

Patricia Mello Marçal Pinto

**Uma metodologia de construção de uma base
de dados de perfis do Twitter rotulados em
bot ou em legítimo**

Niterói - RJ, Brasil

20 de dezembro de 2022

Patricia Mello Marçal Pinto

**Uma metodologia de construção de
uma base de dados de perfis do
Twitter rotulados em bot ou em
legítimo**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dra. Karina Yuriko Yaginuma

Niterói - RJ, Brasil

20 de dezembro de 2022

Patricia Mello Marçal Pinto

**Uma metodologia de construção de uma base
de dados de perfis do Twitter rotulados em
bot ou em legítimo**

Monografia de Projeto Final de Graduação sob o título “*Uma metodologia de construção de uma base de dados de perfis do Twitter rotulados em bot ou em legítimo*”, defendida por Patricia Mello Marçal Pinto e aprovada em 20 de dezembro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Karina Yuriko Yaginuma
Departamento de Estatística – UFF

Profa. Dra. Patrícia Lusié Velozo da Costa
Departamento de Estatística – UFF

Profa. Dra. Jessica Quintanilha Kubrusly
Departamento de Estatística – UFF

Niterói, 20 de dezembro de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

P659m Pinto, Patricia Mello Marçal
Uma metodologia de construção de uma base de dados de perfis do Twitter rotulados em bot ou em legítimo / Patricia Mello Marçal Pinto. - 2022.
55 f.: il.

Orientador: Karina Yuriko Yaginuma.
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Base de dados rotulada. 2. Bots. 3. Twitter. 4. Contas honeypots. 5. Produção intelectual. I. Yaginuma, Karina Yuriko, orientadora. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD - XXX

Resumo

O Twitter é uma plataforma que fornece uma estrutura de uso simples e permite fácil acesso às pessoas, favorecendo à utilização de perfis automatizados, os *bots* sociais, que são contas de mídia social controladas totalmente ou parcialmente por algoritmos de computador. Os *bots* sociais podem ser benignos, quando são inofensivos, ou seja, não são poluidores de conteúdo, ou podem ser maliciosos, quando se passam por usuários humanos com o objetivo de manipular e poluir conteúdo. Além disso, a natureza automatizada dos *bots* maliciosos ajudam a amplificar a desinformação. Dessa forma, este trabalho se propõe a criar uma base de dados de perfis do Twitter rotulados em *bot* ou em legítimo, bem como fazer uma análise comparativa das características de perfis *bots* e legítimos. Foram empregados dois métodos diferentes para a coleta de usuários do Twitter, sendo que em um deles foi utilizado o mecanismo de contas *honeypots* (LEE; EOFF; CAVERLEE, 2011). Após a coleta de usuários, foi realizada verificação manual por meio de um questionário, a fim de obter os limites da pontuação bruta geral do classificador Botometer para definição de usuários legítimos e *bots*. No total, a base de dados contou com 1.279 contas, sendo 679 rotuladas como *bots* e 600 rotuladas como legítimas. Para a base de dados completa, foram extraídas informações de cada conta, assim como de suas publicações. Algumas variáveis analisadas indicaram maior heterogeneidade das contas classificadas como *bots*, ao passo que outras variáveis analisadas mostraram que as contas classificadas como legítimas apresentam maior variabilidade.

Palavras-chave: Base de dados rotulada. *Bots*. Twitter. Contas *honeypots*.

Agradecimentos

Primeiramente, gostaria de agradecer aos meus pais Marco Antonio e Rosangela, que sempre apoiaram as minhas decisões e sempre estiveram presente na minha vida.

À professora Dra. Karina Yuriko Yaginuma por ter aceitado ser minha orientadora, pela paciência, dedicação e encorajamento em todas as etapas deste trabalho.

Ao Instituto de Tecnologia e Sociedade do Rio (ITS Rio) pela auxílio financeiro e oportunidade de fazer parte do projeto e tornar viável este trabalho.

Às professoras Dra. Karina Yuriko Yaginuma e Dra. Patrícia Lusié Velozo da Costa pela chance de integrar o projeto do ITS Rio e por todo aprendizado adquirido.

Ao discente Marco Anthonio Guimarães Esquivel pela troca de informações e ensinamentos durante o projeto do ITS Rio.

Às professoras Dra. Patrícia Lusié Velozo da Costa e Dra. Jessica Quintanilha Kubrusly, que aceitaram fazer parte da banca examinadora deste trabalho, pelas sugestões e elogios.

Aos amigos que fiz na UFF: Caroline, Carla, Vitor e Ingrid.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 10
1.1	Motivação	p. 10
1.2	Revisão Bibliográfica	p. 12
1.3	Objetivos	p. 14
1.4	Organização	p. 15
2	Materiais e Métodos	p. 16
2.1	Twitter e bots	p. 16
2.2	Classificador Botometer	p. 17
2.3	Estudo piloto	p. 23
2.4	Estudo atual	p. 25
2.5	Verificação manual de usuários	p. 26
2.5.1	Verificação manual de usuários bots	p. 27
2.5.2	Verificação manual de usuários legítimos	p. 28
2.6	Procedimentos de coleta de usuários	p. 30
2.6.1	Coleta de usuários bots	p. 30
2.6.2	Coleta de usuários legítimos	p. 31
2.7	Procedimento para extração das publicações dos usuários	p. 33

3	Análise dos resultados	p. 35
3.1	Variáveis coletadas dos usuários	p. 35
3.2	Variáveis coletadas das publicações dos usuários	p. 36
3.3	Resultados	p. 37
3.3.1	Variável número de seguidores	p. 37
3.3.2	Variável número de usuários que a conta segue	p. 38
3.3.3	Variável número de tweets que o usuário curtiu	p. 40
3.3.4	Variável número de tweets publicados pelo usuário	p. 41
3.3.5	Variável idade das contas	p. 42
3.3.6	Variável número de tweets publicados pelo usuário por dia	p. 44
3.3.7	Variável idioma das publicações do usuário	p. 45
3.3.8	Variável local definido pelo usuário para o perfil da sua conta	p. 46
3.3.9	Variável nome de usuário da conta	p. 47
4	Conclusões	p. 49
	Referências	p. 51
	Apêndice 1 – Verificação manual - perguntas	p. 54

Lista de Figuras

1	Esquema de votação para classificadores especializados do Botometer	p. 20
2	Informações do Botometer para a conta @CNN	p. 22
3	Informações detalhadas do Botometer para a conta @CNN	p. 22
4	Publicação fixa de uma conta honeypot	p. 26
5	Boxplot do número de seguidores	p. 38
6	Boxplot do número de usuários que a conta segue	p. 40
7	Boxplot do número de tweets que o usuário curtiu	p. 41
8	Boxplot do número de tweets publicados pelo usuário	p. 42
9	Boxplot da idade das contas (dias)	p. 44
10	Boxplot do número de tweets publicados pelo usuário por dia	p. 45
11	Gráfico de barras do idioma das publicações do usuário	p. 46
12	Gráfico de barras do local definido pelo usuário para o perfil da sua conta	p. 47
13	Gráfico de barras do nome de usuário da conta	p. 48

Lista de Tabelas

1	Resultados do estudo piloto por semana e por conta	p. 25
2	Informações sobre verificação manual de usuários bots	p. 28
3	Informações sobre verificação manual de usuários legítimos	p. 29
4	Informações sobre usuários coletados rotulados como bots	p. 30
5	Informações sobre usuários legítimos coletados	p. 32
6	Informações sobre quantidade de usuários legítimos por tópico e conta .	p. 33
7	Descrição das variáveis analisadas	p. 37
8	Medidas resumo da variável número de seguidores	p. 38
9	Medidas resumo da variável número de usuários que a conta segue . . .	p. 39
10	Medidas resumo da variável número de tweets que o usuário curtiu . .	p. 41
11	Medidas resumo da variável número de tweets publicados pelo usuário .	p. 42
12	Medidas resumo da variável idade das contas (dias)	p. 43
13	Medidas resumo da variável número de tweets publicados pelo usuário por dia	p. 45

1 Introdução

1.1 Motivação

O Twitter¹ é uma plataforma em formato de *microblog* lançada em 2006 e que fornece uma estrutura de uso simples por meio de mensagens curtas, os *tweets* (ELMENDILI et al., 2018). Além disso, apresenta fácil acesso, é gratuito, embora haja versões pagas que ofereçam recursos adicionais, e possibilita aos usuários criar perfis, interagir com outras pessoas, ver e publicar conteúdos de interesse (RECUERO; ZAGO, 2016). Maiores informações sobre formas de comunicação no Twitter podem ser obtidas na Seção 2.1.

Um aspecto do Twitter é que um usuário pode seguir qualquer outro usuário e o usuário que está sendo seguido não precisa seguir o outro de volta. Portanto, qualquer pessoa no Twitter pode seguir ou deixar de seguir qualquer pessoa a qualquer momento, exceto as contas bloqueadas (Twitter, 2022b). Ademais, ser um seguidor permite que o usuário receba todos os *tweets* publicados pelo usuário que se segue. Essas características mostram que o acesso às pessoas na plataforma é muito fácil, com poucas restrições, o que favorece à utilização de perfis automatizados, os *bots* sociais.

Bots sociais são contas de mídia social controladas totalmente ou parcialmente por algoritmos de computador (YANG et al., 2019) . Eles podem gerar conteúdo automaticamente e interagir entre si, com usuários humanos e com outros *bots* (REGATTIERI, 2021).

Os *bots* sociais, que daqui para frente no texto serão citados apenas como *bots*, podem ser classificados em benignos ou maliciosos. No primeiro caso, são inofensivos e ingênuos, ou seja, não são poluidores de conteúdo. Além do mais, desempenham funções úteis como divulgação de notícias e trabalhos acadêmicos (LOKOT; DIAKOPOULOS, 2016), recrutamento e coordenação de voluntários (SAVAGE; MONROY-HERNANDEZ; HÖLLERER, 2016), notificação de desastres naturais, transparência política e monitoramento ambiental

¹<https://twitter.com>

(REGATTIERI, 2021).

Os *bots* maliciosos, por sua vez, podem se passar por usuários humanos com o objetivo de manipular e poluir conteúdo. Esse tipo de *bot* pode ser usado para disseminar informações de baixa credibilidade (YANG; FERRARA; MENCZER, 2022), rumores e discurso de ódio (SHEVTSOV et al., 2022). Ademais, sua natureza automatizada ajuda a amplificar a desinformação (YANG et al., 2019).

Atividades de *bots* maliciosos foram relatadas em áreas importantes como eleições (BESSI; FERRARA, 2016; GORODNICHENKO; PHAM; TALAVERA, 2021; PAULA; MICHALSKI, 2019), vacinas (BRONIATOWSKI et al., 2018), pandemia de COVID-19 (FERRARA, 2020; YANG; TORRES-LUGO; MENCZER, 2020), mudanças climáticas (MARLOW; MILLER; ROBERTS, 2020), mercado de ações (FAN; TALAVERA; TRAN, 2020), propaganda terrorista (FERRARA et al., 2016), entre outros.

Particularmente ao tema eleições, Bessi e Ferrara (2016) descobriram que durante as eleições presidenciais dos EUA em 2016 os *bots* produziram grande quantidade de conteúdo. Os autores sugerem que uma forma vulnerável de espalhar informações era por meio de *retweets*, visto que não foi observada diferença significativa na quantidade de *retweets* que contas humanas geravam ao compartilhar conteúdo produzido por outras contas humanas ou por *bots*. Tal fato indica a eficácia dos *bots* em fazer com que as mensagens fossem recompartilhadas por contas legítimas.

No que se refere às eleições no Brasil, o estudo de Paula e Michalski (2019) mostra que nas eleições presidenciais de 2018 foram utilizados *bots* em grande quantidade, com o objetivo de disseminar *hashtags* em apoio aos candidatos e suas propostas, além de atacar perfis com ideias contrárias.

Dessa forma, devido à facilidade de acesso a outras contas e escalabilidade de propagandas e *fake news*, fica evidente a importância do estudo sobre *bots* poluidores de conteúdo dentro do Twitter, além da criação de uma base de dados confiável para ser utilizada na construção de classificadores de contas em *bot* ou em legítima.

A base de dados a ser criada é do tipo rotulada, ou anotada, e requer a adição de rótulos que identificam certas características ou propriedades aos dados brutos coletados. Neste trabalho, em particular, para a base de dados a ser obtida será necessário pegar informações públicas dos perfis das contas do Twitter e rotular cada conta em *bot* ou em legítima.

1.2 Revisão Bibliográfica

Uma das primeiras bases de dados para detecção de *bots* foi construída por Lee, Eoff e Caverlee (2011). Os autores realizaram um estudo de longo prazo usando *honeypots* sociais que têm como objetivo atrair atividades de *bots* no Twitter através de um perfil automatizado que apresenta como característica a coleta automática de possíveis poluidores de conteúdo de uma forma a não interferir nas atividades de usuários genuínos. No total, 60 contas *honeypots* sociais foram criadas e implementadas no Twitter. A frequência, bem como o conteúdo e tipo de postagem das contas *honeypots* sociais foram controladas pelos pesquisadores que determinaram os tipos de *tweets* a serem publicados: *tweet* textual; resposta “@” a outras contas *honeypots* sociais; *tweet* contendo um *link*; e *tweet* contendo um dos 10 principais tópicos do Twitter. Para publicar tais *tweets* foi necessário coletar 120.000 *tweets* que serviram como amostra (30.000 para cada um dos quatro tipos). Logo que um usuário fazia contato com uma das contas *honeypots* sociais, informações sobre ele e seus *tweets* anteriores eram coletados. Todos os usuários que entraram em contato com uma conta *honeypot* social foram considerados como *bot*. As contas *honeypots* sociais, ou simplesmente contas *honeypots*, ficaram em funcionamento de 30 de dezembro de 2009 a 02 de agosto de 2010. Nesse período foram coletados 36.043 usuários do Twitter, e após remover usuários que seguiam mais de uma conta *honeypot*, 23.869 usuários permaneceram na base de dados.

Outro estudo importante foi o de Cresci et al. (2015), que forneceu uma base de dados rotulada de contas genuínas e seguidores falsos. Para o conjunto de dados de contas genuínas, as atividades tiveram início em 12 de dezembro de 2012 com a criação de uma conta do Twitter (@TheFakeProject), cujo perfil apresentava uma frase: “Follow me only if you are NOT a fake” (Siga-me apenas se você não for um *fake*). Em um período de 12 dias, essa conta foi seguida por 574 perfis, sendo que 469 foram certificados como humanos. Ainda nesse estudo, outro conjunto de dados de contas genuínas consiste em usuários italianos ativos no Twitter com interesse pela política. No período entre 09 de janeiro e 28 de fevereiro de 2013, foram identificadas 84.033 contas únicas no Twitter que usaram a *hashtag* *elezione2013* em seus *tweets* e, após exclusão de contas que pertenciam a blogueiros, jornalistas ou políticos, permaneceram 40.000 contas, aproximadamente. Esse último conjunto foi submetido a uma verificação manual por duas pessoas que abrangeu análise de fotos de perfil, biografia e linha do tempo das contas. Para serem incluídas no conjunto de dados, as contas deveriam ser rotuladas como humanas pelas duas pessoas. A fase de verificação manual durou dois meses e teve como resultado 1.481 contas genuínas.

Os seguidores falsos, por sua vez, foram comprados de três diferentes mercados *online* em abril de 2013. No total, a base de dados contava com 3.900 contas, sendo 1.950 humanos e 1.950 seguidores falsos.

Já o estudo de Varol et al. (2017) teve como objetivo rotular contas como *bot* ou legítimas manualmente, amostradas por decis de pontuação do Botometer², que é um classificador de contas no Twitter. Mais informações sobre o Botometer podem ser obtidas na Seção 2.2. Em um período de três meses a partir de outubro de 2015, foram coletadas 14 milhões de contas de usuários do Twitter, que foram agrupadas segundo a pontuação de *bot* de cada uma. Em seguida, 300 contas de cada decil de pontuação de *bot* foram amostradas aleatoriamente, resultando em 3.000 contas a serem rotuladas manualmente. Tal verificação manual contou com quatro voluntários que analisaram a aparência do perfil, conteúdo produzido e retweetado, bem como as interações com outras contas em termos de *retweets*. Ressalta-se que para classificar as contas, os voluntários não receberam um conjunto de instruções, mas foram mostrados a eles exemplos. Assim, as contas poderiam ter indicação de humana, *bot* ou indecisa, sendo essa última eliminada. Na base de dados ficaram 2.573 contas do Twitter, correspondendo a 826 *bots* e 1.747 humanas.

O estudo de Gilani et al. (2017) também apresentou a criação de uma base de dados rotulada manualmente. Por 30 dias em abril de 2016 e por meio da API (*application programming interface*) do Twitter, foram coletados 65 milhões de *tweets* e 2,9 milhões de contas únicas. Para distinguir entre *bots* e humanos, foram recrutados quatro alunos, sendo que cada conta foi revisada por todos os alunos de forma independente. Os alunos receberam não só o perfil da conta, mas também uma lista de recursos utilizados pela conta ao longo do mês, como, por exemplo, aplicativo do Twitter, navegador etc. Além de verificar o conteúdo das contas, os alunos também levaram em consideração a quantidade e os tipos de recursos utilizados por cada conta, visto que esses podem indicar alguma forma de automação. No total, foi possível anotar 3.062 contas: 1.304 classificadas como *bots* e 1.758 como humanas.

Por sua vez, o estudo de Feng et al. (2021) teve como objetivo a criação de uma base de dados abrangente do Twitter denominada TwiBot-20. Essa visa mitigar problemas encontrados em outras bases de dados de *bots* do Twitter como escassez de dados, falta de diversidade e informações limitadas de usuários. Para obter uma amostra diversificada e representativa do Twitter, foi utilizada busca em largura, começando de diferentes vértices raiz, conhecidos como usuários sementes. Assim, usuários foram tratados como nós e

²<https://botometer.osome.iu.edu>

seus seguidores como arestas para formar um grafo direcionado. Cada usuário semente foi colocado na camada zero do *cluster* de usuário. Usuários na camada $i + 1$ foram expandidos de usuários na camada i ao longo de suas arestas de acompanhamento. Esse processo de expansão terminou na camada três e formou um *cluster* de usuários. TwiBot-20 usou 40 usuários sementes de 4 tópicos distintos: política, negócios, entretenimento e esporte. Além disso, foi utilizada a API do Twitter para obter algumas informações: para cada usuário, os 200 *tweets* mais recentes, bem como 38 itens de propriedade e informações de usuários seguidores e usuários seguidos. Para verificar se um usuário da base de dados é um *bot* ou não, foi necessário analisá-lo manualmente. Para tanto, cinco pessoas foram contratadas e elas seguiram critérios propostos pelos autores do estudo. Para a anotação final de usuários, os seguintes passos foram observados:

- i) Se o usuário é verificado pelo Twitter, então foi considerado como usuário legítimo;
- ii) Para os demais usuários, se quatro dos cinco avaliadores acreditam ser *bot* (ou humano), então é feita a anotação de *bot* (ou humano);
- iii) Para os outros usuários com menos acordo mútuo entre os avaliadores, foi enviada, através do aplicativo do Twitter, uma mensagem direta com perguntas simples em linguagem natural. As respostas foram coletadas, tornando possível determinar de forma manual se o usuário era legítimo ou um *bot*;
- iv) Os usuários indecisos que restaram foram examinados manualmente pelos próprios pesquisadores e apenas foram rotulados os usuários do Twitter quando havia um consenso. Os demais casos eram descartados.

No total, TwiBot-20 contém 229.573 usuários, além de informações semânticas, de propriedade e de vizinhança de cada usuário. Ademais, a base de dados apresenta não somente diversidade geográfica, com usuários da Índia, EUA, Europa e África, mas também diversidade de interesse de usuários, sendo que as *hashtags* mais comentadas foram COVID-19, BlackLivesMatter, SaintsFC, Trump, entre outras.

1.3 **Objetivos**

Este trabalho apresenta dois objetivos. O primeiro e principal objetivo consiste em criar uma base de dados de perfis do Twitter rotulados em *bot* ou em legítimo. Já o segundo envolve fazer uma análise comparativa das características de perfis *bots* e legítimos.

A criação da base de dados faz parte de um projeto em conjunto do Instituto de Tecnologia e Sociedade do Rio ³ (ITS Rio) com as docentes Karina Yuriko Yaginuma e Patrícia Lusié Velozo da Costa e com os discentes Marco Anthonio Guimarães Esquivel e Patricia Mello Marçal Pinto.

1.4 Organização

No Capítulo 2 serão apresentadas informações sobre o Twitter, o classificador Botometer e verificação manual de usuários. Nesse capítulo também serão explicados e detalhados os resultados do estudo piloto e do estudo atual para obtenção da base de dados rotulada. Os procedimentos de coleta de usuários e de extração das publicações dos usuários serão explicitados. No Capítulo 3, as variáveis coletadas dos usuários e das publicações dos usuários serão detalhadas, bem como os resultados das variáveis analisadas das contas rotuladas como *bots* ou como legítimas. No Capítulo 4 serão apresentadas as conclusões deste trabalho, as dificuldades encontradas e propostas para trabalhos futuros.

³<https://itsrio.org/pt/home/>

2 Materiais e Métodos

Neste Capítulo, na Seção 2.1 serão apresentadas formas de comunicação, informações e regras do Twitter. Na Seção 2.2 o classificador Botometer será explicado. A Seção 2.3 apresentará detalhes e resultados do estudo piloto para criação da base de dados rotulada. Na Seção 2.4 as modificações realizadas no estudo piloto e que culminaram no estudo atual serão detalhadas. Na Seção 2.5 serão apresentados as formas de verificação manual de usuários *bot* e legítimos. A Seção 2.6 explicitará os procedimentos de coleta de usuários. Por fim, na Seção 2.7 será explicado o procedimento para extração das publicações dos usuários.

2.1 Twitter e bots

No Twitter as mensagens curtas, os *tweets*, possuem até 280 caracteres e podem conter texto, fotos, *GIFs* e vídeos. Os *tweets* são exibidos na linha do tempo do seu autor ou incorporados em *sites* e *blogs* (Twitter, 2022b).

Para responder um *tweet* de outro usuário utiliza-se um *reply*, enquanto *retweet* é o ato de compartilhar o *tweet* de outra conta para seus seguidores (*followers*). Outro modo de comunicação se dá por meio das mensagens diretas (*direct messages* ou DMs) que são mensagens privadas enviadas de uma conta para outra. Para mencionar (*mention*) outras contas no *tweet* é utilizado o símbolo “@” seguido diretamente pelo nome do usuário. Também é possível curtir (dar *like*) em um *tweet*. Um recurso que ajuda os usuários a categorizar seus *tweets* é uma *hashtag* que pode ser qualquer palavra ou frase imediatamente precedida pelo símbolo “#” (Twitter, 2022b).

Via de regra, os perfis no Twitter são públicos, ou seja, todos podem ver as informações e publicações de um determinado usuário (Twitter, 2022b). Há a possibilidade de fechar o perfil de uma conta e, conseqüentemente, somente os seguidores poderão ter acesso às informações e postagens daquela conta.

Há ainda uma plataforma de desenvolvedor que por meio de uma API permite a análise de diversos aspectos do Twitter.

No segundo trimestre de 2022, o Twitter registrou um número médio diário de 237,8 milhões de usuários ativos, um aumento de 16,6% em relação ao mesmo período do ano anterior (Twitter, 2022c). Em janeiro de 2022, os cinco países com maior número de usuários na plataforma foram EUA, Japão, Índia, Brasil e Indonésia (Statista, 2022a). As três contas com maior número de seguidores em abril de 2022 foram @barackobama, @justinbieber e @katyperry (Statista, 2022b).

Segundo estimativas de Varol et al. (2017), entre 9% e 15% das contas ativas do Twitter exibem comportamento de *bot*.

Vale lembrar que é permitido a criação e utilização de *bots*, desde que eles não violem as regras do Twitter (2022a). Assim, é proibido:

- Uso malicioso da automação para distorcer e atrapalhar a conversa pública, como tentar fazer um termo chegar aos assuntos do momento;
- Amplificação artificial de conversas, inclusive por meio da criação de contas múltiplas ou sobrepostas;
- Gerar, solicitar ou comprar engajamentos falsos;
- Utilizar táticas massivas ou agressivas ao tweetar, engajar ou seguir contas;
- Usar táticas de *spam* em *hashtags*, incluindo o uso de *hashtags* não relacionadas em um *tweet* (também conhecido como “*hashtag cramming*”).

2.2 Classificador Botometer

O classificador Botometer é um projeto conjunto do *Observatory on Social Media* (OSoMe) e do *Network Science Institute* (IUNI) da Universidade de Indiana.

De acordo com Yang, Ferrara e Menczer (2022), o Botometer é uma ferramenta de aprendizado de máquina supervisionado, isto é, necessita de um conjunto de dados rotulados para aprender as características de prováveis contas humanas e *bots*. O Botometer calcula uma pontuação, sendo que pontuações baixas sugerem prováveis contas humanas, enquanto pontuações altas sugerem prováveis contas *bots*.

Há duas formas de utilizar o Botometer: através da página na *internet* e através da API (<https://github.com/IUNetSci/botometer-python>).

Sob o nome BotOrNot, segundo Davis et al. (2016), o Botometer foi lançado em maio de 2014. Com pontuação calculada no intervalo $[0,1]$, o sistema de classificação do BotOrNot analisava 1.150 características do perfil, usando metadados e informações de dados públicos obtidas de conteúdo e padrões de interação. As características foram agrupadas em seis classes principais:

- Características da rede: capturam diversas dimensões dos padrões de difusão de informação. Redes são construídas com base em *retweets*, menções e *hashtags*, sendo que são extraídas informações estatísticas como coeficiente de agrupamento e medidas de centralidade;
- Características do usuário: baseada em metadados do perfil obtidos pela API do Twitter, incluindo tamanho e número de dígitos no *screen name*, tamanho do *user name*, idade da conta (em dias), quantidade de *tweets*, *retweets*, menções e respostas (quantidades total e por hora de postagens);
- Características dos amigos: incluem características descritivas em relação aos contatos de uma conta, como mediana, momentos e entropia das distribuições do número de seguidores, seguidos, postagens, entre outros;
- Características temporais: analisam estatísticas sobre intervalos de tempo (entre dois *tweets* consecutivos, entre dois *retweets* consecutivos e entre duas menções consecutivas);
- Características linguísticas e de conteúdo: têm como base pistas linguísticas processadas por meio de linguagem natural, principalmente marcação de classe gramatical (*part-of-speech tagging*);
- Características de sentimento: verificam como as emoções podem se manifestar nas postagens. São baseadas em algoritmos de análise de sentimentos de uso geral e específico do Twitter. Incluem pontuações de felicidade, de valência, excitação e dominância, e de *emoticons*.

Para treinar o modelo do Botometer, os idealizadores desse classificador utilizaram bases de dados rotuladas em contas humanas e em *bots*, sendo que para essa última foi utilizada a base de dados identificada por Lee, Eoff e Caverlee (2011).

Por meio da API do Twitter para obter o histórico recente da conta, foram coletados até os 200 *tweets* mais recentes do usuário, bem como os 100 *tweets* mais recentes que mencionaram o usuário.

Segundo Varol et al. (2017), o comportamento de contas humanas e *bots* evoluem e novos comportamentos surgem, tornando necessário atualizar o classificador para que ele permaneça preciso. Dessa forma, a versão 2 foi adotada em maio de 2016 e trouxe como novidade a adição de uma base de dados de treinamento. Tal base de dados foi avaliada manualmente, sendo composta de contas de usuários do Twitter que publicam em inglês.

A versão 3 do classificador, por sua vez, foi adotada em maio de 2018, agora com o nome Botometer. Nessa nova versão, conforme Yang et al. (2019), novas pontuações foram apresentadas e incorporou-se novas características e bases de dados.

Como a pontuação é definida no intervalo $[0,1]$, muitos usuários do Botometer interpretavam-na como equivalente à probabilidade da conta ser *bot*. Assim, para resolver esse problema, duas novas pontuações foram adicionadas: reescalonamento linear de pontuações para o intervalo $[0,5]$ e a *Complete Automation Probability* (CAP).

A pontuação CAP é definida como a probabilidade de uma conta com a pontuação fornecida ou superior ser automatizada. Como exemplo, suponha que uma conta apresente uma pontuação de $0,96/1$ (equivalente a $4,8/5$) e CAP de 90%. Esses resultados significam que 90% das contas com pontuação maior do que 0,96 são reconhecidas como *bots*, ou, equivalentemente, 10% das contas com pontuação maior do que $4,8/5$ são reconhecidas como humanas. Portanto, a versão 3 do Botometer apresentou três pontuações: no intervalo $[0,1]$, denominada pontuação bruta (*raw score*), no intervalo $[0,5]$, denominada pontuação de exibição (*display score*) e a CAP no intervalo $[0,1]$.

Ainda na versão 3, novas características foram introduzidas com o objetivo de capturar comportamentos mais sofisticados de *bots*. Exemplos de novas características incluem fusos horários, metadados de idioma, metadados do dispositivo e padrões de exclusão de conteúdo. Além disso, a versão 3 do Botometer conta com o acréscimo de cinco novas bases de dados de treinamento, totalizando sete bases de dados.

Já a versão 4 do classificador, adotada em setembro de 2020, apresenta uma nova arquitetura, denominada *Ensemble of Specialized Classifiers* (ESC). Sayyadiharikandeh et al. (2020) verificaram que contas *bots* apresentam maior heterogeneidade, ou seja, diferentes características, enquanto contas humanas são mais homogêneas. Assim, os autores propuseram um novo método para o Botometer, capaz de treinar diversos classificadores

especializados para tipos distintos de *bots*.

Segundo o Botometer (2022), os *bots* são classificados em:

- *Echo-chamber*: contas que se envolvem em grupos do tipo “seguir de volta” e que compartilham e excluem, em alto volume, conteúdo político;
- *Fake follower*: *bots* comprados com o objetivo de inflar o número de seguidores;
- *Financial*: *bots* que publicam utilizando *cashtags*, que são *hashtags* de propaganda financeira;
- *Self declared*: *bots* do *botwiki.org*;
- *Spammer*: contas rotuladas como *spambots* de várias bases de dados;
- *Other*: *bots* diversos, obtidos a partir de verificações manuais, *feedback* de usuários, etc.

A pontuação é calculada por meio de um esquema de votação para os classificadores de *bot*, conforme pode ser observado na Figura 1. No total são construídos $n + 1$ classificadores, sendo n o número de *bots* distintos. Os classificadores podem ser separados em dois tipos: RF_0 , que detecta comportamento de perfis humanos (versão anterior do Botometer), e RF_1, RF_2, \dots, RF_n , que detectam comportamento de perfis *bots* especializados. Assim, são geradas $n + 1$ pontuações $(s_0, s_1, s_2, \dots, s_n)$, em que s_0 é a pontuação para classificar perfis humanos.

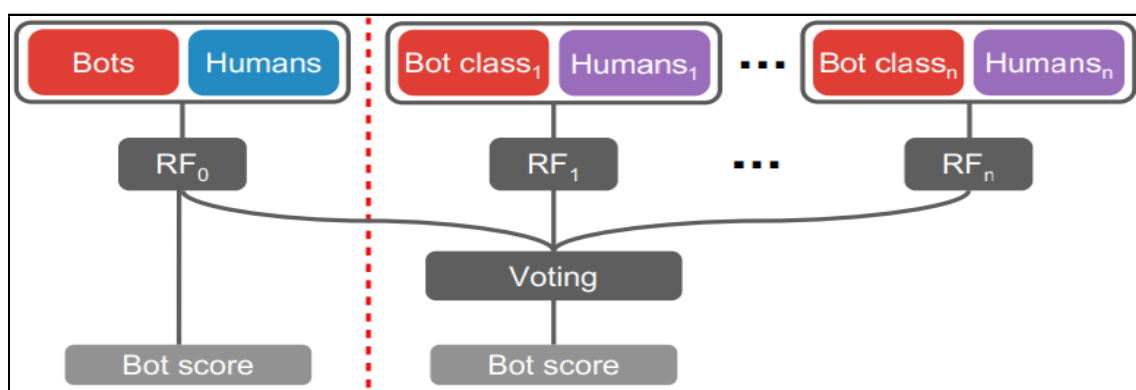


Figura 1: Esquema de votação para classificadores especializados do Botometer
Fonte: Sayyadharikandeh et al. (2020)

Para determinar a classe a que o *bot* pertence, considera-se $i^* = \operatorname{argmax}_i \{s'_i\}$, onde:

$$s'_i = \begin{cases} 1 - s_i, & \text{se } i = 0 \\ s_i, & \text{c.c.} \end{cases}$$

A pontuação de *bot* (*bot score*), também chamada de pontuação geral, é obtida calibrando a pontuação s_i^* usando a escala de Platt (NICULESCU-MIZIL; CARUANA, 2005).

Vale ressaltar que características linguísticas, de conteúdo e de sentimentos têm como base a língua inglesa. Para contas em outro idioma, como alternativa, o Botometer apresenta uma pontuação independente do idioma, que exclui as características linguísticas.

Para a versão 4, diversas bases de dados rotuladas foram adicionadas, totalizando 15 bases.

Uma outra versão do Botometer, o BotometerLite, criada por Yang et al. (2020), utiliza um mínimo de metadados das contas. O BotometerLite é um detector de *bots* escalonável e leve para o Twitter. Ao contrário do Botometer, essa versão necessita apenas de um *tweet* para a detecção de *bot*, o que traz algumas vantagens como a possibilidade de analisar com eficiência grandes quantidades de contas, além de detectar *bots* em dados históricos, enquanto o Botometer retorna apenas o *status* mais recente.

A API do BotometerLite está disponível através do plano ULTRA da API do Botometer Pro mediante pagamento.

Importante destacar que o Botometer é constantemente usado em vários artigos como o classificador mais consistente para fazer rotulação de *bots* para gerar bases de dados. Em uma pesquisa no Google Scholar para verificar o número de artigos publicados entre os anos 2018 e 2022 e que mencionaram a palavra *botometer*, foram obtidos 1.150 resultados para todos os idiomas e 38 em português. Esses resultados corroboram a ideia de que o Botometer é o método de detecção de *bots* mais popular, além de proporcionar reconhecimento ao classificador.

A Figura 2 apresenta as pontuações fornecidas em 21 de junho de 2022 na página do Botometer para a conta @CNN. É possível verificar a pontuação de exibição no intervalo $[0,5]$, bem como a pontuação por tipo de *bot*. Nota-se que a pontuação de exibição ficou em 4,4/5 e que a pontuação CAP foi de 14%, que significa que 14% das contas com pontuação maior do que 4,4 foram rotuladas como humanas. Em “Details” são obtidas informações sobre a conta como descrição, data de criação, data da publicação mais recente, idioma e quantidade de *tweets*, além de gráficos mostrando *tweets* por dia da semana e por hora do dia (considerando os últimos 200 *tweets*). A Figura 3 mostra essas informações.

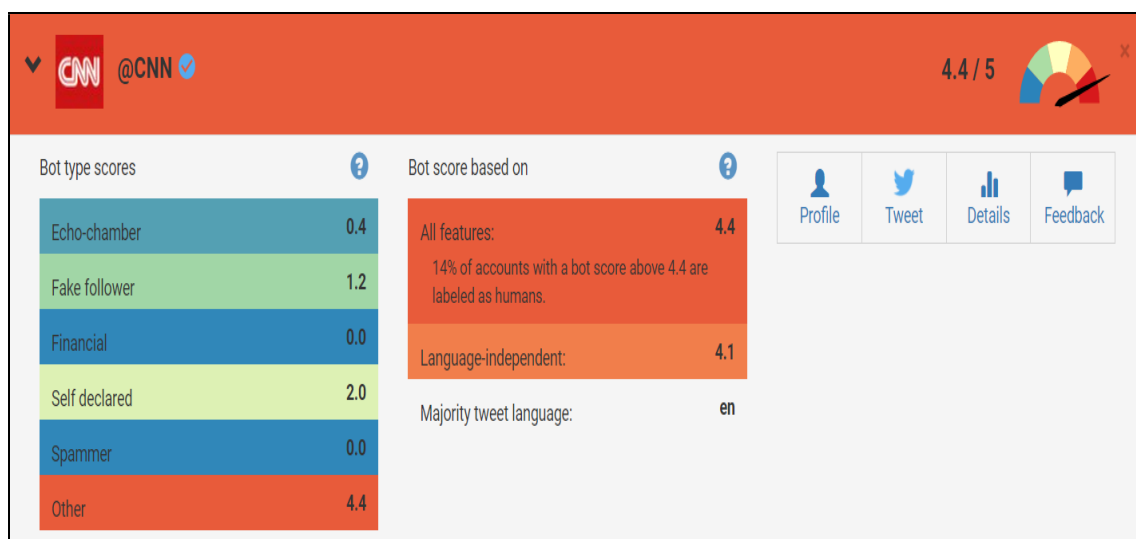


Figura 2: Informações do Botometer para a conta @CNN

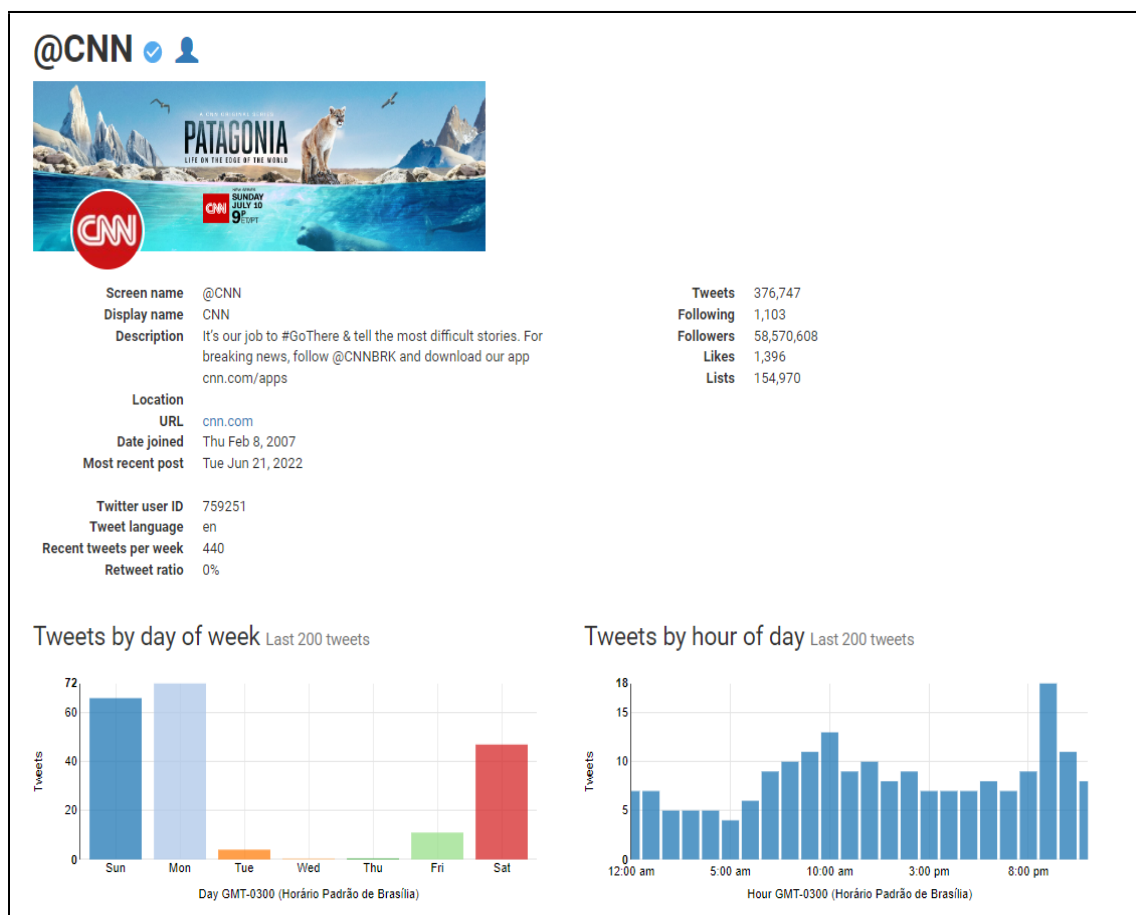


Figura 3: Informações detalhadas do Botometer para a conta @CNN

2.3 Estudo piloto

Para saber se é possível atrair *bots* maliciosos com conteúdo em português, foi necessário realizar um estudo piloto antes de se criar uma base de dados rotulada.

O estudo piloto efetuado baseou-se no trabalho de Lee, Eoff e Caverlee (2011). Para tanto, foram criadas e implementadas no Twitter quatro contas *honeypots* que ficaram em funcionamento por quatro semanas, a partir de 16 de fevereiro de 2022. Tais contas podiam publicar *retweets*, *links*, textos aleatórios e *tweet* contendo um dos principais tópicos. Importante frisar que as contas *honeypots* interagem apenas entre si, não interferindo nas atividades de outros usuários.

O *script* para as publicações das contas *honeypots* estava em linguagem Python e no código era possível encontrar uma função com estruturas condicionais.

Para publicar textos aleatórios, foi utilizada a biblioteca *tweepy* para obter 100 *tweets* diferentes, extraídos da linha do tempo do Twitter, que incluíssem uma determinada palavra. Os textos dos *tweets* foram colocados em um arquivo de texto de forma aleatória, ou seja, para que os textos de um mesmo assunto não aparecessem de forma sequencial. O texto a ser publicado era sempre o primeiro do arquivo e após a sua publicação ele era deletado. Caso o arquivo de texto estivesse vazio, seria preciso coletar outros textos aleatórios.

Para publicar *tweets* contendo um dos principais tópicos, procurou-se os assuntos do momento no Brasil. Em seguida, escolheu-se aleatoriamente e publicou-se o *tweet* contendo um dos principais tópicos juntamente com um texto aleatório definido anteriormente.

Já para publicação de *links*, também foi realizada busca de modo a obter os assuntos do momento no Brasil. Pegou-se a url (*uniform resource locator*) de cada assunto e publicou-se o *link* escolhido aleatoriamente.

Por fim, para *retweets*, uma conta *honeypot* foi selecionada de forma aleatória e seus dez últimos *tweets* foram procurados. Em seguida, um único *tweet* foi selecionado aleatoriamente para ser retweetado e curtido por outra conta *honeypot*.

A quantidade de publicações diárias variava entre 20 e 25. O *script* ficou hospedado em uma plataforma (<https://www.heroku.com/>) que também permitia que ele fosse rodado na nuvem 24 horas por dia.

Informações sobre o *script* e alguns textos aleatórios do estudo piloto estão disponíveis no GitHub do projeto (<https://github.com/pmmp-tcc/Estudo-piloto>).

A Tabela 1 mostra os resultados do estudo piloto por semana e por conta.

Na primeira semana, cada conta *honeypot* postou um tipo de publicação a mais do que outros tipos de publicação. Dessa forma, seis seguidores passaram a seguir uma das contas *honeypot*, sendo que cinco começaram a seguir a conta que postava mais *links*. Devido aos baixos resultados da primeira semana e na tentativa de atrair mais *bots*, todas as contas *honeypots* publicaram mais *links* na segunda semana. Como resultado, foi possível atrair apenas mais dois seguidores. Na terceira semana ficou definido que todas as contas *honeypots* publicariam mais *tweets* contendo um dos principais tópicos. Assim, três usuários passaram a seguir os *honeypots*. Por sua vez, na quarta semana todas as contas *honeypots* postaram mais textos aleatórios. Todavia, não foi possível atrair novos seguidores. No total, em quatro semanas, as contas *honeypots* conseguiram coletar 11 contas, sendo 9 não repetidas.

Como o estudo piloto atraiu poucos *bots*, foi necessário realizar modificações no funcionamento das contas *honeypots* de modo que elas atraíssem maior quantidade desse tipo de perfil.

Tabela 1: Resultados do estudo piloto por semana e por conta

Semana	Conta	Tipo de publicação a mais	Resultado
1	Honeypot 1	Principais tópicos	1 conta
	Honeypot 2	Retweets	Nenhuma conta
	Honeypot 3	Links	5 contas
	Honeypot 4	Textos aleatórios	Nenhuma conta
2	Honeypot 1	Links	Nenhuma conta
	Honeypot 2		1 conta
	Honeypot 3		1 conta
	Honeypot 4		Nenhuma conta
3	Honeypot 1	Principais tópicos	1 conta
	Honeypot 2		Nenhuma conta
	Honeypot 3		Nenhuma conta
	Honeypot 4		2 contas
4	Honeypot 1	Textos aleatórios	Nenhuma conta
	Honeypot 2		
	Honeypot 3		
	Honeypot 4		

2.4 Estudo atual

Paralelamente ao estudo piloto e na tentativa de atrair mais *bots*, realizou-se um teste e duas novas contas *honeypots* foram criadas, sendo que elas publicavam somente *retweets* que incluíssem um termo, podendo ser uma palavra ou uma *hashtag*. Como resultado, em duas semanas, essas duas contas *honeypots* conseguiram coletar 85 contas, quantidade superior à apresentada pelo estudo piloto.

Dessa forma, decidiu-se que as contas *honeypots* somente publicariam *retweets* que incluíssem um termo, que poderia ser uma palavra ou uma *hashtag*. No total dez contas *honeypots* foram implementadas e entraram em funcionamento a partir de abril de 2022.

O *script* para os *retweets* está em linguagem Python e apresenta duas funções para *retweets*: uma contendo apenas *hashtag* que é selecionada de uma lista que inclui apenas *hashtags*, e a outra contendo palavra ou *hashtag* de uma lista que possui palavras e *hashtags*.

Das dez contas *honeypots*, decidiu-se que quatro publicariam *retweet* com *hashtag* e seis publicariam *retweet* com *hashtag* ou palavra.

Inicialmente escolheu-se a conta *honeypot* para realizar a publicação e, em seguida, selecionou-se aleatoriamente a palavra ou *hashtag*. Os 20 *tweets* mais recentes com o termo selecionado foram pesquisados e, de forma aleatória, definiu-se um *tweet* para ser compartilhado e curtido.

A fim de atrair apenas *bots*, cada conta *honeypot* recebeu, por uma conta particular de um dos discentes do projeto, um rótulo de conta automatizada em seu perfil. Ademais, fixou-se um *tweet* na linha do tempo de cada conta *honeypot* informando que qualquer um que seguisse aquele perfil seria considerado como um possível *bot*, conforme mostra a Figura 4.

O *script* ficou em uma instância virtual hospedada na infraestrutura do Google (<https://cloud.google.com>), o que permitiu que fosse executado 24 horas por dia.

Informações sobre o *script* e as listas contendo algumas palavras e *hashtags* encontram-se no GitHub do projeto (<https://github.com/pmmp-tcc/Estudo-atual>).



Figura 4: Publicação fixa de uma conta honeypot

2.5 Verificação manual de usuários

A verificação manual de contas do Twitter foi implementada para analisar os limites da pontuação bruta geral do Botometer para definição dos usuários *bots* e legítimos, e foi realizada por meio de um questionário com perguntas formuladas através de informações obtidas na literatura e da observação de contas *bots* do Twitter.

O questionário continha 21 perguntas no total, sendo que as possíveis respostas eram “Sim”, “Não”, e, para algumas perguntas, “NA” (não aplicável). Cada resposta do questionário tinha uma pontuação, podendo ter valor zero ou um.

A verificação manual começou pela análise do perfil e seus amigos. Essa parte era composta de 12 perguntas no total. Caso a biografia da conta possuísse *links* para redes sociais como Facebook, Instagram e LinkedIn, então essas também eram verificadas com o

objetivo de encontrar relação com o perfil do Twitter.

A outra parte era composta de nove perguntas no total. Aqui era necessário analisar as últimas 20 publicações da conta.

A verificação manual foi concluída somando-se a pontuação de cada resposta. Se os limites da pontuação bruta geral do Botometer definissem usuários legítimos, então a pontuação somada das respostas deveria ser menor do que 11. Caso os limites da pontuação bruta geral do Botometer definissem usuários *bots*, então a pontuação somada das respostas deveria ser maior ou igual a 11.

No Apêndice 1 encontram-se todas as perguntas para a verificação manual.

2.5.1 Verificação manual de usuários bots

A verificação manual de possíveis contas *bots* foi realizada entre 01/07/2022 e 03/07/2022. De 1.003 contas coletadas até esse período, 712 apresentavam pontuação CAP maior ou igual a 0,75. Desse último número, 75 apresentavam pontuação bruta geral do Botometer no intervalo $[0,4; 0,5]$. Tais contas passaram pela verificação manual, sendo 38 verificadas por um avaliador discente do projeto e 37 por outro avaliador discente do projeto.

A Tabela 2 mostra informações sobre o número de contas classificadas manualmente como *bots* e não *bots* segundo a pontuação bruta geral do Botometer. Como exemplo, a primeira linha da Tabela 2 informa que no total 15 contas apresentaram pontuação bruta geral do Botometer igual a 0,40, sendo seis contas classificadas como *bots* e nove classificadas como não *bots*. No geral, 39 contas foram classificadas como *bots*, 30 contas foram classificadas como não *bots* e seis contas estavam suspensas.

Tabela 2: Informações sobre verificação manual de usuários bots

Pontuação	Número de contas bot	Número de contas não bot	Total
0,40	6	9	15
0,41	8	2	10
0,42	3	5	8
0,43	7	6	13
0,44	5	2	7
0,45	2	0	2
0,46	2	2	4
0,47	0	3	3
0,48	2	0	2
0,49	3	0	3
0,50	1	1	2
Total	39	30	69

Assim, devido à alta quantidade de contas verificadas manualmente como não *bots*, ficou decidido que para entrar na base de dados com o rótulo de *bot*, é necessário que a conta tenha pontuação bruta geral do Botometer maior ou igual a 0,5, ou seja, a pontuação deve estar no intervalo $[0,5; 1]$, do contrário a conta é descartada.

2.5.2 Verificação manual de usuários legítimos

Para coletar possíveis usuários legítimos, foi utilizado, no dia 22/10/2022, o mecanismo a ser descrito na Sessão 2.6.2. As contas selecionadas foram @spotifybrasil, @geglobo, @espnbrasil, @tntsportsbr, @cnnbrasil, @globoNews, @senadofederal e @camaradeputados. No total foram coletados 470 usuários e eles tiveram a pontuação bruta geral do Botometer definida no mesmo dia que foram coletados.

Para determinar o corte na pontuação bruta geral dada pelo Botometer para que usuários entrassem na base de dados como legítimos, foi realizada verificação manual de uma amostra aleatória de 50 contas de usuários com pontuação bruta geral no intervalo $[0; 0,2]$.

A verificação manual contou com 2 avaliadores discentes do projeto que analisaram 25 usuários cada.

A Tabela 3 contém informações sobre o número de contas verificadas manualmente como legítimas e não legítimas segundo a pontuação bruta geral do Botometer. Como exemplo, a segunda linha da Tabela 3 informa que cinco contas apresentaram pontuação bruta geral do Botometer igual a 0,01, sendo as cinco contas classificadas como legítimas e nenhuma classificada como não legítima. No geral, 48 contas foram classificadas como legítimas e duas contas foram classificadas como não legítimas. Assim, decidiu-se que usuários com pontuação bruta geral menor ou igual a 0,2 poderiam ser rotulados como legítimos. Logo, o intervalo da pontuação bruta geral do Botometer para usuários legítimos na base de dados ficou definido em $[0; 0,2]$.

Tabela 3: Informações sobre verificação manual de usuários legítimos

Pontuação	Número de contas legítimas	Número de contas não legítimas	Total
0	1	0	1
0,01	5	0	5
0,02	6	0	6
0,03	6	0	6
0,04	2	0	2
0,05	1	0	1
0,06	2	0	2
0,07	2	1	3
0,08	5	0	5
0,09	1	0	1
0,10	3	0	3
0,11	0	0	0
0,12	2	0	2
0,13	2	1	3
0,14	3	0	3
0,15	2	0	2
0,16	2	0	2
0,17	2	0	2
0,18	0	0	0
0,19	0	0	0
0,20	1	0	1
Total	48	2	50

2.6 Procedimentos de coleta de usuários

Dois procedimentos distintos foram adotados para a coleta de usuários, um para cada rótulo da conta.

2.6.1 Coleta de usuários bots

A coleta de usuários *bots* procedeu da seguinte forma:

- usuário candidato a *bot* passou a seguir uma das contas *honeypot* (10 contas *honeypot* implementadas);
- informação da id do usuário foi extraída;
- informações sobre a conta do usuário e pontuação bruta geral do Botometer foram obtidas.

Os usuários coletados eram únicos, ou seja, não houve repetição de usuários.

As informações coletadas sobre a conta do usuário serão descritas na Seção 3.1.

Entre abril e agosto de 2022 foram coletados 1.517 usuários possíveis *bots*, sendo que 679 apresentaram pontuação bruta geral do Botometer no intervalo $[0,5; 1]$.

Em relação à quantidade de contas *honeypots*, foi necessária a criação de 20 contas desse tipo, visto que durante o período de coleta muitas dessas contas violaram as regras do Twitter e conseqüentemente foram suspensas.

A Tabela 4 mostra as datas em que a pontuação bruta geral do Botometer foi obtida e a quantidade de contas passadas no Botometer para cada data. Como exemplo, no dia 14/06/2022, foi obtida a pontuação bruta geral do Botometer para 402 contas.

Tabela 4: Informações sobre usuários coletados rotulados como bots

Data da pontuação do Botometer	Quantidade de contas
14/06/2022	402
15/08/2022	192
26/08/2022	85
Total	679

2.6.2 Coleta de usuários legítimos

A coleta de usuários legítimos teve como base o trabalho de Feng et al. (2021). Para tanto, foram selecionadas 11 contas de quatro tópicos distintos: entretenimento (@netflixbrasil, @spotifybrasil, @primevideobr), esporte (@geglobo, @espnbrasil, @tntsportsbr), notícias gerais (@folha, @cnnbrasil, @globoNews) e política (@senadofederal, @camaradeputados). As etapas de coleta foram as seguintes:

- procurou-se os últimos *tweets* publicados por cada uma das 11 contas mencionadas anteriormente;
- para o último *tweet* publicado por cada uma das 11 contas, extraiu-se a id de 100 usuários que curtiram essa publicação mais recentemente;
- em seguida, extraiu-se a id de 100 usuários que curtiram o penúltimo *tweet* publicado por cada uma das 11 contas. Caso esse *tweet* apresentasse mais de 100 curtidas, então extraiu-se os 100 usuários que curtiram essa publicação mais recentemente. Caso esse *tweet* apresentasse menos de 100 curtidas, então seria necessário extrair a id de usuários que curtiram o antepenúltimo *tweet* publicado. Essa sequência ocorre até coletar 200 usuários para cada uma das 11 contas.
- no total, foram extraídos id de 2.200 usuários;
- em seguida foram extraídas informações das contas dos 2.200 usuários, bem como a pontuação bruta geral do Botometer;

As informações extraídas da conta de cada usuário serão descritas na Seção 3.1.

A Tabela 5 contém informações como tópico, descrição das contas, data e horário de extração das informações e da pontuação bruta geral do Botometer dos 2.200 usuários e número de *tweets* necessários para coletar 200 usuários que curtiram os *tweets* de cada uma das 11 contas. Como exemplo, para o tópico “Política” e para a conta @camaradeputados, a data e horário de extração das informações e da pontuação bruta geral do Botometer ocorreram em 30/10/2022 às 14:00, sendo necessários 42 *tweets* para coletar 200 usuários que curtiram os *tweets* dessa conta.

Tabela 5: Informações sobre usuários legítimos coletados

Tópico	Conta	Data e horário	Nº tweets
Entretenimento	@netflixbrasil	29/10/2022 às 14:48	3
	@spotifybrasil	29/10/2022 às 15:20	9
	@primevideobr	29/10/2022 às 15:04	4
Esporte	@geglobo	29/10/2022 às 14:32	8
	@espnbrasil	29/10/2022 às 14:19	14
	@tntsportsbr	29/10/2022 às 15:46	5
Notícias gerais	@folha	30/10/2022 às 13:19	3
	@cnmbrasil	30/10/2022 às 13:38	7
	@globonews	30/10/2022 às 13:44	3
Política	@senadofederal	30/10/2022 às 13:55	21
	@camaradeputados	30/10/2022 às 14:00	42

Como a base de dados de usuários rotulados como *bot* tem 679 contas, então, para haver um equilíbrio na base de dados total, decidiu-se que a base de dados de usuários legítimos teria 600 contas. Assim, cada tópico contribuiu com 150 contas para a base de dados de usuários legítimos. As 600 contas com pontuação bruta geral do Botometer no intervalo $[0; 0,2]$ foram escolhidas aleatoriamente utilizando o R na versão 4.2.1. Vale ressaltar que em toda base de dados, os usuários são únicos, ou seja, não há usuários repetidos.

A Tabela 6 mostra para cada tópico quanto cada conta contribuiu com usuários legítimos para a base de dados. Como exemplo, para o tópico “Entretenimento”, a conta @netflixbrasil participou com 50 usuários legítimos para a base de dados.

Tabela 6: Informações sobre quantidade de usuários legítimos por tópico e conta

Tópico	Conta	Quantidade de usuários
Entretenimento	@netflixbrasil	50
	@spotifybrasil	50
	@primevideobr	50
Esporte	@geglobo	50
	@espnbrasil	50
	@tntsportsbr	50
Notícias gerais	@folha	50
	@cnnbrasil	50
	@globonews	50
Política	@senadofederal	75
	@camaradeputados	75
Total	11 contas	600

Informações sobre os *scripts* utilizados para coletar usuários legítimos podem ser encontradas em <https://github.com/pmmp-tcc/usuarios-legitimos>.

A base de dados total contou com 1.279 contas, sendo 679 rotuladas como bots e 600 rotuladas como legítimas.

2.7 Procedimento para extração das publicações dos usuários

As publicações extraídas dos usuários poderiam ser *tweet*, *retweet* ou *reply* (resposta).

O procedimento ocorreu da seguinte forma:

- com a informação da id dos usuários, foram extraídas as 50 publicações mais recentes de cada usuário;
- informações sobre cada publicação foram extraídas.

As informações coletadas sobre cada publicação serão descritas na Seção 3.2.

Esse procedimento ocorreu tanto para contas rotuladas como *bot* quanto para contas rotuladas como legítimas.

As datas de coleta das publicações dos usuários foram 31/10/2022 para contas classificadas como *bots* e legítimas e 17/11/2022 e 03/12/2022 para algumas contas classificadas como legítimas.

Informações sobre o *script* utilizado para obter informações sobre as publicações podem ser encontradas em <https://github.com/pmmp-tcc/tweets>.

3 Análise dos resultados

Neste capítulo serão apresentados os resultados obtidos neste trabalho. Na Seção 3.1 serão explicadas as variáveis coletadas dos usuários do Twitter. Na Seção 3.2 serão detalhadas as variáveis coletadas das publicações dos usuários. A Seção 3.3 mostrará os resultados das análises de nove variáveis, sendo seis quantitativas e três qualitativas.

3.1 Variáveis coletadas dos usuários

As variáveis coletadas dos usuários utilizando a API do Twitter ¹ são:

- `id`: número inteiro e único de identificação do usuário (variável qualitativa);
- `screen_name`: nome de usuário exclusivo da conta e que começa com o símbolo “@” (variável qualitativa);
- `location`: local definido pelo usuário para o perfil da sua conta (variável qualitativa);
- `description`: texto da bio que descreve a conta do usuário (variável qualitativa);
- `followers_count`: número de seguidores que a conta tem atualmente (variável quantitativa);
- `friends_count`: número de usuários que a conta está seguindo (variável quantitativa);
- `created_at`: data (dia, mês e ano) e hora UTC (*Coordinated Universal Time*) em que a conta do usuário foi criada (variável qualitativa);
- `favourites_count`: quantidade de *tweets* que o usuário curtiu (variável quantitativa);
- `statuses_count`: quantidade de *tweets* (incluindo *retweets*) publicados pelo usuário (variável quantitativa);

¹<https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/user>

- `time_line`: 20 publicações mais recentes da conta podendo ser *tweet*, *retweet* ou resposta (*reply*) (variável qualitativa).

3.2 Variáveis coletadas das publicações dos usuários

As variáveis coletadas das publicações, que podem ser *tweets*, *retweets* ou *reply* (respostas), utilizando a API do Twitter ² são:

- `user`: usuário que publicou o *tweet* ou *retweet* ou resposta. É um número inteiro e único de identificação do usuário (variável qualitativa);
- `created_at`: data (dia, mês e ano) e hora UTC (*Coordinated Universal Time*) em que a publicação foi criado (variável qualitativa);
- `id_str`: representação numérica que identifica exclusivamente a publicação (variável qualitativa);
- `text`: texto da publicação (variável qualitativa);
- `lang`: identifica o idioma do texto da publicação (variável qualitativa);
- `source`: utilitário usado para publicar o texto, no formato HTML - *HyperText Markup Language* (variável qualitativa);
- `in_reply_to`: identifica o nome de usuário da publicação original, no caso de uma resposta (variável qualitativa);
- `coordinates`: representa a localização geográfica da publicação (longitude e latitude), informada pelo aplicativo do usuário (variável qualitativa);
- `place`: indica que a publicação está associada ao lugar, mas não necessariamente é originária deste lugar (variável qualitativa);
- `is_quote_status`: indica se é um *tweet* citado, isto é, um *retweet* com comentários acima do *tweet* original (variável qualitativa - Sim/Não);
- `retweet_count`: número de vezes que o *tweet* foi retweetado (variável quantitativa);
- `favourite_count`: indica aproximadamente quantas vezes o *tweet* foi curtido por outros usuários do Twitter (variável quantitativa);

²<https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>

3.3 Resultados

As variáveis analisadas e o tipo de cada variável estão detalhadas na Tabela 7.

Tabela 7: Descrição das variáveis analisadas

Variável	Tipo de variável
Número de seguidores da conta	Quantitativa
Número de usuários que a conta segue	Quantitativa
Número de tweets que o usuário curtiu	Quantitativa
Número de tweets publicados pelo usuário	Quantitativa
Idade das contas	Quantitativa
Número de tweets publicados pelo usuário por dia	Quantitativa
Idioma das publicações dos usuários	Qualitativa
Local definido pelo usuário para o perfil da sua conta	Qualitativa
Nome de usuário da conta	Qualitativa

Para cada variável quantitativa, tabela com algumas medidas resumo e figura com gráfico *boxplot* serão apresentadas. Para esse tipo de gráfico, foi necessário remover valores discrepantes para melhor visualização do comportamento da variável. Além disso, para as medidas resumo, foi considerado arredondamento para duas casas decimais, em caso de necessidade. Por sua vez, para cada variável qualitativa, será apresentada figura com gráfico de barras. Todas as análises foram feitas no *software* R na versão 4.2.1.

3.3.1 Variável número de seguidores

Pela Tabela 8 verifica-se que para a variável número de seguidores, os valores encontrados de média e mediana para cada rótulo não são iguais ou próximos. Como a média é uma medida de tendência central influenciada por valores atípicos, a mediana, que é uma medida mais robusta que a média, torna-se mais apropriada. Assim, a mediana das contas rotuladas como *bots* foi maior (843 seguidores) do que as contas rotuladas como legítimas (166 seguidores). Em relação a valores mínimos e máximos, ambos os rótulos têm valor mínimo igual a zero, ou seja, não seguem qualquer outro usuário, enquanto as contas rotuladas como *bots* apresentam valor máximo (165.965 seguidores) superior às contas rotuladas como legítimas (13.260 seguidores). No que se refere ao desvio padrão, as

contas classificadas como *bots* apresentaram valor maior (9.581,85 seguidores), o que indica maior heterogeneidade no número de seguidores. Na Figura 5 estão os *boxplots* do número de seguidores, evidenciando maior homogeneidade das contas classificadas como legítimas em comparação às contas rotuladas como *bots*. Ademais, pode-se concluir que 75% das contas com rótulo de legítimas apresentaram número de seguidores menor ou igual a 500, aproximadamente, ao passo que para as contas com rótulo de *bots*, 75% apresentaram número de seguidores menor ou igual a 2.500, aproximadamente.

Tabela 8: Medidas resumo da variável número de seguidores

Estatística	Rótulo	
	Bot	Legítima
Média	3.110,44	576,62
Mediana	843	166
Valor mínimo	0	0
Valor máximo	165.965	13.260
Desvio padrão	9.581,85	1.323,75

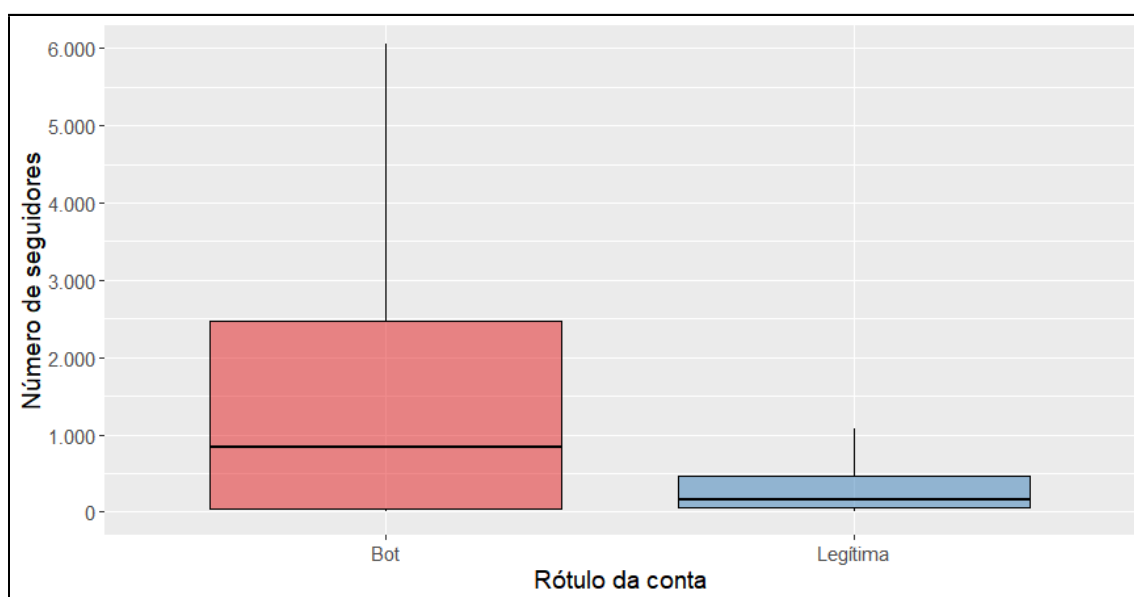


Figura 5: Boxplot do número de seguidores

3.3.2 Variável número de usuários que a conta segue

No que diz respeito à variável número de usuários que a conta segue, nota-se, na Tabela 9, valores de média e mediana diferentes para cada rótulo. Para as contas rotuladas como *bots*, a média ficou em 3.066,94 usuários seguidos e a mediana ficou em 841 usuários

seguidos, enquanto para as contas rotuladas como legítimas, a média foi de 758,48 usuários seguidos e a mediana foi de 402,50 usuários seguidos. Dessa forma, a mediana é mais adequada. Logo, constata-se que a mediana das contas classificadas como *bots* é mais que duas vezes a mediana das contas classificadas como legítimas, mostrando que as contas rotuladas como *bots* seguem maior quantidade de usuários. Observa-se que as contas com rótulo de *bots* apresentaram valor mínimo, valor máximo e desvio padrão iguais a zero, 91.580 e 6.638,99 usuários que a conta segue, respectivamente. Já as contas rotuladas como legítimas apresentaram valor mínimo, valor máximo e desvio padrão iguais a sete, 7.982 e 1.047,25 usuários que a conta segue, respectivamente. Assim, fica claro que as contas rotuladas como *bots* possuem maior variabilidade. A Figura 6 evidencia esse fato e revela que, para as contas com rótulo de legítimas, 25% apresentaram número de usuários que seguem maior ou igual a 900, aproximadamente, e, para as contas com rótulo de *bots*, 25% apresentaram número de usuários que seguem maior ou igual a 3.800, aproximadamente.

Tabela 9: Medidas resumo da variável número de usuários que a conta segue

Medidas resumo	Rótulo	
	Bot	Legítima
Média	3.066,94	758,48
Mediana	841	402,50
Valor mínimo	0	7
Valor máximo	91.580	7.982
Desvio padrão	6.638,99	1.047,25

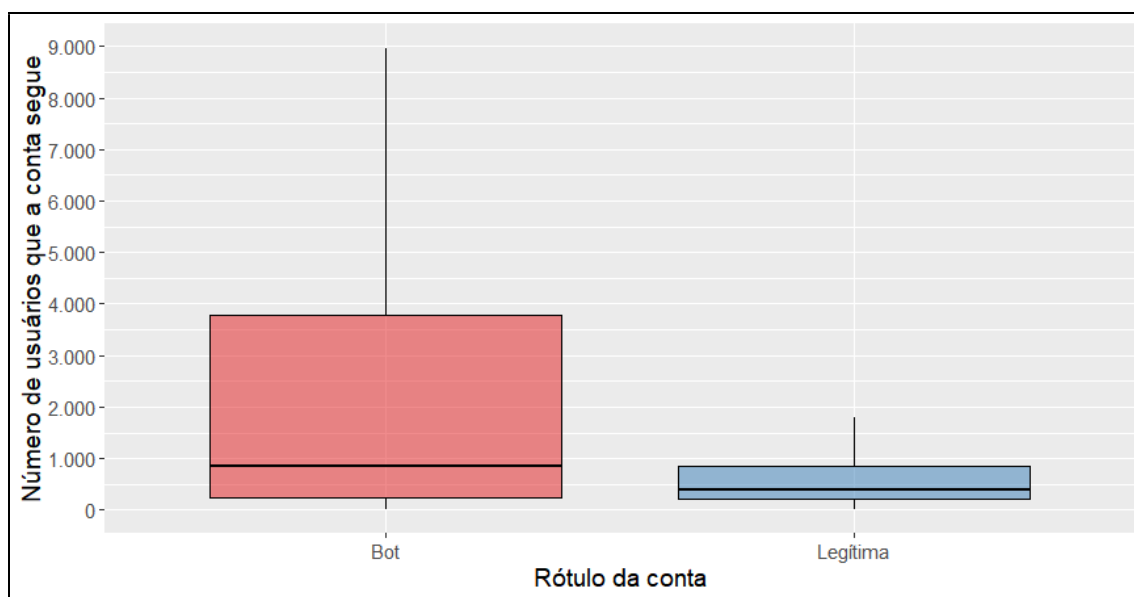


Figura 6: Boxplot do número de usuários que a conta segue

3.3.3 Variável número de tweets que o usuário curtiu

Pela Tabela 10 é possível observar que para a variável número de *tweets* que o usuário curtiu, os valores de média e mediana para cada um dos rótulos são distintos. Portanto, mais uma vez, a mediana torna-se uma melhor medida para descrever a variável do que a média. A mediana das contas rotuladas como *bots* foi de 1.420, ao passo que a mediana das contas rotuladas como legítimas foi de 34.688, ou seja, tais contas parecem curtir mais *tweets* que as contas classificadas como *bots*. Com relação a valores mínimos e máximos, as contas rotuladas como *bots* apresentaram valores 0 e 302.046, respectivamente, e as contas rotuladas como legítimas apresentaram valores 124 e 1.066.453, respectivamente. Os valores de desvio padrão apontam que as contas classificadas como legítimas possuem maior heterogeneidade no que se refere ao número de *tweets* que o usuário curtiu. Na Figura 7 verifica-se maior variabilidade das contas classificadas como legítimas, em que 75% delas apresentaram número de *tweets* que o usuário curtiu menor ou igual a 90.000, aproximadamente. Em contrapartida, para as contas classificadas como *bots*, 75% apresentaram número de *tweets* que o usuário curtiu menor ou igual a 8.000, aproximadamente.

Tabela 10: Medidas resumo da variável número de tweets que o usuário curtiu

Medidas resumo	Rótulo	
	Bot	Legítima
Média	9.615,40	68.865,44
Mediana	1.420	34.688
Valor mínimo	0	124
Valor máximo	302.046	1.066.453
Desvio padrão	24.956,13	96.476,67

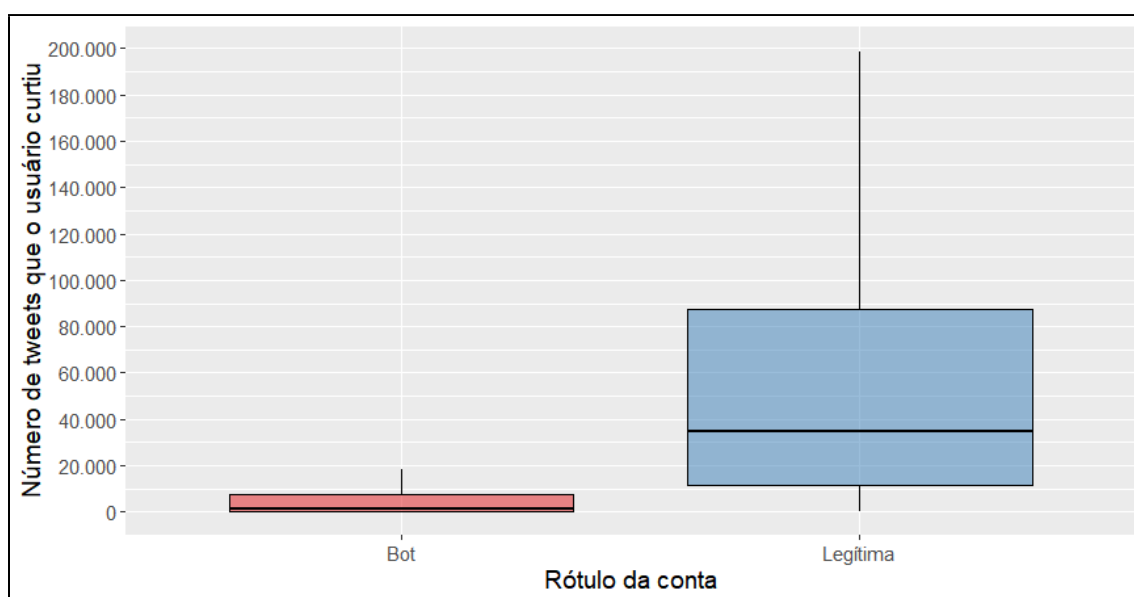


Figura 7: Boxplot do número de tweets que o usuário curtiu

3.3.4 Variável número de tweets publicados pelo usuário

No que se refere à variável número de *tweets* (incluindo *retweets*) publicados pelo usuário, pela Tabela 11 verifica-se que, assim como em outras variáveis, a mediana parece ser uma medida mais adequada do que a média. Logo, conclui-se que as contas rotuladas como legítimas apresentaram mediana maior (3.527 *tweets* publicados pelo usuário) do que as contas rotuladas como *bots* (1.358 *tweets* publicados pelo usuário). Os valores mínimos e máximos foram de um e 318.413 *tweets* publicados pelo usuário, respectivamente, para as contas classificadas como *bots*, e de 54 e 450.651 *tweets* publicados pelo usuário, respectivamente, para as contas classificadas como legítimas. Em relação à variabilidade, as contas rotuladas como *bots* são mais homogêneas, apresentando desvio padrão igual a 26.572,69 *tweets* publicados pelo usuário, enquanto as contas rotuladas como legítimas

apresentaram desvio padrão igual a 31.008,59 *tweets* publicados pelo usuário. Na Figura 8 nota-se que as contas rotuladas como *bots* são mais homogêneas, e pode-se constatar que 75% delas apresentam até 6.200 *tweets* publicados pelo usuário, aproximadamente. Para as contas rotuladas como legítimas, 75% possuem até 11.500 *tweets* publicados pelo usuário, aproximadamente.

Tabela 11: Medidas resumo da variável número de tweets publicados pelo usuário

Medidas resumo	Rótulo	
	Bot	Legítima
Média	9.511,87	13.218,01
Mediana	1.358	3.527
Valor mínimo	1	54
Valor máximo	318.413	450.651
Desvio padrão	26.572,69	31.008,59

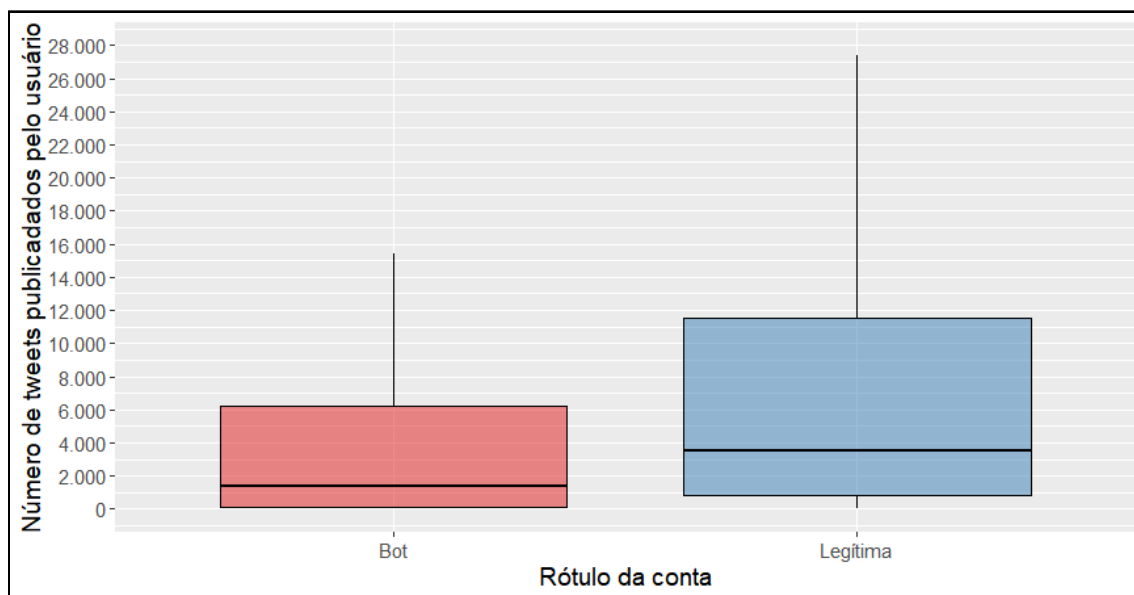


Figura 8: Boxplot do número de tweets publicados pelo usuário

3.3.5 Variável idade das contas

Para obter a variável idade das contas (em dias), foi necessário considerar como data inicial a data de criação da conta, e como data final a data em que foi extraída a pontuação bruta geral do classificador Botometer. Na Tabela 12 nota-se que média

e mediana apresentam valores diferentes para cada rótulo das contas, sendo, portanto, necessário utilizar a mediana no lugar da média. A idade mediana das contas classificadas como legítimas foi de 2.049,50 dias, ao passo que das contas classificadas como *bots* foi de 311 dias. Tal fato indica que as contas rotuladas como *bots* são mais novas, mais recentes do que as contas rotuladas como legítimas. Essa situação ocorre porque as contas rotuladas como *bots* violam as regras do Twitter mais facilmente e, como consequência, são suspensas frequentemente, tornando necessária a criação de novas contas para continuar a poluir conteúdos. Com relação a valores mínimos, as contas classificadas como *bots* e como legítimas apresentaram um dia e 34 dias de idade, respectivamente. Por sua vez, para os valores máximos, as contas classificadas como *bots* e como legítimas apresentaram 5.077 e 5.357 dias de idade, respectivamente. Analisando o desvio padrão, constata-se que as contas rotuladas como legítimas possuem maior variabilidade, com 1.589,96 dias. Pela Figura 9 pode-se concluir que para as contas classificadas como *bots*, 75% delas têm idade de até 1.000 dias, aproximadamente, enquanto para as contas classificadas como legítimas, 75% delas têm idade superior a 1.000 dias, aproximadamente.

Tabela 12: Medidas resumo da variável idade das contas (dias)

Medidas resumo	Rótulo	
	Bot	Legítima
Média	966,06	2.444,36
Mediana	311	2.049,50
Valor mínimo	1	34
Valor máximo	5.077	5.357
Desvio padrão	1.368,91	1.589,96

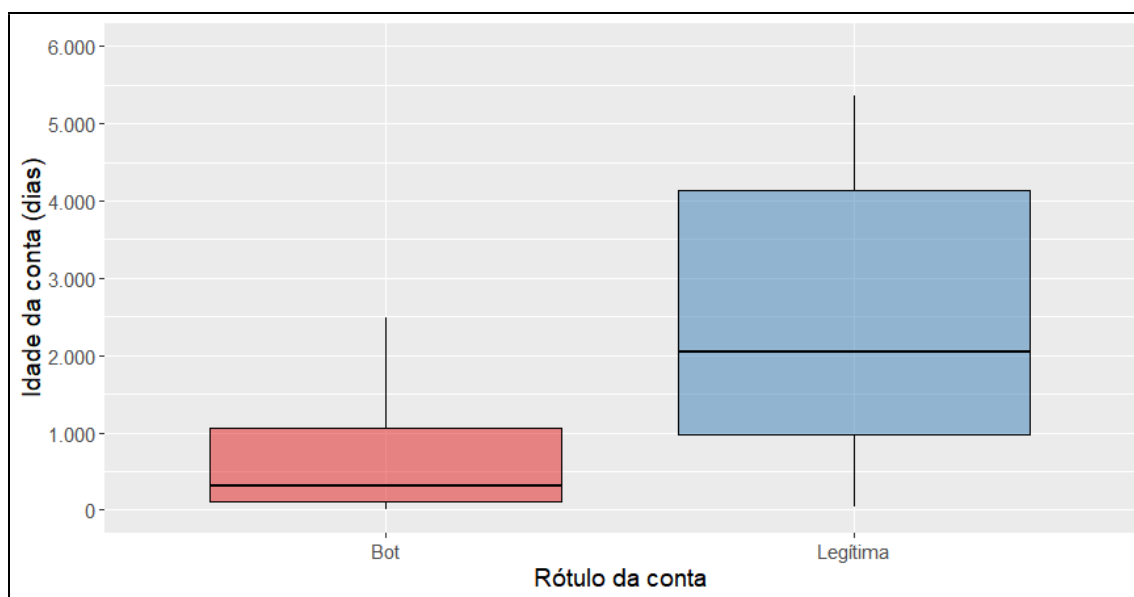


Figura 9: Boxplot da idade das contas (dias)

3.3.6 Variável número de tweets publicados pelo usuário por dia

Para obter a variável número de *tweets* (incluindo *retweets*) publicados pelo usuário por dia, foi necessário dividir a variável número de *tweets* publicados pelo usuário pela variável idade das contas (em dias). A Tabela 13 mostra que para a variável número de *tweets* (incluindo *retweets*) publicados pelo usuário por dia, a média e a mediana de cada rótulo apresentam valores distintos e, portanto, será utilizada a mediana no lugar da média. Assim, observa-se que o número mediano de *tweets* publicados pelo usuário por dia é maior para as contas classificadas como *bot* (3,26) do que para as contas classificadas como legítimas (1,83). Quanto a valores mínimos e máximos, os valores encontrados foram 0 e 538,31 *tweets* publicados pelo usuário por dia, respectivamente, para as contas rotuladas como *bots* e 0,02 e 121,64 *tweets* publicados pelo usuário por dia, respectivamente, para as contas rotuladas como legítimas. Relativamente à dispersão, verifica-se que as contas classificadas como *bots* são mais heterogêneas, apresentando desvio padrão maior (44,01) do que as contas classificadas como legítimas (12,82). Pela Figura 10 é possível confirmar a variabilidade maior das contas rotuladas como *bots*. Ademais, 75% dessas contas possuem número de *tweets* publicados pelo usuário por dia menor ou igual a 20, aproximadamente, ao passo que para as contas rotuladas como legítimas, 75% possuem número de *tweets* publicados pelo usuário por dia menor ou igual a 7, aproximadamente.

Tabela 13: Medidas resumo da variável número de tweets publicados pelo usuário por dia

Medidas resumo	Rótulo	
	Bot	Legítima
Média	18,85	6,47
Mediana	3,26	1,83
Valor mínimo	0	0,02
Valor máximo	538,31	121,64
Desvio padrão	44,01	12,82

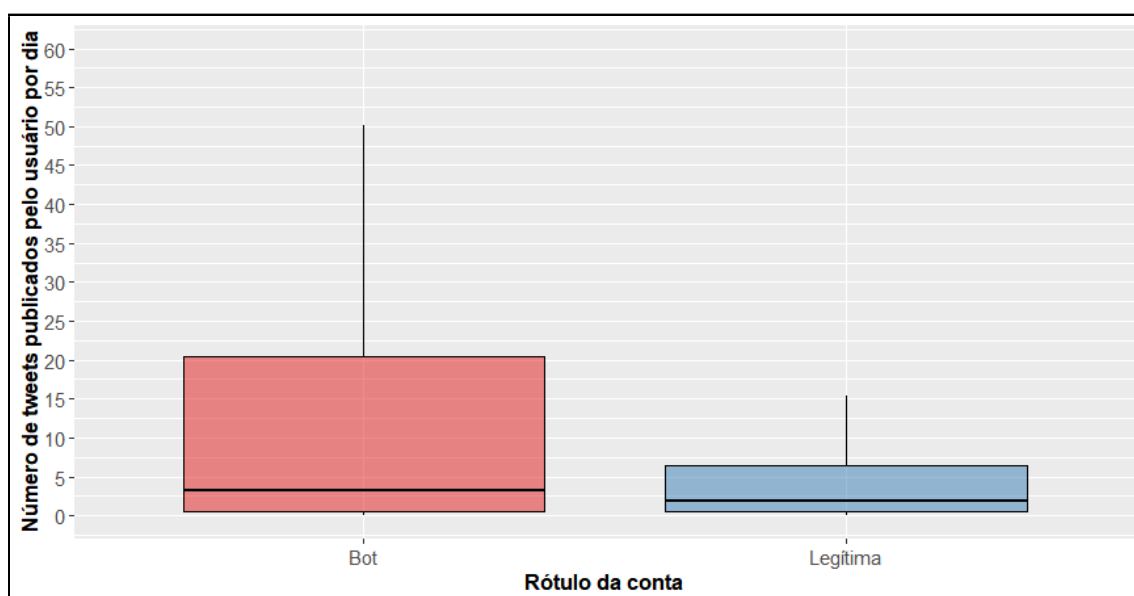


Figura 10: Boxplot do número de tweets publicados pelo usuário por dia

3.3.7 Variável idioma das publicações do usuário

Para a variável idioma das publicações do usuário, foi necessário criar categorias, em que 0 indica se o usuário publicou apenas em português e 1 indica se o usuário publicou somente em outro idioma ou publicou em português e em outro idioma. As publicações do usuário consideradas podem ser *tweets*, *retweets* ou resposta (*reply*). Do total de 1.279 contas da base de dados, foi analisado o idioma de 1.257 contas. As demais contas ou estavam suspensas ou foram deletadas, impossibilitando, portanto, a verificação do idioma. A Figura 11 revela que, em relação às contas que publicam apenas em português, há maior quantidade de contas classificadas como legítimas (470 contas) do que contas classificadas como *bots* (383 contas). Por outro lado, para as contas que publicam em outro idioma, há maior número de contas rotuladas como *bots* (274 contas) do que contas rotuladas

como legítimas (130 contas). Esse último resultado se deve, principalmente, ao uso, pelas contas *honeypots*, de palavras ou *hashtags* em inglês como #followback, #bitcoins, #nft, #followme e #retweet.

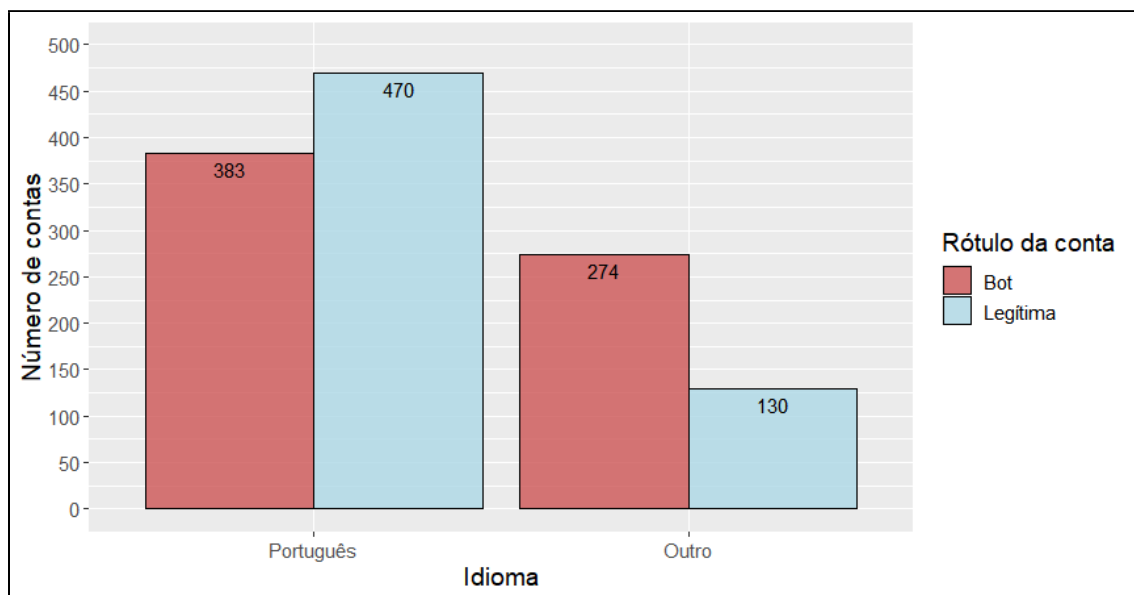


Figura 11: Gráfico de barras do idioma das publicações do usuário

3.3.8 Variável local definido pelo usuário para o perfil da sua conta

Para a variável local definido pelo usuário para o perfil da sua conta, foram criadas três categorias: localização específica, quando o usuário informa cidade e/ou estado; localização superficial, quando o usuário indica apenas o país; e não informada, quando o usuário não informa localização ou a localização não existe ou não é possível identificá-la. A Figura 12 mostra que 316 e 216 usuários de contas classificadas como legítimas e *bots*, respectivamente, informaram a localização específica da conta. Por sua vez, 40 usuários de cada rótulo de conta informaram a localização superficial da conta, totalizando 80 usuários. Usuários de 423 e 244 contas classificadas como *bots* e legítimas, respectivamente, não informaram a localização da conta.

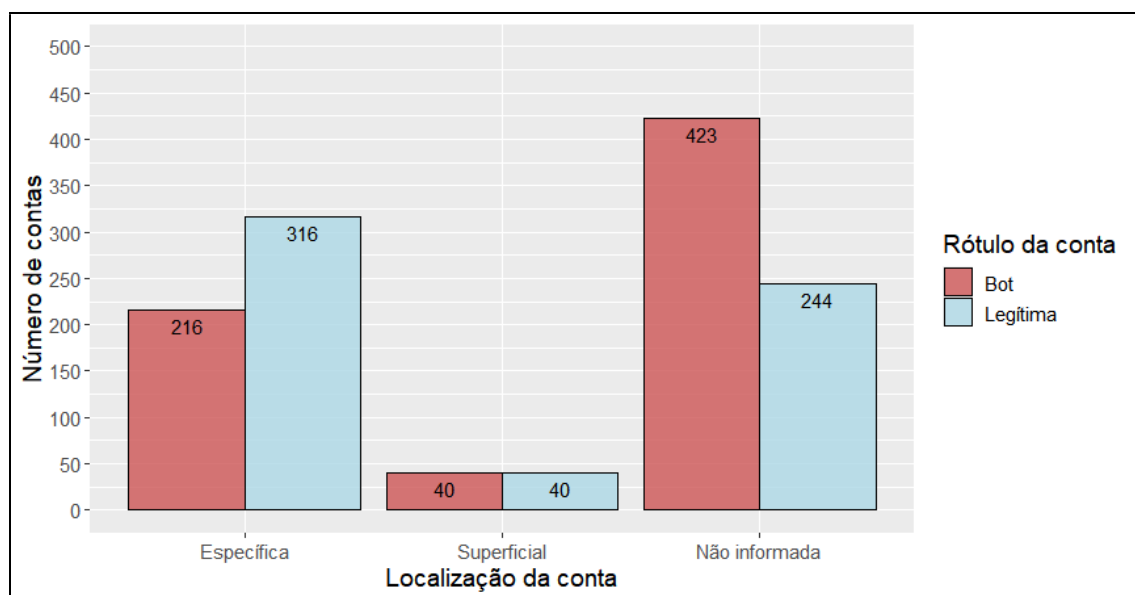


Figura 12: Gráfico de barras do local definido pelo usuário para o perfil da sua conta

3.3.9 Variável nome de usuário da conta

Em relação à variável nome de usuário da conta, foram criadas duas categorias: nome de usuário com até dois dígitos, quando o nome de usuário apresenta quantidade de dígitos menor ou igual a dois (exemplo: @fulano12); e nome de usuário com mais de dois dígitos, quando o nome de usuário apresenta quantidade de dígitos maior do que dois (exemplo: @fulano1234). Observa-se pela Figura 13 que para nome de usuário com até dois dígitos, há maior quantidade de contas classificadas como legítimas (542 contas) do que contas classificadas como *bots* (488 contas). Por sua vez, se o nome de usuário possuir mais de dois dígitos, há maior número de contas rotuladas como *bots* (191 contas) do que contas rotuladas como legítimas (58 contas). Tal fato acontece porque, ao criar uma nova conta, o Twitter sugere nomes de usuários com alguns dígitos e, sendo a conta automatizada, ela possivelmente ficará com a primeira sugestão dada.

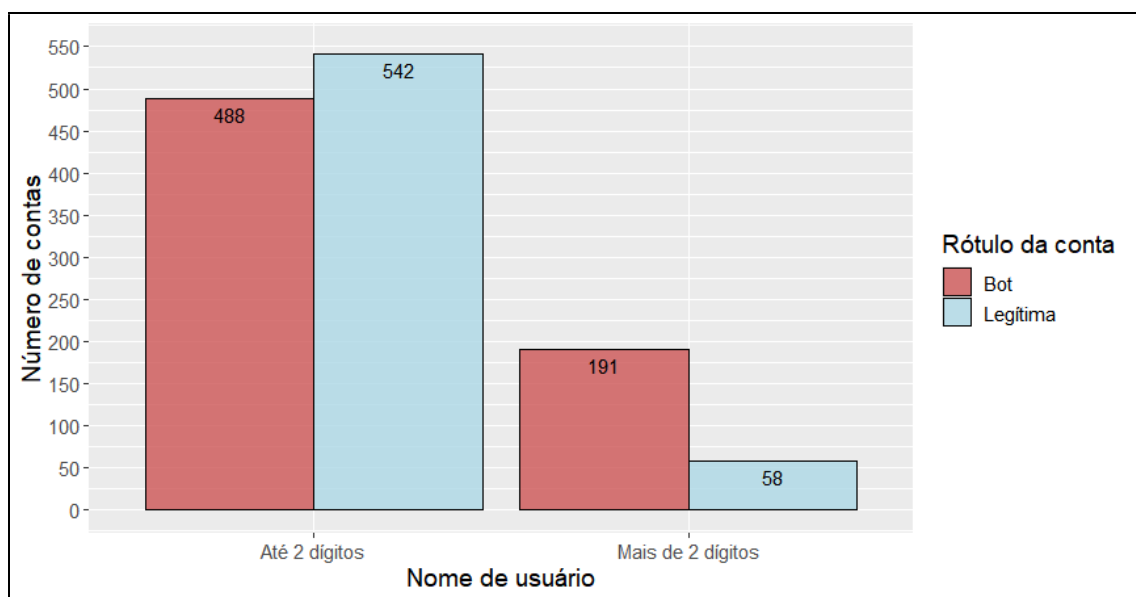


Figura 13: Gráfico de barras do nome de usuário da conta

4 Conclusões

Este trabalho apresentou dois objetivos. O primeiro e principal objetivo consistiu na criação de uma base de dados de perfis do Twitter rotulados em *bot* ou em legítimo. O segundo objetivo envolveu fazer uma análise comparativa das características de perfis *bots* e legítimos.

Para a criação da base de dados, foram empregados dois métodos diferentes. O primeiro método, para perfis rotulados em *bots*, consistiu na utilização de contas *honeypots*, sendo que elas somente publicavam *retweets* que incluíssem um termo, que poderia ser uma palavra ou *hashtag*. O segundo método, para perfis rotulados como legítimos, envolveu selecionar 11 contas de quatro tópicos distintos e extrair informações de usuários que curtiram os *tweets* mais recentes de cada uma das 11 contas.

Em ambos os métodos, após a coleta de possíveis usuários *bots* e legítimos, foram realizadas duas verificações manuais por meio de um questionário, a fim de obter os limites da pontuação bruta geral do classificador Botometer para definição dos usuários legítimos e *bots*. Assim, ficou definido que para entrar na base de dados com o rótulo de *bot*, era necessário que a conta apresentasse pontuação bruta geral dada pelo Botometer no intervalo $[0,5; 1]$. Por sua vez, para entrar na base de dados com o rótulo de legítima, era necessário que a conta apresentasse pontuação bruta geral dada pelo Botometer no intervalo $[0; 0,2]$. No total, a base de dados contou com 1.279 contas, sendo 679 contas rotuladas como *bots* e 600 contas rotuladas como legítimas. Para a base de dados completa, foram extraídas informações da conta, bem como suas publicações, que poderiam ser *tweet*, *retweet* ou *reply* (resposta).

As variáveis analisadas das contas rotuladas como *bots* ou legítimas foram: número de seguidores da conta (variável quantitativa), número de usuários que a conta segue (variável quantitativa), número de *tweets* que o usuário curtiu (variável quantitativa), número de *tweets* (incluindo *retweets*) publicados pelo usuário (variável quantitativa), idade das contas (em dias) (variável quantitativa), número de *tweets* (incluindo *retweets*) publicados

pelo usuário por dia (variável quantitativa), idioma das publicações do usuário (variável qualitativa), local definido pelo usuário para o perfil da sua conta (variável qualitativa) e nome de usuário da conta (variável qualitativa).

Das seis variáveis quantitativas, foi verificado que para três delas (número de seguidores, número de usuários que a conta segue e número de *tweets* publicados pelo usuário por dia), as contas rotuladas como *bots* apresentaram maior heterogeneidade, conforme relatado por Sayyadiharikandeh et al. (2020).

As variáveis qualitativas mostraram que as contas classificadas como legítimas publicaram mais apenas em português, mais forneceram a localização específica e apresentaram maior quantidade de nome de usuários com até dois dígitos do que as contas classificadas como *bots*.

Uma das dificuldades deste trabalho foi a criação de novas contas no Twitter para serem utilizadas como contas *honeypots*, à medida que as contas implementadas fossem suspensas por violarem as regras do Twitter. Outra dificuldade encontrada foi rotular as contas em *bots* ou em legítimas. Para tanto, foi utilizado o classificador Botometer, todavia, para idiomas diferentes do inglês, ele exclui as características linguísticas, fornecendo uma pontuação independente do idioma. Embora o Botometer seja amplamente utilizado na literatura, seria interessante o uso de um classificador específico para contas brasileiras que publicam em português.

Como propostas para trabalhos futuros, recomenda-se analisar possíveis padrões nas publicações de usuários rotulados como *bots*, realizar análise de sentimentos, classificar os *bots* da base de dados em maliciosos ou benignos, implementar mais análises estatísticas para diferenciar os usuários *bots* de usuários legítimos, aumentar a base de dados realizando os métodos e procedimentos relatados neste trabalho, e utilizar a base de dados criada para aprimorar classificadores de contas do Twitter.

Referências

- BESSI, A.; FERRARA, E. Social bots distort the 2016 us presidential election online discussion. *First monday*, v. 21, n. 11-7, 2016.
- Botometer. *FAQ*. 2022. (<http://botometer.osome.iu.edu/faq>), acesso em: 20 de junho de 2022.
- BRONIATOWSKI, D. A. et al. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American journal of public health*, American Public Health Association, v. 108, n. 10, p. 1378–1384, 2018.
- CRESCI, S. et al. Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, Elsevier, v. 80, p. 56–71, 2015.
- DAVIS, C. A. et al. Botornot: A system to evaluate social bots. In: *Proceedings of the 25th international conference companion on world wide web*. [S.l.: s.n.], 2016. p. 273–274.
- ELMENDILI, F. et al. A security approach based on honeypots: Protecting online social network from malicious profiles. *arXiv preprint arXiv:1804.09988*, 2018.
- FAN, R.; TALAVERA, O.; TRAN, V. Social media bots and stock markets. *European Financial Management*, Wiley Online Library, v. 26, n. 3, p. 753–777, 2020.
- FENG, S. et al. Twibot-20: A comprehensive twitter bot detection benchmark. In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. [S.l.: s.n.], 2021. p. 4485–4494.
- FERRARA, E. What types of covid-19 conspiracies are populated by twitter bots? *arXiv preprint arXiv:2004.09531*, 2020.
- FERRARA, E. et al. Predicting online extremism, content adopters, and interaction reciprocity. In: SPRINGER. *International conference on social informatics*. [S.l.], 2016. p. 22–39.
- GILANI, Z. et al. Of bots and humans (on twitter). In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. [S.l.: s.n.], 2017. p. 349–354.

GORODNICHENKO, Y.; PHAM, T.; TALAVERA, O. Social media, sentiment and public opinions: Evidence from# brexit and# uselection. *European Economic Review*, Elsevier, v. 136, p. 103772, 2021.

LEE, K.; EOFF, B.; CAVERLEE, J. Seven months with the devils: A long-term study of content polluters on twitter. In: *Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2011. v. 5, n. 1, p. 185–192.

LOKOT, T.; DIAKOPOULOS, N. News bots: Automating news and information dissemination on twitter. *Digital Journalism*, Taylor & Francis, v. 4, n. 6, p. 682–699, 2016.

MARLOW, T.; MILLER, S.; ROBERTS, J. T. Twitter discourses on climate change: exploring topics and the presence of bots. SocArXiv, 2020.

NICULESCU-MIZIL, A.; CARUANA, R. Predicting good probabilities with supervised learning. In: *Proceedings of the 22nd international conference on Machine learning*. [S.l.: s.n.], 2005. p. 625–632.

PAULA, L. T. de; MICHALSKI, R. Os bots de disseminação de informação na conjuntura das campanhas presidenciais de 2018 no brasil. *Múltiplos Olhares em Ciência da Informação*, v. 9, n. 1, 2019.

RECUERO, R.; ZAGO, G. Em busca das “redes que importam”: redes sociais e capital social no twitter. *Líbero*, n. 24, p. 81–94, 2016.

REGATTIERI, L. L. *ALGORITMIZAÇÃO DA VIDA: o debate sobre Amazônia e incêndios florestais no Twitter em 2020*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2021.

SAVAGE, S.; MONROY-HERNANDEZ, A.; HÖLLERER, T. Botivist: Calling volunteers to action using online bots. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. [S.l.: s.n.], 2016. p. 813–822.

SAYYADIHARIKANDEH, M. et al. Detection of novel social bots by ensembles of specialized classifiers. In: *Proceedings of the 29th ACM international conference on information & knowledge management*. [S.l.: s.n.], 2020. p. 2725–2732.

SHEVTSOV, A. et al. Identification of twitter bots based on an explainable machine learning framework: The us 2020 elections case study. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2022. v. 16, p. 956–967.

Statista. *Leading countries based on number of Twitter users as of January 2022(in millions)*. 2022. (<http://www.statista.com/statistics/242606/>)

number-of-active-twitter-users-in-selected-countries/), acesso em: 20 de junho de 2022.

Statista. *Twitter accounts with the most followers worldwide as of April 2022(in millions)*. 2022. <<http://www.statista.com/statistics/273172/twitter-accounts-with-the-most-followers-worldwide/>>, acesso em: 20 de junho de 2022.

Twitter. *Bot ou não? Os fatos sobre manipulação de plataforma no Twitter*. 2022. <http://blog.twitter.com/pt_br/topics/company/2020/bot-ou-nao-fatos-sobre-manipulacao-de-plataforma-no-twitter>, acesso em: 20 de junho de 2022.

Twitter. *Glossary*. 2022. <<http://help.twitter.com/en/resources/glossary>>, acesso em: 20 de junho de 2022.

Twitter. *Twitter Announces Second Quarter 2022 Results*. 2022. <<http://investor.twitterinc.com/financial-information/quarterly-results/default.aspx>>, acesso em: 05 de dezembro de 2022.

VAROL, O. et al. Online human-bot interactions: Detection, estimation, and characterization. In: *Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2017. v. 11, n. 1.

YANG, K.-C.; FERRARA, E.; MENCZER, F. Botometer 101: Social bot practicum for computational social scientists. *arXiv preprint arXiv:2201.01608*, 2022.

YANG, K.-C.; TORRES-LUGO, C.; MENCZER, F. Prevalence of low-credibility information on twitter during the covid-19 outbreak. *arXiv preprint arXiv:2004.14484*, 2020.

YANG, K.-C. et al. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, Wiley Online Library, v. 1, n. 1, p. 48–61, 2019.

YANG, K.-C. et al. Scalable and generalizable social bot detection through data selection. In: *Proceedings of the AAAI conference on artificial intelligence*. [S.l.: s.n.], 2020. v. 34, n. 01, p. 1096–1103.

APÊNDICE 1 – Verificação manual - perguntas

1. Nome do perfil do usuário e nome do usuário (@) são iguais ou parecidos?
Sim(0)/Não(1)
2. Muitos dígitos no nome do usuário (mais que dois dígitos)? Sim(1)/Não(0);
3. Há imagem no perfil? Sim(0)/Não(1)
 - 3.1 Se sim, a imagem é uma foto pessoal? Sim(0)/Não ou NA(1)
4. Há imagem de fundo? Sim(0)/Não(1)
 - 4.1 Se sim, a imagem é uma foto pessoal? Sim(0)/Não ou NA(1)
5. Quantidade de perfis seguidos (*following*) maior ou igual a quantidade de seguidores (*followers*)? Sim(1)/Não(0)
6. A quantidade de perfis seguidos (*following*) é maior do que 1.000? Sim(1)/Não(0)
7. Há texto na descrição do perfil? Sim(0)/Não(1)
 - 7.1 Se sim, há informações pessoais na descrição? Sim(0)/Não ou NA(1)
 - 7.2 Se sim, o texto é original? Sim(0)/Não ou NA(1)
8. Na aba “Mídia”, há fotos ou vídeos pessoais? Sim(0)/Não(1)

Para as próximas perguntas, necessário analisar as últimas 20 publicações da conta.

1. Há muitos posts sem textos? Sim(1)/Não(0)
2. Faz muito uso de *emojis*? Sim(1)/Não(0)
3. Enaltece muito outros usuários? Sim(1)/Não(0)

4. Faz comentários que são coerentes com o *tweet*? Sim(0)/Não ou NA(1)
5. Só compartilha *links*? Sim(1)/Não(0)
6. Parece só retweetar? Sim(1)/Não(0)
7. Comportamento repetitivo com # ou menções? Sim(1)/Não(0)
8. Comportamento agressivo? Sim(1)/Não(0)
9. Publica apenas um assunto? Sim(1)/Não(0)