XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

DETECÇÃO DE *BOTS* QUE DIVULGAM ARTIGOS CIENTÍFICOS NO *TWITTER*: CONTRIBUIÇÕES PARA APRIMORAMENTO DOS INDICADORES ALTMÉTRICOS

DETECTION OF BOTS PUBLISHING SCIENTIFIC ARTICLES ON TWITTER: CONTRIBUTIONS TO IMPROVE ALTMÉTRIC INDICATORS

Danielle Pompeu Noronha Pontes. UEA.

João de Melo Maricato. UNB.

Modalidade: Trabalho Completo

Resumo: Agentes autônomos computacionais (*bots*) são softwares automáticos, que empregam inteligência artificial e automatizam processos informacionais. As informações (maliciosas ou não) disseminadas por estes agentes nas mídias sociais influenciam a tomada de decisões de diferentes atores sociais. Diferentemente de agentes humanos, os *bots* têm capacidade quase ilimitada de disseminação de informações, podendo dificultar, ou até inviabilizar, a interpretação de indicadores oriundos de mídias sociais. *Bots* automatizados em redes sociais como o *Twitter* aumentam a incerteza sobre os padrões de divulgação científica nas mídias e redes sociais, levantando a dúvida sobre a validade e confiabilidade desses dados para análises altmétricas. Diante disso, este artigo discute e testa maneiras de se detectar a atuação de *bots* em contas do *Twitter* que divulgam artigos científicos. Propõe-se um *workflow* para teste da acurácia em dados de treinamento e outro para a detecção preliminar de contas *bots* no contexto da altmetria, utilizando-se de técnicas de aprendizagem de máquina por meio da plataforma KNIME. O *workflow* proposto classificou 6.094 das 6.301 contas de usuários do *Twitter*, das quais 2.900 (48%) foram classificadas como humanos e 3.194(52%) como *bots.* Embora os dados sejam preliminares, acredita-se ser possível desenvolver metodologias capazes de aumentar a acurácia dos dados de maneira a vislumbrar maior credibilidade de uso da altmetria para a avaliação científica.

Palavras-Chave: Altmetria. Aprendizagem de Máquina. *Bots*. *Twitter*. Mídia social.

Abstract:*. Computational autonomous agents (bots) are automatic software that employ artificial intelligence and automate informational processes. The information (malicious or not) disseminated by these agents on social media influences decision-making. Unlike human agents, bots have an almost unlimited capacity to disseminate information, which can make it difficult or even impossible to interpret indicators from social media. Automated bots on social networks such as Twitter increase uncertainty about the standards of scientific dissemination in the media and social networks, raising doubts about the validity and reliability of this data for altmetric analysis. Therefore, this article discusses and tests ways to detect bots in Twitter accounts that publish scientific articles. A workflow is proposed for testing the accuracy of training data and another for the preliminary detection of bot accounts in the context of altmetrics, using machine learning techniques through the KNIME platform. The proposed workflow classified 6,094 out of 6,301 Twitter user accounts, of which 2,900 (48%) were classified as human and 3,194 (52%) as bots. Although the data are preliminary, it is believed that it is*

possible to develop methodologies capable of increasing the accuracy of the data in order to see greater credibility in the use of altmetrics for scientific evaluation.

Keywords: *Altmetric. Machine Learnign. Bots. Twitter. Social Media*.

1 INTRODUÇÃO

Existe uma diversidade crescente de "ecossistemas sociais" que sustentam o uso acadêmico das mídias sociais, os cientistas estão aproveitando o poder desse tipo de mídia para acelerar o ritmo em que eles estão desenvolvendo e compartilhando conhecimento, tanto no âmbito da comunidade científica como no âmbito do público em geral. As mídias sociais em geral, e o *Twitter* em particular, influenciam o ciclo de vida da publicação científica (DARLING et al., 2013). Para estes autores esta rede social traz benefícios ao aumentar as conexões a redes acadêmicas, o desenvolvimento mais rápido de ideias e discussões entre os pares, e a ampliação da disseminação e discussão do conhecimento científico dentro e fora da academia.

De maneira crescente as comunidades científicas começaram a adotar ativamente plataformas de mídia social para identificar o impacto e as influências da literatura acadêmica em públicos diversos, havendo interesse em saber se elas são capazes de oferecem indicadores complementares, aos indicadores bibliométricos, para avaliação científica (PRIEM e BRADELY, 2010). A captura e medição da circulação da ciência nas mídias e redes sociais têm o potencial de fornecer maneiras complementares de se compreender os fluxos de comunicação científica dentro e fora da academia.

Porém, o ambiente das mídias e redes sociais caracterizada pela Web 2.0, ao mesmo tempo que possibilitaram trocas de informações de maneira mais ágil e democrática, trouxeram dilemas que não existiam anteriormente, como, por exemplo, o uso de robôs automatizados ou comumente chamados de *bots.* Os *bots* não são necessariamente ruins, pois também podem ser utilizados para divulgar informações verdadeiras e relevantes. Mesmo assim, dependendo do contexto de uso eles podem ser considerados prejudiciais, pois podem interferir nos significados dos dados, indicadores e nos processos de tomada de decisão. Um exemplo disso é a produção de indicadores feita a partir dos rastros deixados por agentes humanos e não humanos em redes sociais. A existência de *bots* é uma das dificuldades de se utilizar indicadores de mídias sociais para o estudo e a avaliação dos impactos das produções científicas.

Os impactos da ciência (dentro da própria academia ou na sociedade em geral), as correlações entre citação e menção nas mídias sociais, as coberturas de menções, os tipos de comunidade de atenção, etc. não terão o mesmo significado se forem produzidos com dados de artigos científicos que contenham *bots* e agentes humanos. No caso do *Twitter* (DIDEGAH, MEJLGAARD; SØRENSEN, 2018), os *bots* automatizados aumentam a incerteza sobre os padrões de divulgação científica nessa rede social e levantam dúvidas sobre a confiabilidade dos seus dados, fonte para análises.

A compra de um *bot* para promover o próprio produto acadêmico é visto como inapropriado no contexto acadêmico. Porém, provavelmente, é aceitável que sejam utilizados para promover automaticamente a conscientização sobre importantes questões ambientais. Assim, a existência desse tipo de serviço e a prevalência de postagens feitas por *bots*, levantam questões sobre o uso de postagens de mídia social para avaliar a relevância social da ciência (CROTTY, 2014), bem como, sobre os limites aceitáveis sobre este tipo de divulgação ou sobre o marketing científico como um todo. Outro exemplo de divulgação de produtos acadêmicos nas redes sociais é o realizado pelas revistas científicas, que comumente promovem seus artigos nas mídias e redes sociais. Independentemente dessas questões, a identificação de *bots* pode facilitar análises dos tipos de contas que se tem interesse, diminuindo a ocorrência de vieses.

Diante do exposto, este trabalho se propõe a realizar um estudo preliminar sobre a detecção de *bots* que divulgam artigos científicos no *Twitter*, com o objetivo de propor um *workflow* para teste da acurácia em dados de treinamento e outro para a detecção preliminar de contas *bots no contexto da altmetria*, utilizando-se de técnicas de aprendizagem de máquina por meio da plataforma KNIME.

2 DETECÇÃO DE *BOTS* EM MÍDIAS SOCIAIS, NO *TWITTER E APLICAÇÕES NA ALTMETRIA*

Os *bots* são algoritmos que produzem conteúdos e interagem com os seus usuários. Esses mecanismos são responsáveis por uma proporção significativa das atividades online (FERRARA et al, 2016). Eles são agentes de software, que utilizam dados e informações sociais, com a finalidade de aumentar a sua capacidade de relacionamento, interação, persuasão e influência sobre atores humanos. O que leva à necessidade de atenção para o papel desses mecanismos nos processos e ciclos informacionais (NUNES, 2020).

Os *bots* muitas vezes são utilizados para disseminação maliciosa em massa de informações. Entretanto, eles não possuem apenas características maliciosas, podendo contar com características benignas e neutras. Exemplos de *bots* benignos incluem aqueles que postam automaticamente *tweets* de alertas climáticos, de *chats* para atendimento eletrônico e de disseminação de informações por agências de notícias.

Perfis nocivos são uma ameaça de segurança para redes sociais online, sendo grande o interesse e os esforços para detecta-los e eliminá-los. O procedimento de defesa mais utilizado contra os *bots* consistem em detectar estas contas e deletá-las, sendo realizado durante a criação de conta ou depois que se fundem nas mídias sociais. As principais abordagens para detecção de *bots* se dão pela análise das postagens e pelas informações dessas contas/perfis. Há uma prevalência de trabalhos que utilizam características comportamentais e de informações do usuário para detecção de *bots* maliciosos. Porém, existem diversas abordagens técnicas descritas na literatura.

Segundo Morais e Digiampietri (2021) um aspecto fundamental no processo de criação de mecanismos de detecção de *bots* reside na definição das características destes agentes a serem analisadas. Estas podem ser agrupadas em seis grupos distintos. Baseadas no usuário: número de amigos e seguidores, número de posts produzidos (por exemplo, número de *tweets*), descrição do perfil e configurações; Baseadas nos amigos: Relações com outros usuários da rede, incluindo o envio de menções, respostas e compartilhamento de mensagens, ser mencionado ou ter suas mensagens compartilhadas/encaminhadas; Baseadas na rede: diferentes tipos de interação na rede, considerando a frequência de iteração e coocorrências de hashtags; Temporais: consideram a atividade do usuário durante diferentes intervalos de tempo; Conteúdo: o tipo de linguagem utilizada, comprimento das mensagens e a entropia do texto sendo publicado; Sentimento: A atitude ou o humor de uma conversa ou mensagem.

Segundo Orabi et al. (2020) a maioria da literatura que tem por objetivo detectar *bots* e contas maliciosas, empregou técnicas de aprendizagem de máquina. Este fato sugere que a essa técnica é de grande importância na detecção desse tipo de agente nas mídias sociais. As três categorias principais das técnicas de Aprendizagem de Máquina (AM) são conhecidas como abordagens supervisionadas, não supervisionadas e semissupervisionadas.

Aprendizagem de máquina é uma subárea da Inteligência Artificial (IA). Algoritmos de AM têm sido amplamente utilizados em diversas tarefas, que podem ser divididas em Preditivas e Descritivas. Em tarefas preditivas (aprendizagem de máquina supervisionada), os algoritmos são aplicados a conjuntos de dados de treinamento rotulados para induzir um modelo capaz de predizer, para um novo objeto representado pelos valores de seus atributos preditivos, o valor de seu atributo alvo (FACELI, 2011).

A abordagem não supervisionada é utilizada em tarefas descritivas que, ao invés de predizer um valor, extraem padrões dos valores preditivos de um conjunto de dados. Essa não faz uso do conhecimento do "supervisor externo" como no caso das supervisionadas. A abordagem semissupervisionada é usada quando os dados não são rotulados, mas algumas restrições sobre eles devem ser conhecidas (FACELI, 2011). São utilizadas em tarefas de classificação em que apenas parte dos exemplos de treinamento possui um rótulo de classe (FACELI, 2011).

Martins (2010) e Neves (2020) consideram a Inteligência Artificial (IA) e seus produtos como entidades não humanas que atuam sobre a informação disponível em rede. Ressaltam ainda a importância dessas interações serem estudadas sob a perspectiva da Ciência da Informação. Isto porque estes atores não humanos interagem com seres humanos e podem modificar a realização de alguns processos informacionais e/ou atuar/modificar fluxos informacionais (FERRARA et al., 2016) e até mesmo criar regimes de informação próprios e autônomos.

Os estudos altmétricos não estão imunes a atuação destes *bots*. Segundo Aljohani, Fayoumi, Hassan (2020) é essencial analisar a influência social destes nos dados altmétricos. Os autores consideram fundamental investigar como esses mecanismos disseminam conteúdo e difundem informações e a porcentagem do conteúdo acadêmico produzido por eles nas mídias sociais. Diversos são as redes e mídias sociais existentes na atualidade e, por consequência, existe grande número de indicadores altmétricos que podem ser produzidos por meio delas. Dentre elas, o *Twitter* tem grande importância, pois é a mídia com maior cobertura de indicadores extraídos de agregadores como a plataforma Altmetric, sendo 91% do rastreamento de atividade social monitorado pela plataforma (CHAPMAN et al., 2018).

Inúmeros são os trabalhos encontrados para detecção de *bots* sociais no *Twitter* (MORAIS, 2021). Entretanto, poucos são direcionados para ação desses mecanismos em dados

altmétricos. Um dos estudos que tiveram foco na investigação do impacto destes *bots* na altmetria foi publicado por Haustein et al. (2016). Esses autores buscaram identificar a influência dos *bots* em índices altmétricos. Ao analisarem uma amostra aleatória de 800 contas de usuários do *Twitter*, constataram que 8% dos índices altmétricos foram completamente afetados por contas de *bot* e 5% parcialmente. O estudo ainda mostrou que 9% dos *tweets* para artigos do arXiv foram gerados por contas automatizadas. Descobriram, também, que as contas automatizadas que tuitam sobre temas acadêmicos se comportam de maneira diferente das contas gerais de *bots* do *Twitter*. Os estudos realizados, com contas automatizadas do *Twitter,* revelaram que eles publicam em média de 4,6 a 7,1 *tweets* por dia enquanto contas não automatizadas produzem, em média, 2,2 *tweets* por dia. Os autores também identificaram que os critérios de *tweeting,* utilizados pelos *bots,* são geralmente aleatórios e não qualitativos (eles podem, por exemplo, tuitar aleatoriamente frases predefinidas, como provérbios).

Dentre os estudos identificados, alguns propuseram a detecção de *bots* em redes sociais por meio de técnicas de aprendizagem de máquina, especificamente aplicados a dados altmétricos. Aaljohani, Fayoumi e Hassan (2020) observaram três tipos de técnicas de detecção de *bots* na literatura: método baseado na análise de informações de redes sociais, métodos baseados em inteligência e métodos baseados em aprendizagem de máquina que separam os *bots* e humanos. Os autores estudaram o comportamento e a influência desses mecanismos de mídias sociais em dados altmétricos aplicando mais de uma dessas técnicas de análises no Twitter. As análises revelaram que os *bots* influenciaram 87% dos *tweets* relacionados aos dados altmétricos, interferindo fortemente neste tipo de métrica.

Aaljohani, Fayoumi e Hassan (2020) também aplicaram uma técnica de rede convolucional de grafos para classificação de *bots* em um conjunto de dados altmétricos. Eles chegaram a uma precisão/acurácia de 71% e o F1-score de 0,67 (esse score refere-se a média harmônica entre a precisão e o *recall*, que está muito mais próxima dos menores valores do que uma média aritmética simples). Os autores esclarecem que não encontraram um conjunto de dados rotuladas por humanos para melhorar seu modelo (baseado em inteligência artificial). Assim, concluíram que a diferença entre contas de *bots* e humanos não são claras, sendo complexa a detecção até mesmo por humanos.

Chapman et al. (2018), utilizando-se de outra técnica, também realizou análises no contexto de dados da altmetria no *Twitter*. A pesquisa foi baseada em palavras-chave no *Twitter* com as contas mais ativas (que tuitaram pelo menos 1.000 vezes). Os autores relataram que entre 2.043 contas, 248 (12%) foram identificadas como automatizadas enquanto 305 (15%) foram identificadas como contas de editores ou periódicos (também automatizadas). Assim, 1.795 (88%) das contas encontradas foram classificadas como sendo de humanos e 27% como *bots*(CHAPMAN et al., 2018).

3 PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa caracteriza-se como sendo de natureza aplicada (ao buscar resolver um problema concreto), com enfoque quantitativo (visto que os dados utilizados são numéricos/estatísticos), de cunho exploratório e descritivo (buscando-se compreender e descrever um fenômeno em maior profundidade), utilizando-se de modelo experimental (onde uma ação é efetuada e os efeitos são medidos). Quanto aos procedimentos, utilizou-se técnicas de aprendizagem de máquina e de prova de conceito (ao ser proposto o desenvolvimento de um *workflow* utilizando o KNIME, que utiliza de aprendizagem de máquina, visando a classificação de contas do *Twitter*).

O desenvolvimento desta pesquisa teve como princípio a metodologia CRISP-DM (*CRoss-Industry Standard Process for Data Mining*). O modelo organiza, a partir de um ciclo de vida flexível, o processo de mineração de dados em seis fases. Cada uma dessas seis fases foi adaptada para o formato de um artigo científico, sendo elas: a compreensão de negócios (introdução e referencial teórico do artigo); as etapas compreensão de dados, preparação de dados (procedimentos metodológicos desta pesquisa) e, por fim, as etapas de modelagem, avaliação e implantação (seções resultados, discussões e conclusões do artigo).

3.1 Seleção de perfis no *Twitter (dataset* de teste)

Esta etapa teve como meta identificar contas do *Twitter* que compartilharam artigos na mídia social (ação que produz rastros altmétricos). Assim, primeiramente foi identificado o artigo com maior índice altmétrico, no ano de 2020, na plataforma Altmetric. O artigo selecionado possuía o *altmetric attention score* de 30.776, em 27 de maio de 2022. A busca realizada na plataforma, para o artigo em questão, retornou 72.182 *tweets* originários de 45.097 contas de usuários, as quais totalizaram mais de 18.378.859 seguidores. O título do

artigo é: *Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearer* (DOI: 10.7326/M20-6817) de H Bundgaard e outros.

Devido às limitações de exportação dos *tweets* no ambiente da plataforma Altmetric, as contas foram coletadas através de *web scrapping*. A Altmetric apresenta na web a visualização apenas dos últimos 10.000 *tweets,* dos quais foram coletados 9.824. Desses *tweets* foram coletadas as contas dos usuários, chegando-se a 6.301 contas únicas de usuários da mídia social. Essas contas correspondem ao *dataset* de teste.

3.2 *Dataset* de treinamento

Na aprendizagem de máquina supervisionada o *dataset* de treinamento é um ponto de extrema relevância. Pesquisas com foco na detecção de *bots* (maliciosos ou não) relacionados aos dados altmétricos não são frequentes e não foram identificados *datasets* com dados rotulados (*bot* ou humano), específicos a este contexto. Haustein et al (2016) identificaram manualmente os *bots* em conjunto de dados do *Twitter* e argumentaram que as contas automatizadas no *Twitter* que publicam *tweets* acadêmicos se comportam de maneira diferente das contas gerais de *bots* do *Twitter*. Entretanto não foi possível encontrar o *dataset* utilizado pelos autores.

Dessa forma, utilizou-se, como base para treinamento, o *dataset* disponibilizado por Gutiérrez (2020) disponibilizado no Kaggle[1]. Este conjunto de dados é composto por 37.437 registros de diferentes contas de usuários no *Twitter*. Cada linha contém o ID do usuário e a variável de destino ou resposta. A variável de destino é denotada como *account_type* e tem valores únicos (*bot* ou humano). Este conjunto de dados tem 25.013 contas de usuários anotadas/categorizadas como contas humanas e 12.425 como *bots*.

3.3 Algoritmo de aprendizagem de máquina utilizado

Este trabalho utiliza a técnica de aprendizagem de máquina supervisionada para implementação e teste de um *workflow* visando a detecção de ações de *bots* no *Twitter* em dados altmétricos (artigos compartilhados por contas do *Twitter*). Para isso, utiliza-se o algoritmo denominado K *Nearest Neighbor* (K-NN) ou, em português, K Vizinhos mais próximos, um método baseado em aproximação ou distância de dados a partir de

[1] https://www.kaggle.com/datasets/davidmartngutirrez/twitter-bots-accounts

determinados critérios. O K-NN é um algoritmo considerado não paramétrico, onde a estrutura do modelo é determinada pelo *dataset* utilizado.

Em outras palavras, o K-NN, classifica um determinado objeto (no caso, perfis no *Twitter*) com base em um conjunto de dados de treinamento (neste caso, contas de *Twitter* que foram classificadas previamente como *bot* ou humano). O Algoritmo KNN considera os objetos do conjunto de treinamento mais próximos do ponto de teste, em que K é um parâmetro do algoritmo. Quando o valor de K é maior que 1, para cada ponto de teste, são obtidos K vizinhos.

Cada vizinho é associado a uma classe, sendo possível que diferentes vizinhos sejam agregados de modo a classificar o ponto de teste. Posteriormente, as classes tomam valores em um conjunto discreto e cada vizinho é associado a uma classe. O objeto de teste é classificado na classe mais votada. Dizendo de maneira mais simplista, o algoritmo K-NN encontra padrões de um dado um conjunto de dados (x) que, ao ser executado, treina o sistema tornando-o capaz de classificar o novo conjunto de dados (y), com base nas características e padrões do conjunto de dados (x).

Um dos grandes desafios deste trabalho é identificar os melhores atributos para definir um padrão de comportamento dos *bots.* Outro ponto relevante refere-se à acurácia (taxa de predições corretas) do modelo gerado a partir do *dataset* de treinamento quando aplicado a determinados atributos das contas dos usuários. Assim, foi utilizada aprendizagem de máquina supervisionada para analisar os dados de treinamento e avaliar a acurácia deste modelo quando aplicado ao algoritmo K-NN, tendo como atributos preditivos as características das contas do Twitter: seguidores, amigos e favoritos.

3.4 Análise dos dados com o KNIME

A proposta de um *Workflow* para análise da acurácia do modelo baseado nos dados de treinamento e para classificação de contas do *Twitter* foi realizada por meio da plataforma KNIME. O KNIME é uma plataforma de construção de relatórios, análise e de integração de dados. Ela tem a capacidade de integrar vários componentes para aprendizado de máquina e mineração de dados modularmente. O KNIME possibilita a montagem de nós combinando diferentes fontes de dados, inclusive pré-processamento, modelagem, análise e visualização de dados, com baixa ou nenhuma necessidade de programação. A escolha do KNIME se deu em razão deste conjunto de cartacterísticas, podendo ser relativamente de fácil reprodução

por outros pesquisadores e tomadores de decisão, alem dela ser de acesso livre e de código aberto.

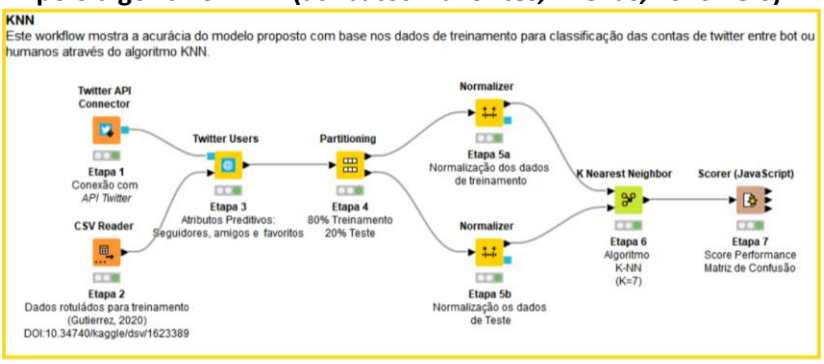
Através desta plataforma foi possível calcular a matriz de confusão. Nesta, foram calculados os indicadores de acurácia total, erro total, coeficiente de Kappa de Cohen, contas corretamente e incorretamente classificadas. Por fim, o KNIME foi utilizado para aplicar o modelo proposto em um *dataset* de teste composto por dados reais de redes sociais acadêmicas/científicas em um *Workflow* para detecção de *bots* e produzir indicadores que indiquem, preliminarmente, a porcentagem de *bots* (contas) e de *tweets* identificados.

4 RESULTADOS E DISCUSSÕES

O *dataset* utilizado para treinamento é composto do ID do usuário e da classificação (bot ou humano). O *dataset* foi disponibilizado em csv e carregado, no KNIME, no nó CSV *Reader*. Para obter os atributos de cada usuário (até então tinha-se apenas o ID do usuário e o rótulo das contas do *Twitter*), considerando as características dos *datasets*, construiu-se uma proposta de *Workflow* para a análise da acurácia do modelo classificador com base nos dados de treinamento (Figura 1). A proposta de *workflow,* para a análise da acurácia dos dados de treinamento, conta com as seguintes etapas:

- Etapa 1 (Nó *CSV Reader*): é feita a leitura do *dataset* de treinamento;
- Etapa 2 (Nó *Twitter API conector*): é estabelecida a conexão com o *Twitter*;
- Etapa 3 (Nó *Twitter User*): são extraídos do *Twitter* os dados dos atributos preditivos selecionados, usando como referência o ID da conta.
- Etapa 4 (Nó *Partitioning*): o *dataset* é dividido, sendo 80% utilizado para treinamento e 20% utilizado para teste;
- Etapa 5 (Nó *Normalizer*): divide-se em 5a e 5b os dados dos dois subconjuntos criados (a e b) na etapa 4 são normalizados para ficarem entre 0 e 1. A normalização de dados é recomendável quando os limites de valores de atributos distintos são muito diferentes, para evitar que um atributo predomine sobre outro;
- Etapa 6 (Nó KNN): é aplicado o Algoritmo KNN onde K = 7 (número de vizinhos mais próximos);
- Etapa 7 (Nó *Scorer*): é criada a Matriz de confusão que aponta a acurácia do modelo.

Figura 1: *Workflow* para análise da acurácia dos dados de treinamento realizada pelo algoritmo K-NN (atributos: *Favorites, Friends, Followers*).



Fonte: Dados da pesquisa

Como resultado deste *workflow* obteve-se a matriz de confusão que, para este dataset de treinamento e estes atributos preditivos, apresentou a acurácia de 66,71% de acerto (Figura 2).

O Problema abordado neste trabalho é considerado binário, uma vez que possui duas classes (*bot* ou humano), com a classe alvo as contas *bots* (ou seja, o valor que se deseja predizer), enquanto contas humanas são as classes negativas. As classes positivas e negativas permitem a definição de medidas específicas relacionadas a cada uma delas tais como Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP), Falso Negativo (FN). Essas medidas são automaticamente calculadas pela matriz de confusão. A Matriz de Confusão é uma forma de apresentar integralmente o desempenho de um algoritmo de classificação binária que relaciona as classes desejadas com as classes preditas. Desta forma tem-se:

- VP: 770 contas da classe positiva (*bots*) foram classificados como positivo (*bots*);
- VN: 4019 contas da classe negativa (humanos) foram classificados como negativo (humanos);
- FP: 794 contas da classe negativa (humanos) foram classificados como positivo (*bots*);
- FN: 1596 contas da classe positiva (*bots*) foram classificados como negativo (humanos).

Em relação à acurácia (*Overal Accuracy*) pode-se dizer que a taxa global de sucesso do algoritmo, que é o número de classificações corretas dividido pelo número total de classificações é de 66,71% e em contrapartida a taxa de erro (*Overal Error*) é de 33,29%. A estatística Kappa (*Cohen's Kappa*) foi de 0,176 demonstrando que não há concordância entre as predições do classificador com a classe correta denotando a necessidade de melhoria no modelo e no *dataset* de treinamento.

Figura 2: Matriz de confusão dos dados de treinamento gerada pelo nó *Score*

	bot (Predicted)	human (Predicted)	
bot (Actual)	770	1596	32.54%
human (Actual)	794	4019	83.50%
	49.23%	71.58%	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
66.71%	33.29%	0.176	4789	2390

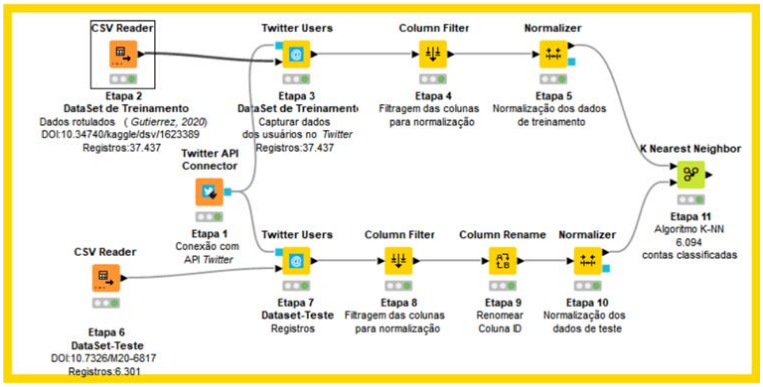
Fonte: Dados da pesquisa

Considerando que a classe alvo deste experimento são as contas *bots*, cuja análise foi realizada com base nos valores apresentados anteriormente, o KNIME apresenta duas medidas associadas ao conceito de relevância: a Precisão (0,492) que mede a qualidade ou exatidão do algoritmo, ou seja, a quantidade de objetos recuperados que são relevantes e o *recall* (0,325), que mede sua completude, ou seja, a quantidade de objetos relevantes que foram recuperados. Em resumo, a precisão é a probabilidade de um item recuperado ser relevante, ao passo que o *recall* é a probabilidade de recuperação de um item relevante.

A medida-F, ou F-Score, usada para avaliar o desempenho de ferramentas de busca e classificação com base na precisão e no *recall*, apresenta um valor baixo (0,39) em relação ao valor obtido por Aaljohani, Fayoumi e Hassan (2020) que obteve um F1-score de 0,67 o que sugere a possibilidade de que a técnica de rede convolucional de seja mais promissora que o K-NN.

Uma vez estabelecidos os atributos que serão utilizados no algoritmo de aprendizagem de máquina supervisionado e a acurácia do método é, então, realizada a etapa de modelagem. A proposta de *Workflow* para a classificação de contas do *Twitter* utilizando o KNIME é apresanda na Figura 3. Neste *workflow* os dados rotulados (*dataset* de treinamento) são utilizados para treinamento do algoritmo K-NN e os dados dos *tweets* coletados na plataforma Altmetric são os objetos classificados (*dataset* de teste). Neste fluxo, além da recuperação das informações dos 37.437 usuários dos dados de treinamentos também foram coletadas as informações de 6.094 das 6.301 contas de usuários contidos no *dataset* de teste.

Figura 3: *Workflow* para classificação de contas do *Twitter* – Maquinade Aprendizagem Supervisionada K-NN (Atributos Preditivos: *Favorites, Friends, Followers*).



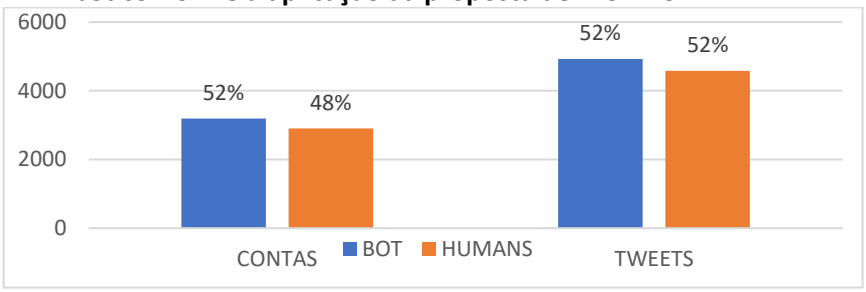
Fonte: Dados da pesquisa

A proposta de *workflow* para a classificação das contas do *Twitter* para dados altmétricos conta com 11 etapas descritas a seguir:

- Etapa 1 (Nó *CSV Reader*): é feita a leitura do *dataset* de treinamento composto pelo Id da conta e o rótulo (*bot*/humano);
- Etapa 2 (Nó *Twitter API conector*): é estabelecida a conexão com o *Twitter*;
- Etapa 3 (Nó *Twitter User*): são extraídos do *Twitter* os dados dos atributos preditivos selecionados, usando como referência o ID da conta.
- Etapa 4 (Nó *Column Filter*): são selecionadas as colunas (ID, account_type, seguidores, amigos, favoritos) que serão utilizadas;
- Etapa 5 (Nó *Normalizer*): os dados das colunas selecionadas na etapa 4 são normalizados para ficarem entre 0 e 1;
- Etapa 6 (Nó *CSV Reader*): é feita a leitura do *dataset* de teste composto pelo login da conta;
- Etapa 7 (Nó *Twitter User*): recupera os valores dos atributos preditivos das contas no *Twitter* com base no login;
- Etapa 8 (Nó *Column Filter*): são filtradas apenas as colunas relacionadas aos atributos preditivos e o id;
- Etapa 9 (Nó *Column Rename*): altera o nome da coluna *User_id* para Id ficando compatível com o *dataset* de treinamento.
- Etapa 10 (Nó *Normalizer*): os dados das colunas selecionadas na etapa 8 são normalizados para ficarem entre 0 e 1;
- Etapa 11 (Nó KNN): é aplicado o Algoritmo KNN onde K = 7.

Como resultado o *workflow* proposto classificou 6.094 das 6.301 contas de usuários do *Twitter* contidas no *dataset* de treinamento (207 contas retornadas do altmetrics, não foram encontradas pelo nó *Twitter* Users do KNIME (Etapa 7). Das 6.094 contas classificadas 3.194(52%) foram classificadas como *bots* e 2.900 (48%) como humanos. As contas *bots* geraram 4,925 (52%) postagens de acordo com o *dataset* analisado (Gráfico 1).

Gráfico 1: Resultado, em porcentagem, da classificação de contas do Twitter em humano e bot conforme a aplicação da proposta de Workflow



Fonte: Dados da pesquisa

5 CONSIDERAÇÕES FINAIS

Este trabalho apresenta algumas evidências em relação a importância de estudar os efeitos colaterais dos *bots* nos estudos altmétricos. Muito tem sido estudado sobre a detecção de *bots* em mídias sociais, mas, poucos são os estudos voltados especificamente para a interferência desses agentes em dados altmétricos. Foi possível explorar como essas contas do Twitter afetam negativamente os processos de mensuração relacionados às métricas alternativas, bem como as suas potencialidades e limitações para avaliação de produtos acadêmicos.

Ao explorar formas de detectar a atuação dos *bots* em mídias sociais, foi possível, com a presente pesquisa, apresentar uma proposta preliminar para detecção de *bots* aplicados a dados altmétricos, utilizando algoritmos de aprendizagem de máquina supervisionada e a ferramenta KNIME. Os workflows apresentados são de relativamente fácil elaboração, podendo ser utilizados em outras pesquisas sem a necessidade de conhecimentos avançados em tecnologias de informação, favorecendo a aplicação em outros tipos e abordagens de pesquisa.

Acredita-se que o desenvolvimento de tecnologias de identificação de *bots* (maliciosos ou não) poderão trazer outras abordagens de pesquisa e a possibilidade de utilização efetiva dos indicadores altmétricos para monitoramento e avaliação científica. Os *bots,* no caso dos indicadores altmétricos, trazem uma visão distorcida e incerta da realidade. A sua segmentação de maneira automatizada, com nível de acurácia alto, possibilitará a realização de uma gama de estudos altmétricos com vieses reduzidos. Apenas para exemplificar, é muito provável que as correlações entre o *Twitter* e citações apresentadas por algumas pesquisas, seriam diferentes caso houvesse a eliminação dos *bots* das amostras.

Para pesquisas futuras, observa-se a necessidade de se estudar as características dos *bots* das contas relacionadas a dados altmétricos. As contas que tuitam sobre ciência podem ter características diferentes daquelas que não tuitam temas acadêmicos, sendo relevante análises qualitativas que as classifiquem como *bot* de humano (não apenas no *Twitter*, mais em outras mídias sociais). Com isso, haverá a possibilidade de se produzir um *dataset* de treinamento capaz de classificar com mais acurácia o *dataset* de teste. É preciso, ainda, desenvolver pesquisas com vistas a identificar padrões que forneçam maior acurácia dos dados, incluindo a testagem e comparação de algoritmos e de outras abordagens de aprendizagem de máquina (supervisionadas, não supervisionadas e semissupervisionadas). Vislumbra-se, por fim, a necessidade de estudos que avaliem a potencialidade do KNIME (e outras ferramentas), para análise de um conjunto mais amplo de dados, de diferentes mídias sociais, com diferentes formatos de entrada de dados, de modo a possibilitar a realização de estudos e aplicações efetivas da altmetria para a avaliação acadêmica e científica.

REFERÊNCIAS

ALJOHANI, Naif Radi; FAYOUMI, Ayman; HASSAN, Saeed Ul. Bot prediction on social networks of *Twitter* in altmetrics using deep graph convolutional networks. **Soft Computing**, *[S. l.]*, v. 24, n. 15, p. 11109–11120, 2020. DOI: 10.1007/s00500-020-04689-y. Disponível em: https://doi.org/10.1007/s00500-020-04689-y. Acesso em: 10 jun. 2022.

CHAPMAN, Colina et al. Exploiting Social Networks of *Twitter* in Altmetrics Big Data. **Scientometrics**, 39, n. 1, p. 175, 2018. Disponível em: https://doi.org 10.1016/j.ipm.2022.102945. Acesso em: 27 jan. 2022.

CROTTY, David. Altmetrics: Finding meaningful needles in the data haystack. **Serials review**, Londres, v. 40, n. 3, p. 141-146, 2014. DOI: 10.1080/00987913.2014.947839. Disponível em: https://doi.org/10.1080/00987913.2014.947839 . Acesso em: 10 jun. 2022.

DARLING, Emily; SHIFFMAN, David; CÔTÉ, Isabelle; DREW, Joshua. The role of *Twitter* in the life cycle of a scientific publication. **Ideas in Ecology and Evolution**, Ontario, v. 6, p. 32–43, 2013. Disponível em: https://doi.org/10.4033/iee.2013.6.6.f. Acesso em: 10 jun. 2022.

DIDEGAH, Fereshteh; MEJLGAARD, Niels; SØRENSEN, Mads P. Investigating the quality of interactions and public engagement around scientific papers on *Twitter*. **Journal of Informetrics**, Amsterdã, v. 12, n. 3, p. 960–971, 2018. Disponível em: https://doi.org/10.1016/j.joi.2018.08.002. Acesso em: 10 jun. 2022.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon Ferreira de. **Inteligência artificial:** uma abordagem de aprendizado de máquina. Rio de Janeiro, LTC , 2011.

FERRARA, Emilio; VAROL, Onur; DAVIS, Clayton; MENCZER, Filippo;; FLAMMINI, Alessandro. The rise of social *bots*. **Communications of the ACM**, Nova Iorque, v. 59, n. 7, p. 96–104, 2016. DOI: 10.1145/2818717. Disponível em: https://dl.acm.org/doi/10.1145/2818717. Acesso em: 10 jun. 2022.

GUTIÉRREZ, David Martín. ***Twitter Bots* Accounts [Data set]**. 2020. Disponível em: https://doi.org/10.34740/KAGGLE/DSV/1623389. Acesso em: 16 abr. 2022.

HAUSTEIN, Stefanie; BOWMAN, Timothy D.; HOLMBERG, Kim; TSOU, Andrew; SUGIMOTO, Cassidy R.; LARIVIÈRE, Vincent. Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. **Journal of the Association for Information Science and Technology**, Hoboken, Nova Jersey, v. 67, n. 1, p. 232–238, 2016. Disponível em: https://doi.org/10.1002/asi.23456. Acesso em: 10 jun. 2022.

MARTINS, Agnaldo Lopes. Potenciais aplicações da Inteligência Artificial na Ciência da Informação. **Informação & Informação**, *[S. l.]*, v. 15, n. 1, p. 1–16, 2010. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/3882. Acesso em: 10 jun. 2022.

MORAIS, Daniel Marques Gomes; DIGIAMPIETRI, Luciano Antonio. Methods and Challenges in Social *Bots* Detection: A Systematic Review. **ACM International Conference Proceeding Series**, Nova Iorque, p. 21–28, 2021. Disponível em: https://doi.org/10.1145/3466933.3466973. Acesso em: 10 jun. 2022.

NEVES, Bárbara Coelho Neves. Inteligência artificial e computação cognitiva em unidades de informação: conceitos e experiências. **Logeion:** Filosofia da Informação, Rio de Janeiro, v. 7, n. 1, p. 186–205, 2020. Disponível em: http://revista.ibict.br/fiinf/article/view/5260/5012. Acesso em: 10 jun. 2022.

NUNES, Amanda Maria de Almeida. **MÁQUINAS SOCIAIS E A DESINFORMAÇÃO EM REDE:** o papel das entidades de software na formação de opinião na Internet. 2020. Universidade Federal de Pernambuco, Recife, 2020. Disponível em: https://repositorio.ufpe.br/handle/123456789/38549. Acesso em: 27 abr. 2022.

ORABI, Mariam; MOUHEB, Djedjiga; AL AGHBARI, Zaher; KAMEL, Ibrahim. Detection of *Bots* in Social Media: A Systematic Review. **Information Processing and Management**, Amsterdã, v. 57, n. 4, 2020. Disponível em: https://doi.org/10.1016/j.ipm.2020.102250. Acesso em: 10 jun. 2022.

PRIEM, Jason; HEMMINGER, Bradely H. Scientometrics 2.0: New metrics of scholarly impact on the social Web. **First Monday**, Chicago, 2010. Disponível em: https://doi.org/10.5210/fm.v15i7.2874. Acesso em: 10 jun. 2022.