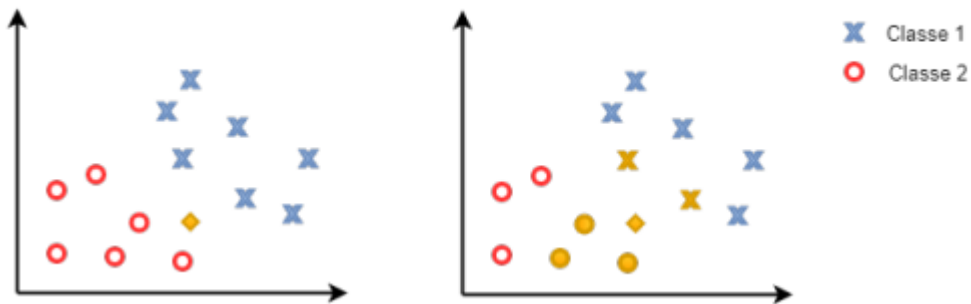


Figura 2 – Exemplo de *k*NN com $k = 5$ [11].

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 [17].

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 [18].

Os cálculos das quatro métricas podem ser visualizados nas equações 1 a 4 [18]. 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 [19].

Como apresentam Caseli e Nunes [19], 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 da 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 [19].

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 (sentenciação), a separação de palavras (tokenização), a tokenização em subpalavras (vetorização de *subtokens*) e a normalização de palavras (lematização e radicalização) [19].

A sentenciação 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 [19].

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 [19].

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” [19].

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) [19].

¹ <https://chat.openai.com/>

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 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 inclusas 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* [19].

```
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 para os modelos de AM

Esta seção apresenta as variáveis de treinamento escolhidas para apresentação aos modelos de Aprendizado de Máquina supervisionados, as quais 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*.

A verificação da relevância das variáveis morfológicas seguiu a iniciativa apresentada por Miranda [20] 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 substantivos: dado pelo número de substantivos dividido pela quantidade total de *tokens* no comentário;
4. Número médio de pontuações: dado pelo número de pontuações 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 também em Alothali *et al.* [3], as informações escolhidas da plataforma *Instagram* para análise dos perfis 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 (embora mantenha 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 falsas 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. Já o número de seguidores e seguidos, por exemplo, podem indicar à busca de elevação da popularidade de usuários [6], e, portanto, são boas métricas de perfil.

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 MATERIAIS E MÉTODOS

Este capítulo descreve os materiais utilizados e a metodologia seguida ao longo deste trabalho. A Seção 3.1 contém a Revisão Sistemática da Literatura, consistindo na busca de publicações científicas na área de detecção de *bots* com uso de PLN e AM em redes sociais; a Seção 3.2 descreve o processo de coleta de dados na plataforma *Instagram*; a Seção 3.3 apresenta a aplicação de técnicas de PLN para o processamento dos comentários; a Seção 3.4 detalha o *script* desenvolvido para implementação dos modelos de AM.

3.1 Revisão Sistemática da Literatura

Para este trabalho foi realizada uma Revisão Sistemática da Literatura (RSL) a fim de verificar publicações científicas que envolvessem 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.

A primeira busca foi feita na plataforma Google Acadêmico. Com o uso das palavras-chave “*Processamento de Linguagem Natural*” and “*bot*”. Foram retornados 899 resultados, sendo selecionados dois: um Trabalho de Conclusão de Curso e um artigo, dentre os 20 primeiros cujos títulos e resumos mais se aproximavam da temática trabalhada.

Posteriormente foi realizada a pesquisa em Língua Inglesa com os termos “*Instagram*” and “*bot*” e “*Twitter*” and “*bot*”, com 30.900 e 727.000 resultados, respectivamente. A seleção foi dada pela leitura dos títulos dos 50 primeiros resultados para cada busca. Foram descartados trabalhos fora da temática buscada e também lidos 9 resumos para a primeira busca e 5 para a segunda. Foi então realizada a escolha de 4 artigos, de forma a englobar diferentes metodologias para a análise do problema.

Por fim, foi feita uma busca com os termos “*neural network*” and “*bot*” para verificar também exemplos de aplicações de redes neurais desenvolvidas para esse tipo de problema. Foi escolhido dentre os 20 primeiros dos 65.200 resultados retornados o artigo com título de maior relação à aplicação buscada.

Nessa etapa foram escolhidas ao todo 7 publicações, nas quais foram 6 artigos e um Trabalho de Conclusão de Curso que tratavam da temática de interesse. Dessas

publicações, 5 eram referentes ao estudo de *bots* na plataforma *Twitter* e 2 no *Instagram*. As informações dos artigos selecionados estão descritas na Tabela 1.

Tabela 1 – Artigos selecionados na revisão sistemática da literatura.

Nº	Autor/Ano	Revista\Evento	Idioma	Qualis
1.	Lorenceti e Salton (2022)	<i>Brazilian Journal of Development</i>	Português	C
2.	Rodríguez-Ruiz (2020)	<i>Computers & Security</i>	Inglês	A1
3.	Alothali <i>et al.</i> (2018)	<i>International Conference on Innovations in Information Technology (IIT)</i>	Inglês	Sem avaliação
4.	Akyon e Kalfaoglu (2019)	<i>Innovations in Intelligent Systems and Applications Conference</i>	Inglês	Sem avaliação
5.	Tunç <i>et al.</i> (2023)	<i>Journal of Polytechnic</i>	Inglês	Sem avaliação
6.	Kudugunta (2018)	<i>Information Sciences</i>	Inglês	A1

O primeiro artigo “*Detecção de fake news em um Tweet utilizando Machine Learning e Processamento de Linguagem Natural*” de Lorenceti e Salton [21] teve como objetivo detectar *fake news* no *Twitter* utilizando técnicas de PLN e AM, destacando durante o processo a presença de *bots* na propagação das notícias falsas.

A coleta dos textos foi feita por meio da API do *Twitter*. A limpeza das informações para deixá-las uniformes e contínuas incluiu as etapas de remoção de duplicidades e de dados incompletos. A classificação dos *tweets* foi feita de forma manual e os modelos de AM treinados foram a Árvore de Decisão, o KNN, o SGD, o Linear SVC e o SVC.

O referido trabalho não conseguiu alcançar o objetivo de classificar os *tweets*. Lorenceti e Salton [21] apresentam como possível causa a existência de apenas 8 *tweets* com informações falsas dentre os 344 do *dataset*.

O segundo artigo “*A one-class classification approach for bot detection on Twitter*” de Rodríguez-Ruiz *et al.* [22] 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. Rodríguez-Ruiz *et al.* [22], também pondera que, apesar da maior parte dos mecanismos propostos terem obtidos resultados satisfatórios, esses requerem dados rotulados de perfis automatizados para 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.

Nesse contexto, Rodríguez-Ruiz *et al.* [22] apresentam como a detecção de *bots* pode ser realizada sob uma perspectiva de classificação de classe única. Tal metodologia busca analisar apenas dados de usuários legítimos com a finalidade de encontrar comportamentos deviantes. Os resultados obtidos apontaram uma acurácia acima de 90%, indicando que classificadores desse tipo podem servir como uma forma inicial de detecção, capaz de perceber padrões de possíveis novos *bots*.

O **terceiro artigo** “*Detecting Social bots on Twitter: A Literature Review*” de Alothali *et al.* [3] apresenta uma visão geral do estado da arte na detecção de *bots* no *Twitter*, citando diferentes métodos utilizados. São apresentadas diversas metodologias comumente aplicadas, como métodos de AM supervisionados e não supervisionados, e o *crowdsourcing* que consiste na utilização da inteligência humana para diferenciar contas *bots* de usuários comuns, normalmente com o envolvimento de um grupo de indivíduos. Tal abordagem busca envolver a escolha de diferentes critérios de forma a justificar com certo grau de certeza a classificação dada.

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 citam, 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 [23].

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 [23].

Como conclusão, Alothali *et al.* [3] destacam o desafio de desenvolver algoritmos robustos e precisos devido aos mecanismos sofisticados utilizados pelos *bots* para evitar detecção. Além disso, pontuam sobre como o constante aumento do volume de dados em redes sociais torna ainda mais complexo a pesquisa nesse cenário.

O quarto artigo “*Instagram Fake and Automated Account Detection*” de Akyon e Kalfaoglu [6] apresenta algumas diretrizes para a verificação de contas falsas e contas *bots*. Características como a ausência de foto de perfil, nomes de usuários incomuns e confusos, com quantidade elevada de dígitos numéricos, por exemplo, são grandes indicativos para esses tipos de contas. Alto número de seguidos e baixo número de seguidores também podem se enquadrar nas características do grupo, visto que um dos principais objetivos dessas ferramentas é a manipulação de métricas da plataforma.

Akyon e Kalfaoglu [6] aplicaram sobre um *dataset* coletado do *Instagram* os modelos *Support Vector Classifier (SVC)*, *Naive Bayes*, *Regressão Logística* e redes neurais. Obtiveram como melhor resultado uma precisão de 91% e um F1-Score de 86% para o SVC.

O quinto artigo “*Classification of Fake, Bot, and Real Accounts on Instagram Using Machine Learning*” de Tunç *et al.* [24] desenvolve um processo utilizando PLN e AM para classificação de contas do *Instagram*, como falsas, *bots* ou controladas por usuários reais. É apresentada a possibilidade de abordagem do tratamento de forma diferente para perfis falsos e *bots* devido as suas especificidades.

Na seção de trabalhos correlatos, Tunç *et al.* [24] ponderam sobre mesmo com o fato de o *Instagram* possuir cerca de 3.4 vezes mais usuários ativos que o *Twitter*, a bibliografia de detecção de *bots* na primeira plataforma é muito mais limitada.

A metodologia da pesquisa seguiu alguns passos para classificação dos perfis. Inicialmente foi realizado o processo de *Web Scrapping* para coleta dos dados na rede social. Os perfis *bots* e perfis falsos foram obtidos de forma manual pela análise de comentários em postagens e diálogo com empresas que utilizam contas *bots*, respectivamente.

De resultados, o trabalho garantiu a criação de um *dataset* balanceado de perfis do *Instagram*, além de apresentar a performance de diferentes modelos de AM, destacando o *Random Forest* como o mais preciso. Como ponto de destaque, Tunç *et al.* [24]

abordam a importância de expandir a etapa de classificação dos perfis com ferramentas de *crowdsourcing* a fim de garantir maior confiança.

O **sexto artigo** “*Deep neural networks for bot detection*” de Kudugunta e Ferrara [9] demonstra a utilização de uma rede neural profunda na detecção de perfis automatizados e teve como principal objetivo verificar a possibilidade de classificar com boa precisão se um *tweet* foi postado por usuário comum ou um *bot*. Nesse trabalho foram coletados tanto os dados da plataforma quanto os conteúdos dos *tweets* na classificação dos perfis.

Como resultado foi obtido uma acurácia de 96% ao analisar apenas mensagens avulsas e de 99% ao incluir análises em nível de perfil. Kudugunta e Ferrara [9] concluem que tais resultados indicam a eficiência de abordagens de Aprendizado de Máquina para a resolução desse problema. Além disso, destacam a necessidade do desenvolvimento de métodos pouco custosos e eficientes de detecção, dado o grau de sofisticação dos *bots* atualmente.

O **Trabalho de Conclusão de Curso** “*Detecção de Bots no Twitter através de Técnicas de Processamento de Linguagem Natural*” de Miranda [20] apresenta 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*. Foram utilizadas diversas ferramentas como o PEGABOT², o algoritmo *Naive Bayes* e algumas bibliotecas para Processamento de Linguagem Natural como o pacote *spaCy* e o *LeIA* (*Léxico para Inferência Adaptada*).

O PEGABOT é um exemplo de ferramenta disponível *online* para a verificação de perfis em tempo real, 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.

Ao analisar dados textuais no *Twitter*, Miranda [20] 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

² <https://pegabot.com.br/>

bot, oito eram *emojis*, ao passo que esse valor caía para três para usuários reais. Já para o caso das métricas da plataforma, essas constituíram a maior diferença entre os dois tipos de perfis, sendo os com mais impacto o número de publicações e a idade da conta. A maior acurácia obtida após a aplicação do modelo foi de 77%.

Como segunda etapa da RSL, foi realizada também uma busca no site da Biblioteca Digital de Teses e Dissertações do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) com as palavras-chave “*Processamento de Linguagem Natural*” and “*bots*”, sendo encontrados apenas dois resultados: uma Tese de Doutorado e uma Dissertação de Mestrado, porém ambos foram descartados por não apresentarem o tema de interesse no título.

Em seguida a busca foi feita com as palavras-chave “*Aprendizado de Máquina*” and “*bots*”, porém, para ampliar os resultados da pesquisa, visto que foram retornados apenas 8 resultados, foi utilizado somente o termo “*bots*”. Tal busca retornou 54 Dissertações de Mestrado e 14 Teses de Doutorado, das quais foram descartadas 50 dissertações e todas as teses, pois não apresentavam o tema de interesse no título. Uma das dissertações não retirada nesta etapa estava com seu acesso indisponível no momento da consulta do texto, sendo também excluída. Após a leitura dos resumos das três restantes, todas foram selecionadas.

A **primeira Dissertação** “*Um modelo baseado em regras para a detecção de bots no Twitter*” de Leite [25] teve como principal objetivo a construção de um modelo para identificar e caracterizar, por meio de regras, o comportamento de *bots* e contas genuínas no *Twitter*.

Para a pesquisa foi utilizado modelos de AM supervisionados. Foi feita uma busca na literatura disponível para a escolha de uma base de dados já rotulada, devido à complexidade de se analisar manualmente uma quantidade muito grande de perfis e as limitações das APIs públicas de detectores de *bots*.

Ao examinar a aplicação tanto da Árvore de Decisão quanto do processo de indução de regras, Leite [25] pontua que além do alto poder de classificação de *bots*, essas técnicas também auxiliam na abertura de novas discussões acerca das principais características que os diferem das contas genuínas, como dados do perfil do usuário e seu comportamento.

A **segunda Dissertação** “*Deteccção de bots no Twitter: uma abordagem utilizando agrupamento*” de Lira [26] buscou propor uma nova abordagem para deteção de *bots* no *Twitter*, incorporando técnicas não supervisionadas junto às supervisionadas.

Lira [26] realizou uma revisão sistemática para verificar os métodos e técnicas existentes para deteção de *bots* na plataforma *Twitter*. Ao todo foram lidos 45 artigos. Como resumo do resultado do processo, esse destacou a grande popularidade do uso de AM para o referido problema, com maior foco da literatura nos modelos supervisionados. Também destaca a maior quantidade de pesquisas no *Twitter* e a existência de sistemas disponíveis para uso público.

Segundo Lira [26], a maior dificuldade no processo de deteção de *bots* ocorre na comparação dos resultados de diferentes soluções, pois os conjuntos de dados não seguem um critério claro na identificação do que é ou não um *bot*, visto que a verdade absoluta é em muitos casos impossível de se obter. Portanto, para ele, um sistema de uso geral de *bots* deve preferencialmente utilizar dados simples e ser aplicado sobre diferentes *datasets*.

Foi aplicado o algoritmo *k-Means* em três diferentes experimentos com combinações dos conjuntos de dados. Para verificar a solução foi realizada a comparação com uma aplicação do modelo *Random Forest*. Como principal contribuição, Lira [26] destaca a importância de os pesquisadores investirem certo tempo em experimentos no conjunto de dados, não inserindo todo o *dataset* no processo de treinamento, de forma a obter métricas mais realistas e evitar uma contribuição negativa no desempenho.

A **terceira Dissertação** “*Bots sociais: implicações na segurança e na credibilidade de serviços baseados no Twitter*” de Freitas [27] buscou identificar características de perfis *bots* no *Twitter*, como padrões de comportamento e estilo de escrita. Também foi realizada a criação de uma estratégia baseada em modelos de AM supervisionados sobre um *dataset* contendo 19.115 contas *bots* e 91.118 usuários legítimos.

Para a análise das características de perfil foram aplicadas funções estatísticas de forma a realizar a verificação de cada variável. Já para os padrões de escrita, Freitas [27] fez o cálculo de diferentes métricas, como tamanho do vocabulário de cada usuário e a diferença entre dois *tweets* por meio do cálculo da distância do cosseno.

3.2 Coleta dos dados

A coleta dos dados foi realizada por meio da recuperação manual de perfis e comentários em postagens de lojas de produtos de Informática. O acesso foi feito com o uso da ferramenta de exploração do *Instagram*, que permite ao usuário visualizar uma listagem contínua de *posts*, e da visualização de *stories* patrocinados, contendo publicidades de diferentes lojas e produtos. Exemplos de ambas as funcionalidades estão ilustrados na Figura 4.



Figura 4 – Exemplo da ferramenta de explorar e *stories* patrocinado.

As contas classificadas como *bots* foram retiradas de comentários em postagens de lojas encontradas unicamente por meio de propagandas patrocinadas da plataforma e que indicavam baixa confiabilidade.

Diversas das lojas apresentadas normalmente anunciavam produtos com valores muito abaixo do esperado pelo mercado e com descontos elevados para pagamento via Pix, além de realizar todo seu processo de venda por meio da disponibilização de um link da plataforma *Whatsapp*.

Após uma maior investigação das postagens feitas pelas lojas, da interação nos comentários e de uma busca de maiores dados das empresas foi possível notar a existência de possíveis tentativas de golpes.

Devido ao custo elevado de tempo para se realizar o processo de análise dos perfis e classificação desses, o ambiente encontrado facilitou a interpretação de um perfil como *bot* ou não com maior grau de certeza.

Durante a análise dos comentários e seus respectivos usuários foi percebido um baixo nível de autenticidade nas interações. Os comentários seguiam um padrão entre as diferentes postagens e às vezes sendo realizados pelos mesmos perfis de forma repetida.

Já as contas reais foram acessadas, principalmente, de páginas de lojas conhecidas pelo autor deste trabalho na cidade de Londrina, mas também foram verificadas lojas de outras cidades nas quais foi possível confirmar a sua veracidade por relatos oriundos do círculo social do autor e pela busca dos dados da empresa.

Não foram coletadas interações (respostas aos comentários de outros usuários) e sim verificados apenas os comentários principais, direcionados diretamente à postagem. Conteúdos que faziam referência aos usuários da plataforma, como menções (incluir no comentário o perfil de outra pessoa por meio do uso do caractere @ seguido do nome de usuário) foram removidos a fim de manter a privacidade.

Ao todo foram coletados 101 comentários e perfis para o modelo, sendo 71 reais e 30 *bots*. Não houve distinção no processo de seleção dos comentários, sendo esses escolhidos de maneira aleatória. O corpus contendo todos os comentários está disponível no Apêndice A. Os dados foram armazenados em um arquivo *xlsx* no qual cada coluna corresponde a uma das variáveis selecionadas.

3.2.1 Classificação dos perfis

A classificação dos perfis como *bots* foi feita pela união das respostas obtidas após a apresentação dos dados coletados a três estudantes do curso de Ciência da Computação da UEL, incluso o próprio autor.

Dado a importância de se rotular corretamente os perfis para o bom resultado do modelo, foram descartados do *dataset* os perfis nos quais não houve unanimidade de classificação. Por envolver apenas três análises diferentes, incluir tais perfis não oferecia um grau de confiabilidade suficiente. A escolha do local de coleta de cada tipo de perfil buscou simplificar o processo de classificação, garantindo resultados mais confiáveis.

Os perfis foram avaliados com base em diferentes critérios. As contas classificadas como *bots* e que possuíam um perfil público permitiram a verificação de um

número maior de características que justificassem tal apontamento. Em quase todos os casos possuíam fotos aparentemente reais, mas essas, normalmente, tinham todas as mesmas datas de postagem na plataforma, indicando a busca por estabelecer rapidamente um perfil para utilização.

Uma das contas analisadas, por exemplo, possuía 38 fotos, porém todas haviam sido postadas no dia 26 de janeiro de 2024 e não envolviam qualquer interação que pudessem auxiliar na verificação de relações comuns da rede. A lista de curtidas das postagens incluía usuários sem foto de perfil, dígitos no nome de usuário e caracteres de outro alfabeto, características de alto indicativo de contas falsas utilizadas unicamente para aumento de métricas, conforme apontado por Akyon e Kalfaoglu [6].

Dados como quantidade de seguidores e seguidos possuíam um padrão característico e os comentários realizados normalmente envolviam perguntas com relação à loja, para serem respondidas de forma a induzir confiança. Um exemplo de pergunta observada foi “a loja possui CNPJ?”, prontamente respondida pela loja com um registro que apesar de aparentemente real era constituído de informações falsas, cujo fato foi percebido após uma verificação mais criteriosa.

3.3 Processamento dos comentários

O processamento dos dados textuais foi realizado por meio da biblioteca *spaCy*. Foi desenvolvido um script em *Python* para leitura dos comentários do arquivo *xlsx*, tratamento e processamento pelo *spaCy*.

Para a apresentação do texto ao *spaCy* abreviações foram manualmente alteradas para sua forma real. Foram analisados aspectos como quantidade de palavras, uso de verbos, adjetivos, substantivos e demais classes apontadas por essa ferramenta. A quantidade média de emojis em cada comentário foi inserida de maneira manual no conjunto de dados. Por fim, foi feita a verificação de termos populares para cada tipo de perfil.

As variáveis obtidas da análise textual foram repassadas para os algoritmos de AM.

3.4 Algoritmos de Aprendizado de Máquina

Foi feita a aplicação de modelos supervisionados *Random Forest* e *k-NN*, 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].

A primeira etapa para o algoritmo foi a criação de um *script* em *Python* para a aplicação dos modelos e a manipulação dos dados. O arquivo *xlsx* contendo as colunas correspondentes a cada *feature* do *dataset* teve sua leitura realizada com auxílio da biblioteca *pandas*³, uma ferramenta de código aberto de análise e manipulação de conjunto de dados feita em linguagem *Python*.

A biblioteca de aprendizado de máquina em *Python* e *scikit-learn* foi utilizada para a implementação dos modelos supervisionados *Random Forest* e *k-NN* e para o cálculo das métricas de avaliação do modelo.

O algoritmo desenvolvido neste trabalho e o conjunto de dados coletado foi disponibilizado na plataforma *Github*⁴.

³ <https://pandas.pydata.org/>

⁴ <https://github.com/kristianok/Script-AM-e-PLN>

4 RESULTADOS

Neste capítulo serão apresentadas as análises dos resultados obtidos após a aplicação dos algoritmos sobre o *dataset* deste trabalho. Foram incluídas a análise da morfologia dos comentários e das métricas da plataforma *Instagram* e as estatísticas de performance dos modelos de AM.

4.1 Análise morfológica

A análise morfológica consistiu na verificação das palavras presentes nos comentários e das classes gramaticais apontadas pelo *spaCy*. Conforme apresenta a Figura 5, foi possível notar algumas semelhanças e diferenças nos trechos escritos por *bots* e usuários reais.

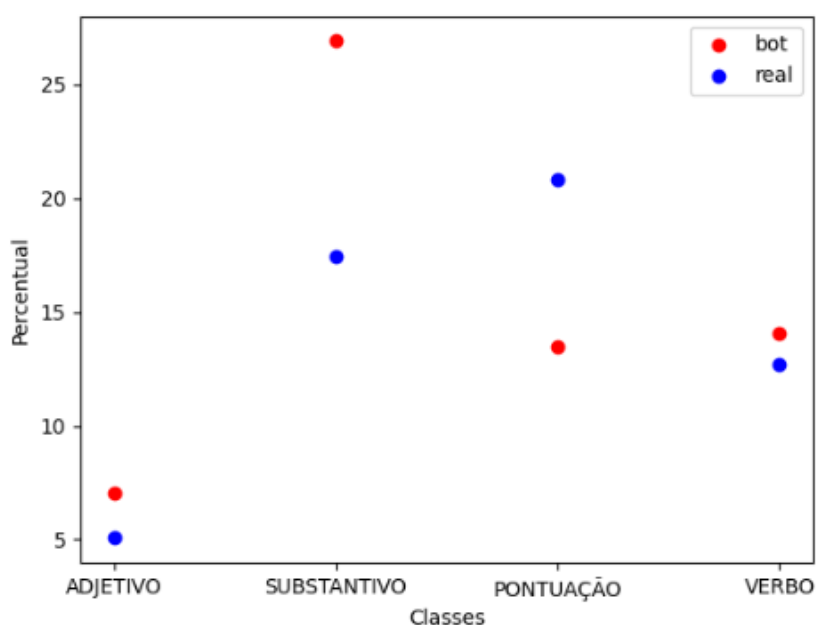


Figura 5 – Percentual de classes gramaticais nos comentários.

A ocorrência de verbos e adjetivos de ambos os perfis teve valores muito próximos. Para os adjetivos houve uma proporção de 7% para *bots* e 5% para os usuários reais e para os verbos um percentual de 12,5% e 14%, respectivamente.

No caso de substantivos e pontuação no texto, foi percebida uma maior diferença. Os primeiros constituíram 27% das palavras utilizadas pelos *bots*, em contraste a 17% para os usuários reais. Já para a pontuação tal cenário se inverte, um total de 13% dos tokens para os *bots* e 21% para os reais.

Além da maior quantidade de pontuação presente nos comentários dos usuários reais foi possível perceber uma característica que justifica tal diferença. Ambos os tipos de perfis incluíam perguntas a respeito da loja e dos produtos anunciados. No entanto, trechos de comentários dos usuários reais como “Onde fica ???” e “Super indico!!!! Rapidez, comprometimento e qualidade!!!!”, indicaram a utilização da repetição de pontos de exclamação ou interrogação para enfatizar dúvidas a respeito da loja ou entusiasmo na realização de uma compra. Comportamento não percebido nos comentários de *bots*.

Outra diferença encontrada foi a quantidade de emojis utilizada por cada um dos perfis. Apenas um comentário dentre os trinta de usuários *bot* incluiu o uso de *emojis*, ao passo que para os clientes reais estes estavam presentes em 23 dos 71 comentários, envolvendo também a repetição sequencial para maior ênfase. Os emojis mais utilizados foram o rosto apaixonado (😍), fogo (🔥) e palmas (👏). Tal informação representou uma diferença considerável dos resultados encontrados por Miranda [20], visto que em sua pesquisa o uso de emojis por usuários *bots* era consideravelmente maior.

Com relação aos termos mais comuns na Figura 6 e Figura 7 encontram-se nos gráficos de barra as 15 palavras mais utilizadas por *bots* e usuários reais, respectivamente. Temos que para ambos o termo mais comentado foi “loja”, relacionado diretamente às perguntas e avaliações a respeito dos produtos anunciados.

Uma palavra de destaque para os usuários reais é “valor”. Em lojas reais é comum postagens serem constituídas apenas por comentários a respeito do preço, visto que em muitos casos estes não são divulgados abertamente e se trata de um tópico de interesse para os clientes da loja.

No caso das lojas falsas, dado que uma das táticas de seus anúncios é colocar um preço muito abaixo como ponto de destaque para atrair usuários, esse termo não teve aparição considerável em seus comentários.

Por fim, o estrangeirismo “*top*” ao fazer uma observação sobre os produtos foi o terceiro termo mais comum nos comentários de usuários legítimos, mas não apareceu nos trechos escritos por *bots*.

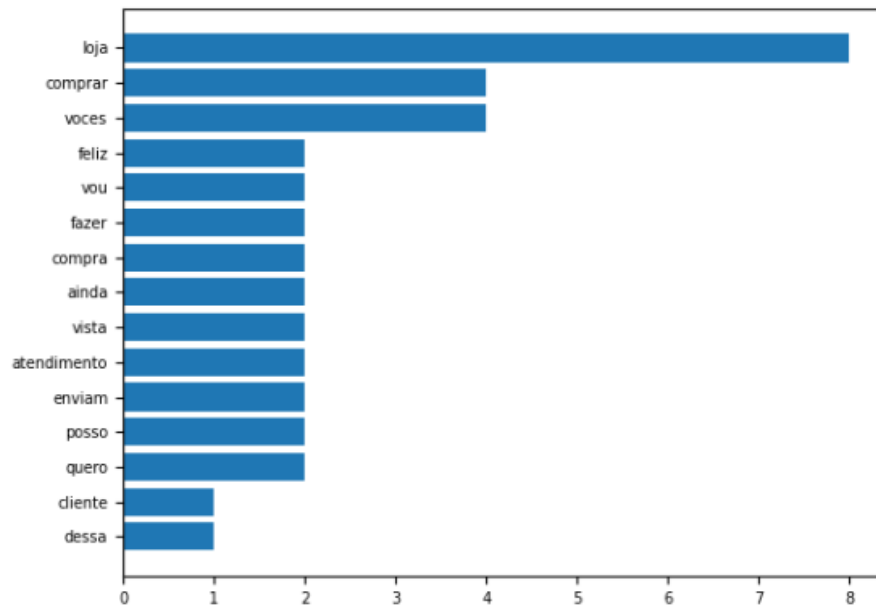


Figura 6 – Palavras mais escritas por bots.

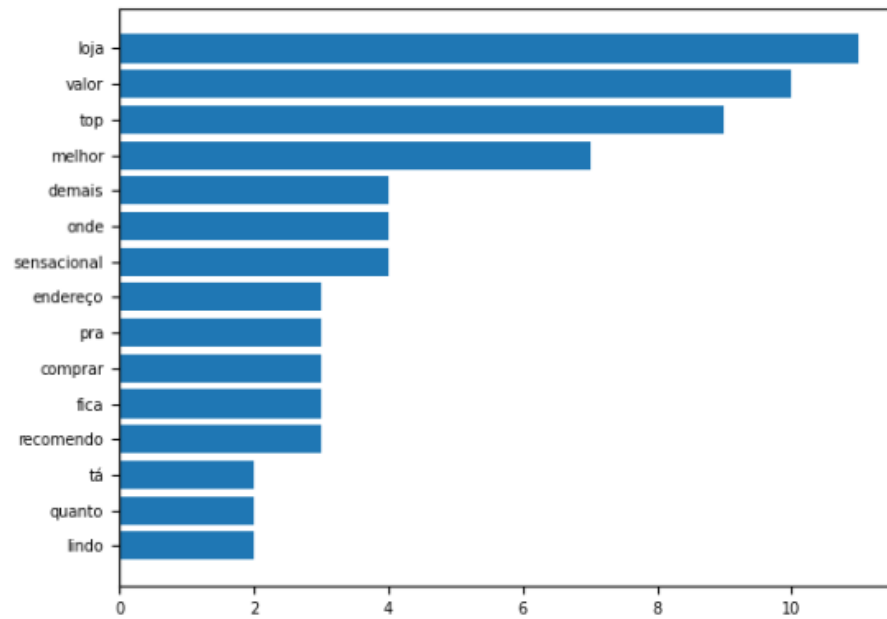


Figura 7 – Palavras mais escritas por usuários reais.

4.2 Influência das variáveis

Para verificar a influência das variáveis no *dataset*, a quantidade de dígitos numéricos no nome de usuário não foi uma variável com relevância para a diferenciação entre

os perfis. Dos usuários reais, 4% apresentaram números no nome enquanto para *bots* esse valor era de 6%. Tal padrão indica um maior nível de detalhamento da elaboração dos perfis *bots*, dado que, como apresentado por Akyon e Kalfaoglu [6], essa variável tem maior impacto na análise de contas falsas, visto que, diferentemente dos *bots*, são voltadas unicamente ao impulsionamento de métricas de outros usuários. O mesmo foi percebido na análise da presença de foto de perfil, apenas 4 de todos os perfis não a possuíam, sendo desses 1 *bot* e 3 não *bots*.

Em média, a quantidade de curtidas nos comentários de ambos os tipos de perfis também se manteve próxima, conforme observado na Figura 8.

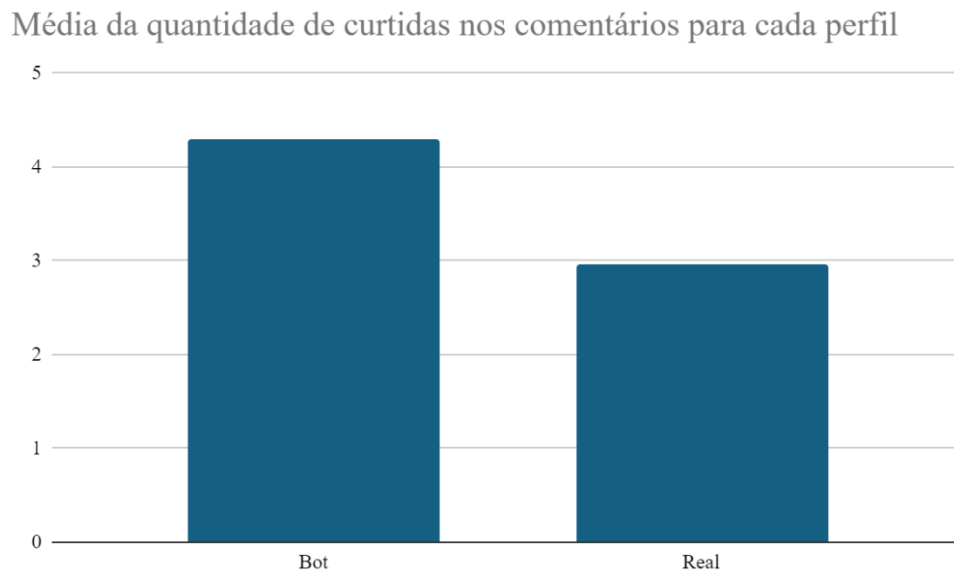


Figura 8 – Média da quantidade de curtidas nos comentários para cada perfil.

Em relação à idade da conta, os dados encontrados seguiram os padrões apontados por Miranda [20] no *Twitter*. Todas as contas *bots* haviam sido criadas após o ano de 2020, indicando uma idade muito próxima entre esses perfis. Além disso, a maior parte teve seu cadastrado realizado em 2023, ou seja, haviam ingressado muito recentemente na plataforma. O gráfico contendo a comparação da quantidade de perfis *bots* criados em cada ano está ilustrado na Figura 9.

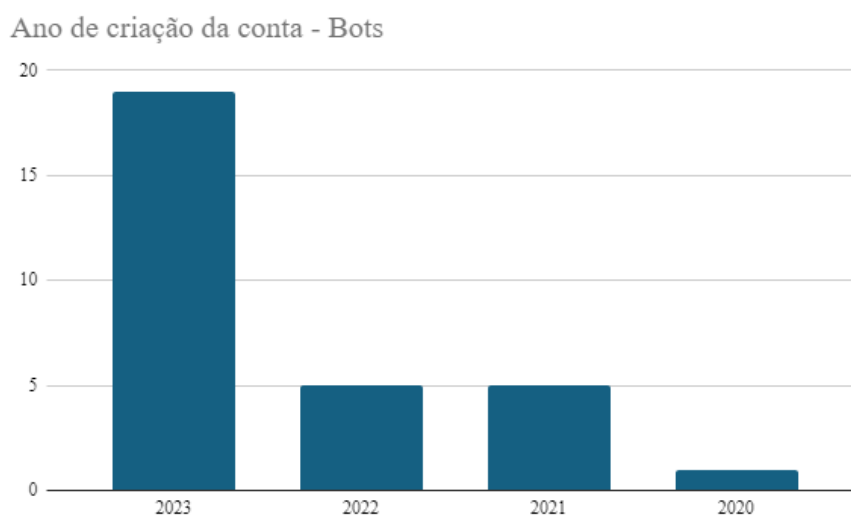


Figura 9 – Gráfico do ano de criação das contas *bots*.

Para os usuários reais, a variação no ano de criação da conta foi muito mais expressiva, conforme ilustrado na Figura 10. A presença de anos anteriores a 2020 foi algo comum. O ano de 2013 apareceu a maior quantidade de vezes, destacando uma maior idade desse tipo de perfil.

De maneira geral, usuários reais apresentam idade muito variadas quando comparadas aos *bots*, sendo essa uma informação esperada. Contas automatizadas são constantemente desenvolvidas para a execução de novas tarefas. Além disso, podem não apresentar um ciclo de vida muito longo devido a constantes restrições de acesso exercidas pela plataforma por violarem os termos da plataforma.

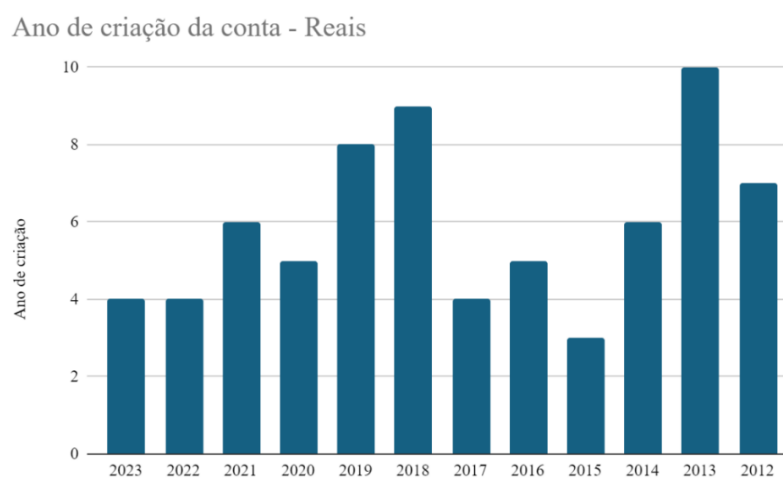


Figura 10 – Gráfico do ano de criação das contas de usuários reais.

A privacidade da conta indicou uma baixa quantidade de contas *bots* privadas. Conforme apresenta a Figura 11, 25 das 30 contas *bots* eram públicas.

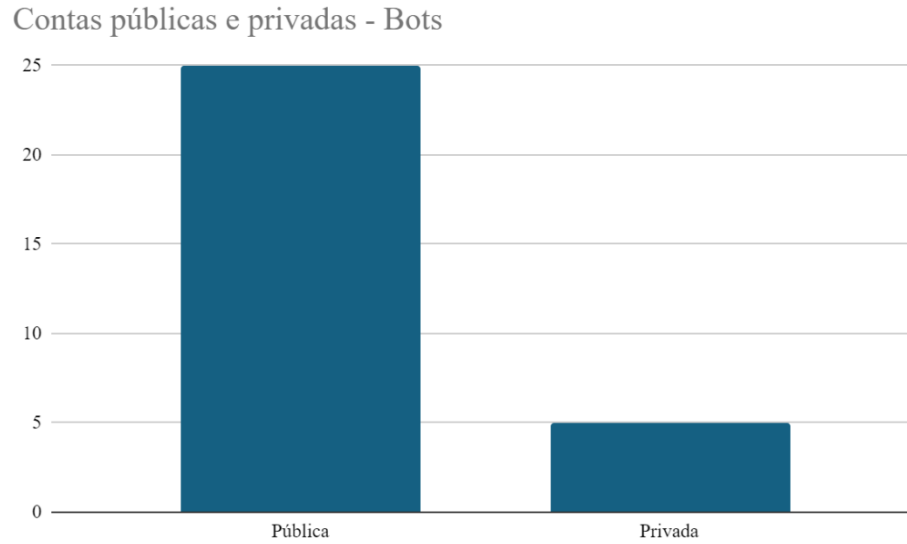


Figura 11 – Gráfico contendo a quantidade de perfis *bots* privados e não privados.

No caso dos perfis legítimos, a diferença entre a quantidade de contas privadas e públicas foi menor. Por meio da Figura 12 é possível perceber a proporção semelhante de ambas as configurações.

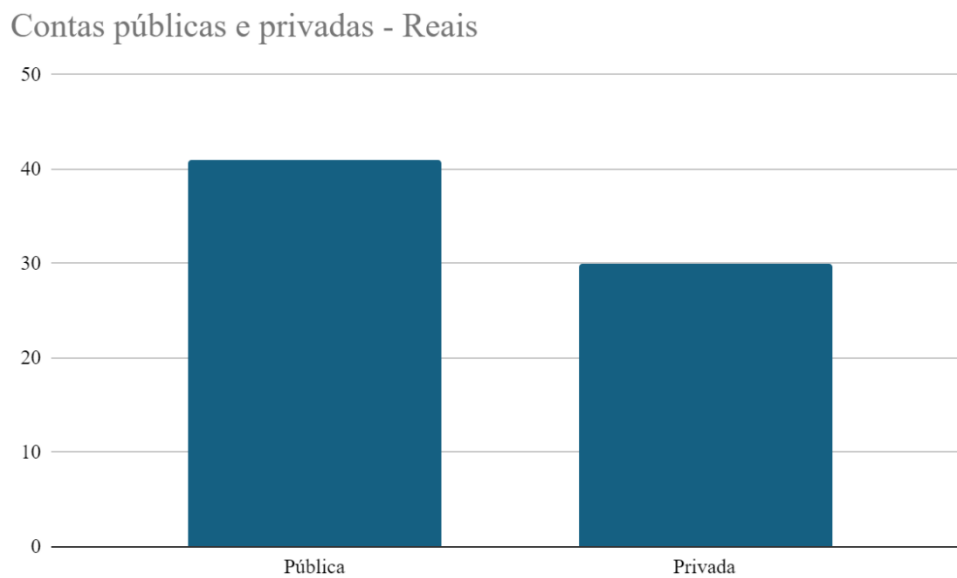


Figura 12 – Gráfico contendo a quantidade de perfis reais privados e não privados.

Os perfis *bots* apresentaram uma quantidade considerável de postagens, sendo 58 a quantidade máxima de envios feitos por esse tipo de perfil e 24 a média de *posts* realizados.

Para usuários reais, esse número seguiu uma quantidade menos padronizada. Os perfis coletados incluíram desde usuários com nenhuma postagem até contas com mais de 1000 envios. Normalmente, essas últimas representavam figuras de algum nível de influência local, com expressividade considerável nas redes e que se interessavam pelas lojas da região.

A quantidade de caracteres na “*bio*” (seção do perfil contendo uma descrição do usuário, normalmente envolvendo informações pessoais ou profissionais relevantes) dos dois tipos de perfis se manteve próximo. Além disso, durante as observações no processo de montagem do conjunto de dados foi percebido que o conteúdo textual da “*bio*” envolveu dados como cidade, nome, interesses e descrições pessoais e esses seguiram um padrão semelhante para *bots* e não *bots*.

Essa observação indica o empenho desse tipo de *bot* na busca por mascarar sua real finalidade. A postagem de fotos aparentemente reais e a abertura do perfil como público teve como finalidade passar maior credibilidade aos usuários, visto que a confirmação da veracidade dos clientes é normalmente feita pela verificação rápida dos perfis referentes às avaliações nas postagens da loja, com destaque a visualização dos *posts*, foto de perfil e “*bio*” desses usuários.

Como último ponto de análise foi verificado o número de seguidos e seguidores das contas, sendo percebidos valores muito característicos para cada tipo de perfil. A média do número de seguidos encontrado foi de 43 para as contas *bots* e 1.200 para os usuários reais. O baixo valor para os primeiros foi contrário aos apontados por Miranda [20] e Tunç *et al.* [24]. No entanto, estava em concordância com os interesses desse tipo de *bots*. Não eram utilizados para aumentar as métricas numéricas de outras contas, mas buscavam ter certa relevância para seus comentários com o objetivo de aumentar o grau confiança para a loja na perspectiva dos outros usuários.

Com relação ao número de seguidores, os *bots* possuíam um valor muito elevado e destoante com relação ao seu número de seguidos. Para os não *bots*, o número médio de seguidores era de 1.600. Com tais resultados foi possível verificar um balanceamento desses valores em perfis reais. Já no caso dos *bots* notou-se o contrário, uma quantidade muito maior de seguidores em comparação aos usuários seguidos. Dos 30 perfis

bots analisados, 19 seguiam menos de 30 outros perfis da plataforma e possuíam em média 18 mil seguidores.

4.3 Modelos de Aprendizado de Máquina

O conjunto de treinamento do modelo correspondeu a aproximadamente 80% dos dados do *dataset*. Tal parcela continha 58 perfis reais e 22 perfis *bots*. O restante dos 21 perfis foi aplicado na etapa de validação da eficiência de cada algoritmo.

A separação dos dados foi feita utilizando a função “*train_test_split()*” da biblioteca *sklearn* com os argumentos “*test_size=0.2*” e “*random_state=0*”. O último foi utilizado para permitir a replicabilidade dos testes, garantindo a mesma divisão de dados nas execuções.

Com a realização da análise da relevância das variáveis coletadas foram organizadas duas configurações diferentes dos dados para apresentação ao algoritmo.

Para o primeiro modelo foram utilizadas todas as colunas do *dataset*. No segundo modelo, as variáveis “Dígitos numéricos no nome de usuário”, “Possui foto de perfil?”, “Quantidade de caracteres na *bio*”, “Quantidade de curtidas do comentário” e número médio de adjetivos e verbos foram descartadas, visto que sobre uma primeira análise indicavam ter pouca influência na classificação dos perfis. Tal configuração buscou permitir a confirmação da relevância dessas variáveis também pelo modelo.

Os resultados obtidos para os dois modelos podem ser visualizados na Tabela 2 e Tabela 3.

Tabela 2 – Métricas estatísticas de avaliação do primeiro modelo

Modelo	Acurácia	<i>Recall</i>	<i>F1-Score</i>	Precisão
k-NN	0.9	0.875	0.875	0.875
Random Forest	1.0	1.0	1.0	1.0

Tabela 3 – Métricas estatísticas de avaliação do segundo modelo

Modelo	Acurácia	<i>Recall</i>	<i>F1-Score</i>	Precisão
k-NN	0.9	0.875	0.875	0.875
Random Forest	1.0	1.0	1.0	1.0

O melhor modelo encontrado correspondeu ao classificador *Random Forest*. Tal resultado seguiu o padrão esperado conforme apontado pela biografia [3].

Não foi possível verificar por meio da comparação da predição dos dois modelos a baixa relevância das variáveis destacadas, visto que o descarte dessas não ocasionou na melhora do processo de classificação do algoritmo *k-NN* ou na piora na classificação do *Random Forest*.

No caso do *Random Forest*, ambos os modelos foram capazes de predizer corretamente todos os perfis *bots*. Já o *k-NN* foi capaz de classificar corretamente apenas 90% dos perfis.

Apesar de o algoritmo *Random Forest* obter 100% de acurácia na classificação de um usuário como *bot* ou não, o que é um valor incomum para esse tipo de problema, é importante destacar a uniformidade dos perfis *bots* coletados. A baixa quantidade de dados e o ambiente restrito favoreceram bons resultados, porém com um uso geral muito limitado. Se aplicados a outros conjuntos de dados, os modelos treinados podem apresentar uma piora significativa de performance.

5 CONSIDERAÇÕES FINAIS

O combate ao uso de *bots* é um problema complexo e a utilização desse tipo de ferramenta de forma maliciosa nas redes sociais traz grandes repercussões negativas. O uso mal intencionado da plataforma *Instagram* para a aplicação de golpes na venda de produtos analisado por este trabalho representa uma situação que compromete a segurança dos usuários e a organização natural da rede.

O objetivo proposto por este trabalho de desenvolver um modelo de Aprendizado de Máquina capaz de classificar um perfil do *Instagram* como *bot* foi alcançado. No entanto, não é possível confirmar a classificação das contas coletadas como uma verdade absoluta, dado que o processo de *crowdsourcing* realizado por este trabalho não garante total confiança, havendo a possibilidade da existência de classificações incorretas no conjunto de dados.

Além disso, devido à necessidade da coleta e classificação manual dos dados, a quantidade de registros no *dataset* foi limitada, sendo restrita à temática de postagens de lojas de equipamentos de Informática.

Foi possível apontar padrões para cada tipo de conta com base na análise dos comentários e das variáveis coletadas da plataforma. Dados como o número de seguidores e seguidos, e idade da conta indicaram diferenças significativas entre os perfis *bots* e reais. Percebeu-se durante o tratamento do texto a discrepância de comportamento no uso de pontuação e emojis e os termos mais comuns.

Os resultados encontrados, apesar de serem voltados a apenas uma pequena parcela do ambiente nos quais os *bots* estão inseridos, buscou complementar a investigação de sua presença no *Instagram*, com foco em perfis em Língua Portuguesa.

Tendo em vista que a aplicação de golpes por meio de anúncios patrocinados de lojas falsas se tornou uma prática comum na plataforma, as análises realizadas por este trabalho incluíram diretrizes importantes para a verificação de contas legítimas nesse ambiente.

Para a realização de pesquisas futuras alguns pontos de expansão deste trabalho podem ser considerados, dentre eles a ampliação do conjunto de dados com mais perfis, conforme apresentado por Lira [26], é essencial para um algoritmo de uso geral de detecção de *bots*. Também se destaca a possibilidade de uma pesquisa mais detalhada dos comentários, verificando elementos sintáticos e semânticos, por exemplo.

Por fim, com o objetivo de formar de conjuntos de dados mais confiáveis e com maior variedade para uso no desenvolvimento de algoritmos e análises, pode-se realizar o *crowdsourcing* em maior escala, envolvendo mais pessoas e temáticas no processo de classificação dos perfis da plataforma *Instagram*.

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., ZAKI, N.; MOHAMED, E. A.; ALASHWAL, H. Detecting Social Bots on Twitter: A Literature Review. *Proceedings of International Conference on Innovations in Information Technology, 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. 28p.
- [6] AKYON, F. C.; ESAT KALFAOGLU, M. Instagram Fake and Automated Account Detection. *Proceedings of Innovations in Intelligent Systems and Applications Conference, 10.*, Izmir, 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, Dec. 2021.
- [8] FARIA, C. R. *Ferramenta Carolina para identificação de pragas e doenças na cultura da soja utilizando Processamento de Linguagem Natural*. Programa de Pós-Graduação em Ciência da Computação. Universidade Estadual de Londrina. Dissertação de Mestrado em Ciência da Computação. Londrina, 2021. 163p.
- [9] KUDUGUNTA S, E. F. Deep neural networks for bot detection. *Information Sciences*, v.

467, n. 1, p. 312-322, Oct. 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*. Programa de Pós-Graduação em Ciência da Computação. Universidade Estadual de Londrina. Dissertação de Mestrado em Ciência da Computação. Londrina, 2020. 70p.
- [12] MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre Aprendizado de Máquina. In: REZENDE, S. O. (Ed.) *Sistemas inteligentes: fundamentos e aplicações*. 1ª. ed. Barueri: Manole, v. I, 2003. p. 89-114.
- [13] MYLES, A. J.; FEUDALE R. N.; LIU, Y.; WOODY N. A.; BROWN S. D. An introduction to decision tree modeling. *Journal of Chemometrics*, v. 18, n. 6, p. 275–285, June 2004.
- [14] MITCHELL, T. M. *Machine Learning*. 1. ed. New York: McGraw-Hill Education, 1997.
- [15] BREIMAN, L. Random Forests. *Machine Learning*, California, v. 45, n. 1, p. 5-32, Oct. 2001.
- [16] CUNNINGHAM, P.; DELANY, S. J. k-Nearest Neighbour Classifiers - A Tutorial. *ACM Computing Surveys*, v. 54, n. 6, p. 1-25, July 2021.
- [17] MILLER, Z.; DICKINSON B.; DEITRICK W.; WEI H. Twitter spammer detection using data stream clustering. *Information Sciences*, v. 260, p. 64-73, Mar. 2014.
- [18] ERICKSON, B. J.; KITAMURA, F. Magician's Corner: 9. Performance Metrics for Machine Learning Models. *Radiology Artificial Intelligence*, v. 3, n. 3, p. e200126, May 2021.
- [19] 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.
- [20] MIRANDA, G. A. *Detecção de Bots no Twitter Através de Técnicas de Processamento de Linguagem Natural*. Instituto de Matemática e Estatística, Universidade Federal Fluminense. Trabalho de Conclusão de Curso em Estatística. Niterói, 2022. 48p.
- [21] LORENCETI, A. D.; SALTON, G. D. Detecção de fake news em um tweet utilizando machine learning e processamento de linguagem natural. *Brazilian Journal of Development*, v. 8, n. 6, p. 43581-43599. Maio. 2022.

- [22] RODRÍGUEZ-RUIZ, J. et al. A one-class classification approach for bot detection on Twitter. *Computers & Security*, v. 91, n. 101715, Apr. 2020.
- [23] SUBRAHMANIAN, V. S.; AZARIA, A.; DURST, S.; Kagan, V.; GALSTYAN, A.; LERMAN, K.; ZHU, L.; FERRARA, E.; FLAMMINI, A.; MENCZER, F. The DARPA *Twitter Bot Challenge*. *Computer*, v. 49, n. 6, p. 38-46, June 2016.
- [24] TUNÇ, Ü.; ATALAR, E.; GARGI, M. S.; ERG ÜL A. Z. Classification of Fake, Bot, and Real Accounts on Instagram Using Machine Learning. *Journal of Polytechnic*, v. 27, n. 2, p. 479-488. Dec. 2023.
- [25] LEITE, M. A. G. L. *Um modelo baseado em regras para a detecção de bots no Twitter*. Programa de Pós-Graduação em Educação. Universidade Federal dos Vales do Jequitinhonha e Mucuri. Dissertação de Mestrado em Educação. Diamantina, 2019. 74p.
- [26] LIRA, D. B. *Detecção de bots no Twitter: uma abordagem utilizando agrupamento*. Programa de Pós-Graduação em Sistemas de Informação. Escola de Artes, Ciências e Humanidades, Universidade de São Paulo. Dissertação de Mestrado em Sistemas de Informação. São Paulo, 2022. 67p.
- [27] FREITAS, C. A. S. de. *Bots sociais: implicações na segurança e na credibilidade de serviços baseados no Twitter*. Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas, Universidade Federal de Minas Gerais. Dissertação de Mestrado em Ciências. Belo Horizonte, 2014. 82p.

APÊNDICE A – CORPUS COMENTÁRIOS

feliz em tambem ser cliente dessa loja
muito obrigado pela atencao no whatsapp
qual valor do 15 pro max ?
vou fazer minha compra semana que vem espero estar de promoçao ainda
feliz com a compra de hoje!
a loja tem cnpj ?
Possuem nota fiscal?
Qual o prazo de entrega?
Como faço para comprar?
Fica quanto à vista?
Qual o endereço da loja ?
Loja incrível 🐱 amei o atendimento
Promoção por tempo limitado?
Tem desconto para pagamento à vista?
Ainda disponível?
O custo benefício desse aparelho e excelente!
o atendimento de voces sao otimo!
parabens pela grande loja
voces enviam para o rj ?
posso fazer retirada ?
quais as formas de pagamentos ?
qual horario de funcionamento da loja ?
a loja esta aberta?
posso ir ate a loja ?
voces enviam o copo de brinde ?
amiga linda !!
quero comprar
valores ?
nao me canso de babar no feed de voces assim que meu salario cair vou comprar
quero comprar em atacado
qual endereço?
Qual o processador e o valor dele
Tá esse preço mesmo?
Top de mais
Tá de quanto
Qual o valor ? 😞
Quanto
só para tirar uma duvida vocês entregam teclados?
Lindo demais! 💎
Ótima loja, comprei uma placa de vídeo e já vai fazer 3 anos que estou com ela, funcionando perfeitamente.

Amei!! 🤝❤️
O melhor do mercado!
Ótimo atendimento!! Obrigado pela atenção 🤝🤝
Amei, muito obrigada 😊👉🤝
Top demais, atendimento impecável 📈
Olá, como faço pra comprar
Muito obrigada pelo carinho ❤️
Ainda tem?
De onde vocês são ?
Qual o endereço da loja?
Endereço da loja física ?
Quero
Boa noite. onde fica a loja
Excelente
A melhoorr 😊🤝🤝
Loja top! Recomendo!!
Loja top sensacional a melhor do Brasil!
Melhor lugar pra comprar Apple
Qual valor ?
Qual o valor
Quero esse 14 pro max! 🙏 qual o valor?
Vocês são sensacionais 🤝
Vcs entrega??
Valor? É lacrado? Tem garantia?
não conhecia essa loja, ameiii
Pode me passar as formas de pagamento?
Novo?
Top o melhor 💧
Super indico!!!! Rapidez, comprometimento e qualidade!!!!
Sensacional!! 😊
O melhor!!!
Onde fica sua loja
Show 😊
Valor?
Chamei no WhatsApp
Perfeitossss ❤️❤️❤️❤️
Qual valor ??
🤝Toppp
Tem na loja ?
Sensacional 💧
Muito bom, só notebook de qualidade! Já pretendo compra um com você, sucesso!
A placa é sensacional
op demais,estou curtindo muito o meu
Tenho interesse. Qual o valor ?

Uma excelente loja, recomendo.
Os melhores 💧 🙌
Mouse espetacular, recomendo demais!
Melhor Loja, melhores equipamentos e preços maravilhosos!! 😄 😄
Top 🙌
Linda de mais 😄 😄 😄 😄
Onde fica ???
olha isso, espera alguns anos que eu compro.
Top heim, quantos sai uma dessa?
a maior 💧 💧
Tem o link da loja de vcs ?
Quem é da cidade vai conseguir retirar na loja?
não achei na lista para comprar.
Opa manda pra ca 😄
Muito lindo esse gabinete. 😄
Top em
Qual valor?