

# Detection of Big Five personality traits in Twitter user's profiles based on textual posts

Laura Damaceno de Almeida<sup>1</sup>, Denise Goya<sup>1</sup>

<sup>1</sup>Centro de Matemática, Computação e Cognição, Universidade Federal do ABC  
Av.dos Estados, 5001 - Santo André, SP, Brasil

l.damaceno@ufabc.edu.br, denise.goya@ufabc.edu.br

**Abstract.** *Personality describes people's behavior and can influence their choices and decision-making. One of the most consolidated metrics for personality traits is Big five. However, few works were produced using this technique and portuguese texts shared on Twitter. The objective of the present work is to construct a dataset with tweets in portuguese labeled with the dominant personality trait and and evaluate its potential use in classic machine learning models. In the experiment carried out, it was found that with the inclusion of the SMOTE technique, the performance of machine learning algorithms improved, and the best result was with Logistic Regression with TF-IDF unigram.*

**Resumo.** *Personalidade descreve o comportamento das pessoas e pode influenciar suas escolhas e tomadas de decisão. Uma das métricas mais consolidadas para traços de personalidade é o Big Five. Poucos trabalhos foram produzidos para detecção dos mesmos a partir de textos em português compartilhados em redes sociais. O objetivo do presente trabalho consiste na construção de um conjunto de dados com tweets em português rotulados com o traço de personalidade dominante e verificar o seu potencial de uso em modelos de aprendizado de máquina clássicos. No experimento realizado, os algoritmos de aprendizado de máquina apresentaram desempenho superior com a inclusão da técnica SMOTE e o melhor resultado foi Regressão Logística com TF-IDF unigram.*

## 1. Introdução

A personalidade é um aspecto fundamental da vida humana. Ela descreve o padrão de pensamento, sentimento e características que predizem e descrevem o comportamento de um indivíduo e também influencia as atividades da vida diária, incluindo emoções e preferências [Majumder et al. 2017]. A previsão da personalidade é uma área de estudo em que uma pessoa é categorizada em uma classe, de acordo com traços de personalidade [Kunte and Panicker 2019].

As redes sociais ganharam grande popularidade entre indivíduos, porque com ela podem compartilhar seus pensamentos, sentimentos e atividades. Elas agregam uma grande quantidade de informações sobre seus usuários, que podem ser usadas para prever traços de personalidade, faixa etária, primeiras impressões e interesses dos mesmos [Salem et al. 2019]. Além disso, as redes sociais desempenham um papel vital na comunicação humana e melhoram os aplicativos de negócios [Flack and D'Souza 2014]. Uma das redes com significativo volume de usuários é o Twitter, que permite o compartilhamento de opiniões e pensamentos através de *tweets* [Genina et al. 2020]. Em janeiro

de 2022, o Brasil superou 19,05 milhões de perfis nesta rede e tornou-se o quarto país com mais usuários [Statista 2022].

Tradicionalmente, as únicas maneiras de identificar personalidades eram por meio de testes de personalidade baseados em questionários, ou através de conversas com psicólogos. Esses métodos, no entanto, são caros, demorados, propensos a erros e trabalhosos [Kaushal and Patwardhan 2018]. Com o avanço de técnicas de aprendizado de máquina foram abertas oportunidades para a automatização desse processo. Resultados de predição de personalidade podem ajudar a melhorar sistemas de recomendação, como no estudo de [Khan et al. 2020] que buscou recomendar o melhor filme para os usuários com base na personalidade e nas últimas avaliações. Também demonstraram potencial para detecção de *cyberbullying* [Balakrishnan et al. 2019] e identificação de tendências à disseminação de *fake news* e rumores nas redes sociais [Chen and Wang 2020, Sangoju et al. 2022].

Existem diversas métricas para determinar a personalidade de uma pessoa, como MBTI, DISC e *Big Five* (também conhecido como Modelo de cinco fatores ou FFM) [Digman 1990]. Esse último é o mais utilizado na detecção de personalidade no Twitter e será adotado neste estudo. O *Big Five* categoriza as personalidades das pessoas em: Abertura (Ab), Conscienciosidade (C), Extroversão (E), Afabilidade (A) e Neuroticismo (N). Ainda que o modelo *Big Five* seja muito amplo para explicar todas as possíveis diferenças de personalidade, foi suficientemente estudado e testado em um grande número de culturas e é considerado confiável para medir traços de personalidade [Ferreira 2018].

O objetivo do presente trabalho consiste na construção de um conjunto de dados com *tweets* em português rotulados com o traço de personalidade *Big Five* dominante e verificar o seu potencial de uso em modelos de aprendizado de máquina clássicos.

Este artigo tem a seguinte organização: a seção 2 aborda os trabalhos relacionados; na seção 3 é apresentada a metodologia usada para o desenvolvimento da pesquisa. Na seção 4 é realizada uma análise nos resultados experimentais encontrados. Na seção 5 tem a conclusão com considerações finais.

## 2. Trabalhos relacionados

Com a finalidade de visualizar o estado da arte, identificar os trabalhos e técnicas mais atualizadas na detecção de traços de personalidade usando o Twitter e aprendizado de máquina, foi conduzido um mapeamento sistemático sobre artigos dos últimos 4 anos (de 2019 a 2022) nas bases IEEE Xplore, ACM DL, *Web of Science* e *Springer*. A string de busca usada foi “*personality AND (detection OR detecting OR prediction OR predicting OR classification OR classifying) AND twitter*”.

Para fazer a detecção de traços de personalidade é necessário ter uma base de dados, e muitas das pesquisas encontradas consideraram construí-la a partir de questionários (como *Big Five Inventory* ou BFI) compartilhados no Twitter para voluntários responderem, como nos estudos de [Pratama and Maharani 2021]. [Lucky et al. 2021] e [KN and GavriloVA 2021] utilizaram conjuntos de dados já prontos de estudos anteriores. Ademais [Christian et al. 2021] utilizaram a rotulação manual realizada por especialistas. Entretanto, não foi encontrado conjunto de dados com *tweets* em português com o respectivo traço de personalidade, sendo necessário realizar neste estudo o compartilhamento do questionário BFI em busca de voluntários.

Nos trabalhos de [Cahyani and Faishal 2020], [Moreno et al. 2019], [Tutaysalgir et al. 2019], [Kunte and Panicker 2019] e [Usher and Dondio 2020] foram coletados apenas dados textuais contidos no *tweets* dos usuários e técnicas de transformação de processamento de linguagem natural como TF-IDF, ademais lematização e *stemming* foram aplicados antes dos algoritmos de classificação. No estudo de [Ergu et al. 2019] foi analisada a performance dos modelos considerando os últimos 50 e 25 *tweets* postados e depois 50 e 25 *tweets* aleatórios dos perfis de cada usuário, o melhor resultado foi com a coleta dos últimos 50 *tweets*.

Dentre os algoritmos e técnicas empregados pelos trabalhos analisados, os mais usados e com potencial de bons resultados foram *Naive Bayes* (NB), SVM, Árvore de decisão (AD), *Random Forest* (RF), KNN e Regressão logística (RL).

Considerando o *Big Five*, poucas pesquisas documentaram quais técnicas de seleção de variáveis foram usadas. Sem elas, o desempenho da predição pode ser prejudicado. Cita-se como exceção [Ergu et al. 2019] que empregou correlação de *Pearson* e *Chi-square* e, ao final, o modelo SVM alcançou um dos maiores valores de acurácia dentre todos os artigos analisados neste mapeamento, com um valor de 0.978.

Um dos problemas enfrentados nas pesquisas foi o desbalanceamento entre as classes de traços de personalidade que pode tornar os modelos de classificação tendenciosos. Em [Indira and Maharani 2021], por exemplo, havia mais pessoas abertas do que extrovertidas. A fim de contornar esse problema, o estudo de [Lucky et al. 2021] aplicou MLSMOTE (sobreamostragem de minoria sintética *multi-label*), com melhoria do desempenho dos classificadores testados em mais de 19% em acurácia e *F1 Score*.

Apesar dos estudos citados, foi identificado que pesquisas voltadas para a predição de traços de personalidade em *tweets* em português são bem escassas e a falta de uma base pública com essas informações dificulta a evolução e o andamento de estudos nessa área. No trabalho de [Marques Junior 2018] é considerado apenas o traço de personalidade narcisista. [Mota et al. 2022] analisaram a identificação de traços de personalidade no Twitter em ambientes de Participação Eletrônica. Ademais, [Buiar et al. 2019] consideraram apenas dados textuais coletados da plataforma educacional Moodle e [Camati 2021] usou a base de dados facebook-BR, que é composta por textos compartilhados no Facebook.

### 3. Metodologia

Para chegar no objetivo proposto da pesquisa, a metodologia é composta por quatro etapas (Figura 1). Os principais passos, incluindo a descrição da coleta de dados, a estratégia adotada para calcular os traços de personalidade, o processo de pré-processamento de texto, treino e teste dos algoritmos de Aprendizado de Máquina são detalhados nas próximas subseções.

#### 3.1. Coleta de dados e cálculo do traço de personalidade

A coleta de dados se iniciou a partir do compartilhamento do questionário *Big Five* nas redes sociais, que foi traduzida e adaptada para português por [Andrade 2008], cujo nome é IGFP-5 e sua validade foi testada por [Ferreira 2018], [Smiderle et al. 2020] e [da Silva et al. 2020].

O questionário é constituído por 44 perguntas, além disso outras foram incluídas para coletar informações demográficas referentes ao usuário como: nome do perfil no

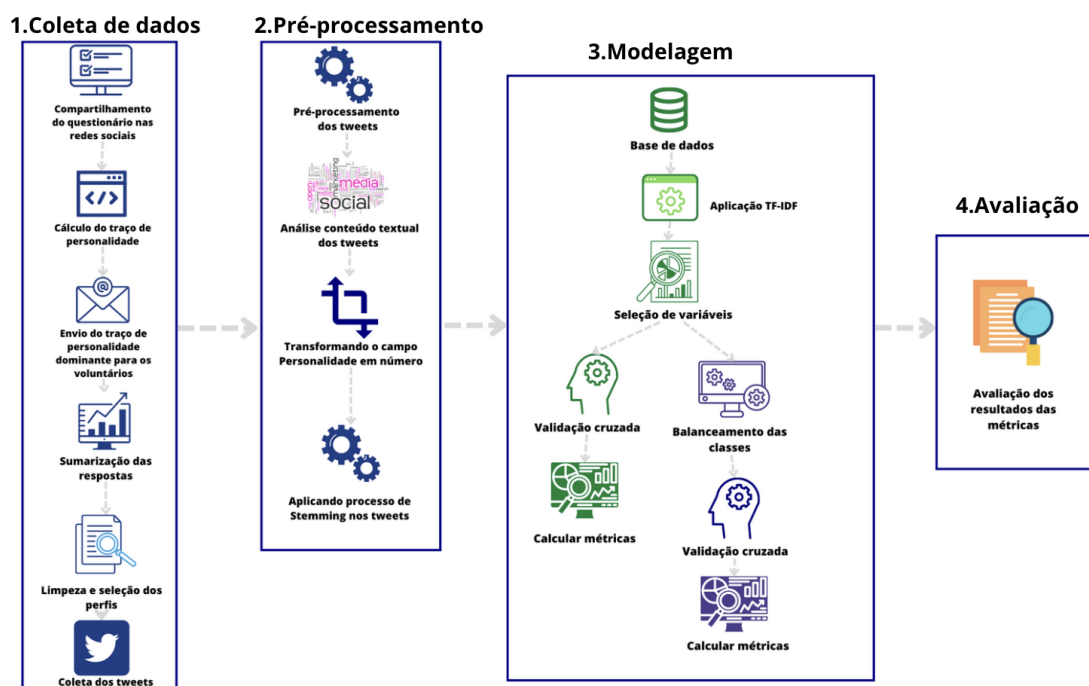


Figura 1. Etapas utilizadas na pesquisa

Twitter, gênero e idade da pessoa, para assim ser realizada uma análise prévia com a finalidade de compreender o perfil das amostras, verificar se há desbalanceamento entre as classes de traço de personalidade. Após isso, a etapa limpeza e seleção dos perfis foi aplicada (subseção 4.1) e as últimas 60 interações dos usuários no Twitter (seja *tweets*, respostas e *retweets*) foram coletadas de cada perfil, usando a API do Twitter, pois conforme conclusão de [Ergu et al. 2019] quando o número de *tweets* e usuários aumentam, o desempenho dos modelos aumentam na mesma proporção.

Para cada resposta informada pelo voluntário, foi atribuída uma pontuação. O cálculo do escore final foi realizado, conforme o estudo de [Ferreira 2018]. As respostas do questionário são dadas em escala Likert de 5 pontos a fim de calcular o escore em cada traço. Cada pergunta poderia ser respondida com as seguintes opções: Concordo totalmente, concordo parcialmente, neutro, discordo parcialmente e discordo totalmente.

Seguindo o raciocínio do estudo de [Ferreira 2018], se a relação entre pergunta e traço é positiva (polo +), a pontuação atribuída será 4 para concordo totalmente, 3 para concordo parcialmente, 2 para neutro, 1 para discordo parcialmente e 0 para discordo totalmente, conforme [Ferreira 2018]. Por outro lado, se a relação é negativa (polo -), concordar com a afirmação significa diminuir o valor do escore naquele traço. Assim, a pontuação atribuída é o inverso do polo positivo, ou seja, 0 para concordo totalmente, 1 para concordo parcialmente e assim por diante. Sendo assim o escore em cada traço de personalidade é dado por:  $score = (\sum p_i)/n$ , em que  $p_i$  é a pontuação da resposta na pergunta  $i$  e  $n$  é a quantidade de perguntas respectivas daquele traço. Essa etapa foi finalizada com o encaminhamento de um email individual para todos os voluntários, com o traço de personalidade dominante seguido de uma explicação, extraída dos artigos.

### 3.2. Pré-processamento dos dados textuais

Na etapa de pré-processamento foram utilizadas técnicas de processamento de linguagem natural (NLP), cujas postagens textuais foram transformadas em letras minúsculas, então acentos, caracteres especiais (como *hashtags*, *urls*, menções de outros usuários) e *stop words* (como preposições e artigos) foram removidos para processamento dos modelos. Além disso, as postagens que estavam em outra língua, como o inglês ou espanhol, foram traduzidas manualmente para o português, visto que poderia afetar o desempenho dos modelos de Aprendizado de Máquina para a língua portuguesa.

Uma vez que os *tweets* originais e respostas são dados textuais é necessário convertê-los para números, as técnicas TF-IDF e *stemming* foram utilizadas neste processo. O parâmetro “*n-gram*” foi ajustado no TF-IDF (*Term Frequency–Inverse Document Frequency*) para avaliar o desempenho dos modelos de classificação considerando *unigram* e *bigram*.

### 3.3. Seleção de variáveis

Após a transformação dos dados textuais uma grande quantidade de colunas foi criada para cada palavra ou conjunto de palavras encontradas nos textos. Se for fornecido como entrada todas essas colunas para os algoritmos de Aprendizado de Máquina podem afetar negativamente o desempenho, visto que teriam muitas palavras ruídos que não ajudariam os algoritmos a diferenciar uma personalidade da outra. A redução do número de atributos ou colunas pode melhorar o desempenho do modelo, reduzir o custo computacional e tornar os resultados obtidos mais compreensíveis [Faceli et al. 2011].

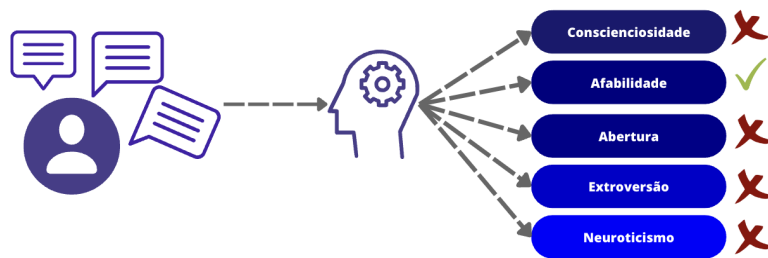
Considerando que as variáveis de entrada têm um formato numérico (peso que o TF-IDF atribuiu para cada uma) e a variável Personalidade é categórica (de 1 a 5), o método estatístico ANOVA foi considerado nesse estudo. Este método de filtro univariado usa a variância para descobrir a separabilidade dos recursos individuais entre as classes [Effrosynidis and Arampatzis 2021]. E com isso consegue-se identificar a correlação entre as variáveis e os traços de personalidade.

### 3.4. Aprendizado de máquina

O Aprendizado de Máquina (AM) utiliza grandes conjuntos de dados e identifica os padrões entre as variáveis de entrada e saída. Essas técnicas podem descobrir associações anteriormente desconhecidas, gerando novas hipóteses e direcionando pesquisadores e recursos para melhores tomadas de decisão [Noorbakhsh-Sabet et al. 2019].

O problema a ser resolvido nesta pesquisa trata-se de uma classificação, ou seja, com base nos últimos textos do usuário será predito seu traço de personalidade dominante (Figura 2). Portanto há mais de duas saídas possíveis, sendo conhecido na literatura como um problema multiclasse, porque há 5 traços de personalidade possíveis para o modelo de classificação selecionar.

Durante o mapeamento sistemático foi identificado que os algoritmos de AM mais utilizados para classificação de traço de personalidade foram: *Naive Bayes* (NB), *Support Vector Machine* (SVM), Árvore de decisão (AD), *K-Nearest Neighbors* (KNN), Regressão Logística (RL) e *Random Forest* (RF), como nos trabalhos de [Cahyani and Faishal 2020] [Ergu et al. 2019] [Moreno et al. 2019] e [Kunte and Panicker 2019]. Ademais, na



**Figura 2. Modelo multiclasse para detecção do traço de personalidade dominante**

documentação da biblioteca de AM do python, Scikit-Learn <sup>1</sup>, é informado que no mínimo 50 amostras são necessárias para utilizar esses algoritmos, portanto essa foi a meta durante o compartilhamento do questionário na coleta de dados.

Os modelos foram treinados e testados utilizando a técnica de validação cruzada (*cross-validation*) *k-fold* com *k* igual 10, pois essa técnica fornece uma estimativa melhor da taxa de erro [King et al. 2021]. E foram calculadas as métricas *F1-score* que é a medida harmônica entre precisão e *recall*, MCC (Coeficiente de Correlação de Matthews) para medir a qualidade das classificações multiclasse e sendo um bom indicador de modelos de previsão desbalanceados [Grandini et al. 2020], ademais o AUC (*Area Under Curve*) foi calculado para medir o desempenho de classificação. De acordo com o estudo de [Wardhani et al. 2019], AUC é a métrica que apresenta maior robustez quando os dados estão desbalanceados, portanto na hora de selecionar o melhor algoritmo um peso maior será atribuído para essa métrica.

Ao final desse processo, para identificar a melhor abordagem de classificação do traço de personalidade dominante em perfis de usuários do Twitter com base nas últimas 60 postagens textuais em português, este presente estudo avaliou os algoritmos de AM em dois cenários diferentes, onde no cenário 1 os dados foram fornecidos sem nenhum balanceamento, e no cenário 2 os dados foram balanceados utilizando a técnica SMOTE (*Synthetic Minority Over-sampling Technique*).

No SMOTE foi utilizado o método *over-sampling* para criar novas amostras das classes minoritárias [Tan et al. 2022] até que fiquem com a mesma quantidade que a majoritária. Esse método em conjunto com a detecção de traços de personalidade já foi adotado em trabalhos que obtiveram resultados positivos [Ryan et al. 2023, Lucky et al. 2021]. Apesar de [Lucky et al. 2021] utilizar o MLSMOTE (*Multilabel Synthetic Minority Over-sampling Technique*), neste estudo será considerado o SMOTE visto que o problema tratado é multiclasse (Figura 2), o MLSMOTE é usado apenas para problemas *multi-label*, ou seja, informar dois traços de personalidade daquele usuário com base no conteúdo textual.

#### 4. Resultados e análises

Nesta seção são mostradas as análises e resultados obtidos durante as etapas descritas na seção da Metodologia.

<sup>1</sup><https://scikit-learn.org>

#### 4.1. Perfil dos usuários selecionados e seus traços dominantes

Ao todo, 77 voluntários preencheram o questionário e concordaram com o Termo de Consentimento Livre e Esclarecido (TCLE), onde na Figura 3 pode-se perceber que grande parte deles têm entre 20 e 30 anos, aproximadamente 83% trabalham na área de tecnologia e 75% dos voluntários são do gênero feminino.

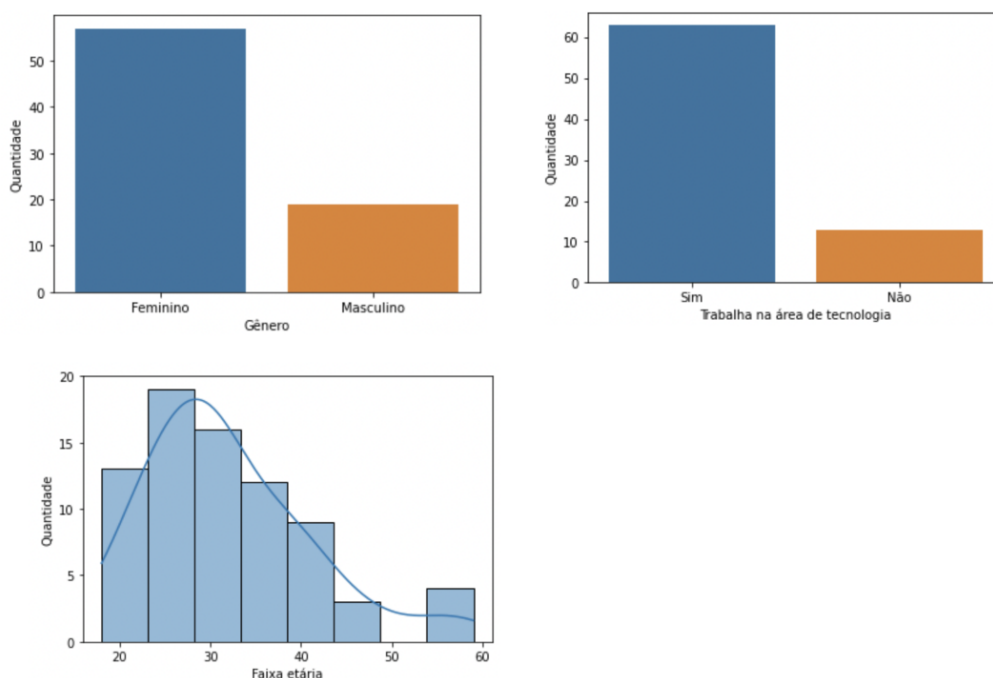


Figura 3. Perfil dos voluntários

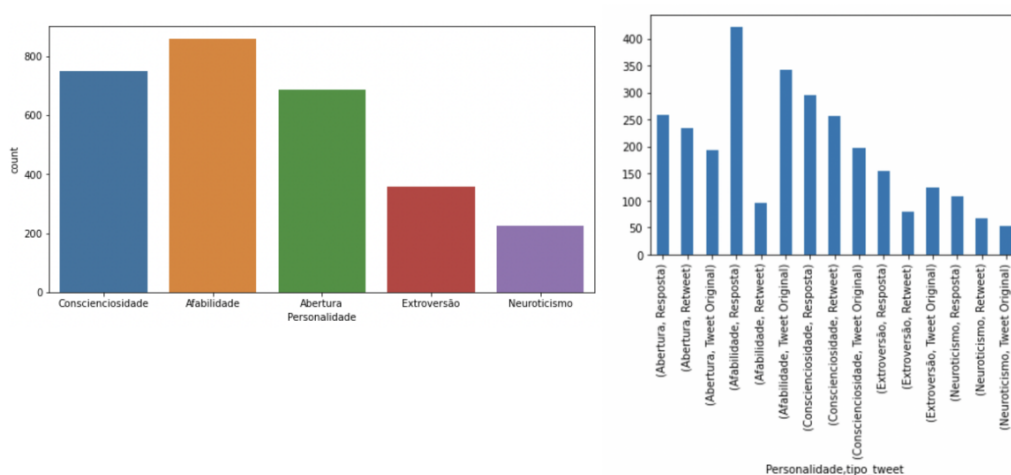
Aproximadamente 30% dos voluntários tiveram como traço dominante a Afabilidade, 26% dos voluntários Conscienciosidade, 26% Abertura, 11% Extroversão e o traço com menor representatividade foi o Neuroticismo com 7% dos voluntários. Durante a etapa de extração de dados do Twitter, foi identificado que nem todos preencheram o campo do nome do usuário no Twitter corretamente, muitos colocaram urls inválidas ou seus usuários de outras redes sociais como o LinkedIn, reduzindo a quantidade válida de usuários avaliados para 65. Portanto foi adotada a estratégia de entrar em contato por email com os mesmos solicitando uma *url* válida ou o nome do usuário do Twitter, entretanto apenas 1 voluntário respondeu informando que não tem conta no Twitter.

Após esse processo foi utilizado um código em python em conjunto com a biblioteca Tweepy <sup>2</sup> para conectar com API do Twitter <sup>3</sup> e coletar as últimas 60 interações textuais dos usuários, entretanto foi encontrado que 10 voluntários tinham suas contas privadas ou fechadas, cuja API retornava o erro “Perfil não encontrado”. Então foi encaminhado um email para esses usuários solicitando que abrissem suas contas por um tempo até a coleta de dados e foi reforçado que tanto os nomes deles quanto seus usuários seriam anonimizados.

<sup>2</sup><https://www.tweepy.org>

<sup>3</sup><https://developer.twitter.com/en/products/twitter-api>

A base final ficou composta por 54 usuários ativos no Twitter, com suas últimas 60 interações: *tweets* originais, respostas e *replies*. Todavia foi considerado apenas textos de *tweets* originais e respostas, porque foi levantada a premissa de que essas formas de interação os usuários têm mais oportunidade de demonstrar sua personalidade, visto que são textos formulados originalmente pelos mesmos.



**Figura 4. Quantidade de *tweets* coletados por traço dominante *Big Five***

Conforme mostrado na Figura 4, a base de dados textual é composta por mais textos da Afabilidade (correspondendo a 28.45% da base), seguido pela Conscienciosidade compondo 26.21%, Abertura com 24.99% e Neuroticismo representa a menor parcela da base de dados, compondo 7.89% dos textos coletados, portanto com esses valores pode-se perceber que a base em questão contém desbalanceamento entre a quantidade de traços de personalidade, assim como nos trabalhos [Indira and Maharani 2021] e [Lucky et al. 2021].

Na Tabela 1 pode-se observar que Neuroticismo é o traço que apresentou maior média em relação à quantidade de menções em seus textos (representado pelo uso do caractere @), com valor de 1.06. A Extroversão (compondo 12.46% da base textual) foi o traço que teve a maior média em relação a quantidade de *hashtags* utilizadas nos textos, com valor de 0.35 (Tabela 1).

**Tabela 1. Quantidade de menções e *hashtags* usadas por traço de personalidade**

Personalidade	Menções				Hashtags			
	Média	25%	75%	Máximo	Média	25%	75%	Máximo
Conscienciosidade	0.84	0	1	8	0.09	0	0	3
Afabilidade	0.77	0	1	5	0.03	0	0	3
Abertura	0.81	0	1	7	0.10	0	0	5
Extroversão	0.75	0	1	10	0.35	0	0	10
Neuroticismo	1.06	0	1	7	0.08	0	0	1

#### 4.2. Dados textuais e traço de personalidade dominante

Após a etapa de pré-processamento foi identificado que apesar do Neuroticismo compor a menor parcela da base final, é o traço de personalidade que apresenta uma quantidade



maior de palavras utilizadas nas respostas (Tabela 2), mais de 75% das respostas têm mais de 20 palavras, indicando que eles utilizam muitas palavras para interagir com outros usuários do Twitter. Esse comportamento pode ser explicado por conta do ambiente que as redes sociais fornecem, permitindo aos usuários maior controle sobre como desejam se apresentar do que em relação aos contextos *offline*, outra tendência proeminente é a aparente motivação daqueles com neuroticismo de alto traço para usar a mídia social para interagir socialmente [Bowden-Green et al. 2021].

**Tabela 2. Quantidade de palavras por tipo de tweet e personalidade**

Personalidade	Tweet Original				Resposta			
	Média	25%	75%	Máximo	Média	25%	75%	Máximo
Conscienciosidade	28	13	7	20	26	10	5	15
Afabilidade	32	13	8	18	26	9	3	14
Abertura	42	11	4	17	53	10	4	9
Extroversão	33	14	7	20	26	10	5	17
Neuroticismo	25	12	6	20	25	11	4	20

Quando analisados os conteúdos textuais de cada um dos traços de personalidade pode-se perceber que pessoas cujo traço dominante é Extroversão há muita variação de palavras e vocabulários, porque a *wordcloud* está mais variada Figura 5 d. Ademais, demonstra que essas pessoas se comunicam bastante, são bem sociáveis, movimentam debates, disseminam rápido às informações que encontram, representado pelas palavras “acabou publicar” e “ainda não” mostrando que gostam de interagir.



**Figura 5. Wordcloud das palavras mais frequentes por traço de personalidade**

Pessoas com Conscienciosidade sendo seu traço dominante tendem a ser bem organizadas e consistentes, gostando de realizar planejamento [Usher and Dondio 2020] e pode-se identificar esses traços na respectiva *wordcloud* (Figura 5 a) que mostra palavras que demonstram suas preocupações com prazos e data comemorativas (como “feliz aniversário”, “dia hoje”), também estando condizente com [Bunker and Kwan 2021] foi encontrado muitos assuntos voltados para pedir ajudar (como “preciso ajuda”, “indicação

cursos”), indicando que esses indivíduos procuram plataformas digitais, como as redes sociais, para encontrar experiências ou oportunidades que não possuem *offline*.

Afabilidade representa pessoas que gostam de manter boas relações com outras pessoas, são bem amigáveis [Usher and Dondio 2020] e pode-se encontrar essas características na Figura 5 b, cujas palavras mais utilizadas são as que expressam cordialidade, além de indicar uma tendência a entrar em contato com pessoas de opiniões opostas (como: “não sei”, “alguém consegue” e “quero saber”).

Em relação a Abertura pode-se observar que há muitas palavras voltadas a disseminação de informações e socialização, como: “vale pena”, “vai ter”, “juntem mim”, “você deveria”. Além disso pode-se notar características da busca de informações, representado pela palavra “não sei”(Figura 5 c). Foi identificado o interesse desse público em assuntos técnicos da área de tecnologia (representado pela palavra “inteligência artificial”), assim como em outros traços de personalidade como: Extroversão, Afabilidade e Conscienciosidade.

Neuroticismo refere-se a pessoas emocionalmente instáveis que são mais propensas a experimentar estresse e nervosismo [Usher and Dondio 2020], e observando o conteúdo textual na Figura 5 d, sobressai o uso da palavra “sofrer” e “não”(“não faz”, “não usar”, “não fere”), entretanto ficou muito difícil de conseguir identificar outras características. Levantando a hipótese que devido a baixa representatividade do Neuroticismo na base textual, pode ter afetado a extração de *insights*, representando apenas 7.88% da base, mostrando que provavelmente os algoritmos de AM que foram testados poderiam apresentar problemas na identificação desse traço, levantando a oportunidade de utilizar características do perfil desses usuários como: quantidade de seguidores e quantidade de pessoas que estão seguindo.

### 4.3. Modelagem

Antes dos dados serem fornecidos como entrada para os algoritmos, foi realizada a etapa de seleção de variáveis com o método ANOVA para identificar quais palavras conseguem ajudar a diferenciar os traços de personalidade (Figura 6 e Figura 7). Nessa etapa as variáveis de entrada foram reduzidas de 1500 para 149 considerando *unigram*, e 187 para *bigram*. Na Figura 6 pode-se perceber que há palavras que precisarão ser tratadas porque dependendo do contexto podem ser ruídos e afetar o desempenho dos algoritmos de AM como: *sky* (palavra em inglês) e *avatar* (filme ou personagem de um jogo).

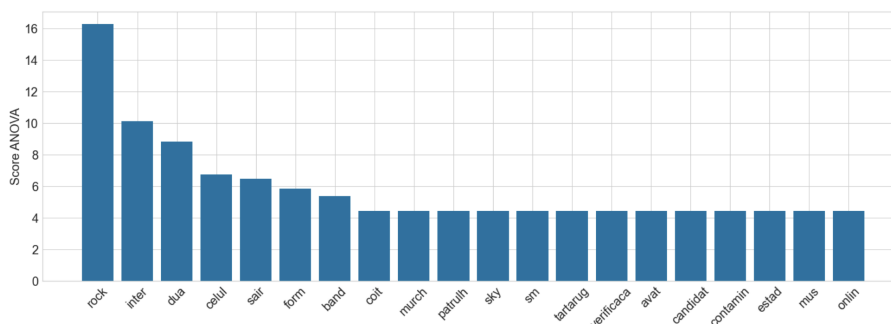
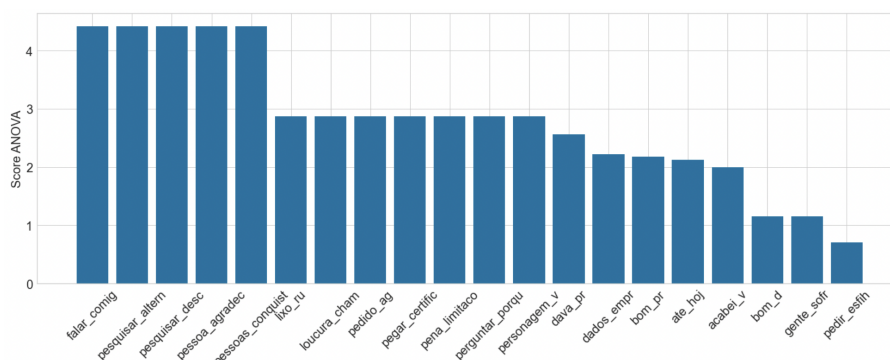


Figura 6. 20 Variáveis textuais (*unigram*) mais importantes



**Figura 7. 20 Variáveis textuais (*bigram*) mais importantes**

Devido ao desbalanceamento entre os traços de personalidade presentes na base, encontrados na subseção 4.1, os algoritmos foram treinados e testados utilizando a validação cruzada em dois cenários diferentes, cujo cenário 1 os dados foram fornecidos para os algoritmos de AM sem nenhum balanceamento, no cenário 2 os dados foram balanceados utilizando a técnica SMOTE antes de serem fornecidos para os algoritmos de AM. Com a aplicação do SMOTE a base de dados aumentou de 54 para 80.

#### 4.4. Avaliação dos resultados

Para avaliar o desempenho dos modelos construídos em ambos os cenários, foram consideradas as métricas AUC, MCC e *F1-score*. Na Tabela 3 pode-se perceber que o desempenho da maioria dos modelos no cenário 1 não foi bom, seja utilizando principalmente *bigram* pois o percentual de acertos foi baixo comparando com o *unigram*. Esse resultado levantou a hipótese que a base de dados é muito pequena para que os algoritmos consigam tirar uma inferência assertiva entre as personalidades e as palavras utilizadas, os melhores algoritmos foram a Regressão Logística considerando TF-IDF *unigram* com 0.84 de AUC, 0.38 de MCC e 0.45 de *F1-score*, *Random Forest* TF-IDF *unigram* com 0.77 de AUC, 0.36 de MCC e 0.39 de *F1-score*. Reforçando que o desbalanceamento entre as classes dos traços de personalidades impacta no desempenho dos modelos.

Os resultados do cenário 2 na Tabela 3 confirmam a hipótese levantada, porque o balanceamento das classes melhorou o desempenho de todos os algoritmos, assim como nas conclusões de [Ryan et al. 2023] e [Lucky et al. 2021]. O algoritmo que apresentou maior equilíbrio entre as métricas selecionadas foi a Regressão Logística considerando TF-IDF *unigram*, com 0.97 de AUC, 0.78 de MCC e 0.76 de *F1-score*.

A técnica SMOTE utilizando a abordagem *over-sampling* que criou mais amostras fictícias para a base de dados, foi bem importante para a melhora do valor das métricas dos modelos, assim como na conclusão tirada por [Sekarningtyas et al. 2021] que identificaram que quanto maior for a base de treino utilizado na classificação, melhor fica a performance dos algoritmos.

## 5. Conclusão

As redes sociais permitem o compartilhamento de pensamentos, sentimentos e atividades de seus usuários; seus dados podem ser usados para prever traços de personalidade, que

**Tabela 3. Desempenho dos classificadores com validação cruzada com 10-fold**

Modelo	Sem SMOTE			Com SMOTE		
	AUC	MCC	<i>F1-Score</i>	AUC	MCC	<i>F1-Score</i>
RF + TF-IDF Unigram	0.7758	0.3638	0.3975	0.9296	<b>0.7223</b>	<b>0.7363</b>
AD + TF-IDF Unigram	0.5911	0.1986	0.3722	0.7128	0.4681	0.5379
NB + TF-IDF Unigram	0.6991	0.3419	0.4383	<b>0.9726</b>	0.6659	0.6354
RL + TF-IDF Unigram	<b>0.8441</b>	<b>0.3880</b>	<b>0.4545</b>	<b>0.9789</b>	<b>0.7831</b>	<b>0.7642</b>
SVM + TF-IDF Unigram	0.7566	0.4789	0.4853	0.9390	0.6938	0.6797
KNN + TF-IDF Unigram	0.6088	0.2551	0.3668	0.8363	0.5408	0.5339
RF + TF-IDF Bigram	0.7230	0.2675	0.2977	0.9031	0.6242	0.5992
AD + TF-IDF Bigram	0.6369	0.2675	0.2964	0.8136	0.5006	0.5033
NB + TF-IDF Bigram	0.445	0.0745	0.1893	0.8054	0.4842	0.4445
RL + TF-IDF Bigram	0.6594	0.1751	0.2011	0.8976	0.6749	0.6444
SVM + TF-IDF Bigram	0.6647	0.2635	0.2491	0.7859	0.5351	0.4308
KNN + TF-IDF Bigram	0.6113	0.1361	0.2238	0.8175	0.4172	0.4328

podem ser usados para melhorar sistemas de recomendação, detecção de *cyberbullying* e identificação de tendências à disseminação de *fake news* e rumores nas redes sociais.

O objetivo do presente trabalho consistiu na construção de um conjunto de dados com *tweets* em português rotulados com o traço de personalidade *Big Five* dominante e verificar o seu potencial de uso em modelos de aprendizado de máquina clássicos. Para isso foi realizada uma sequência de etapas na coleta de dados que se iniciou no compartilhamento em redes sociais do questionário pré-validado *Big Five* (IGFP-5); os traços de personalidade foram calculados seguindo o cálculo feito por [Ferreira 2018]. Em seguida, os dados foram analisados com a finalidade de obter uma dimensão e *insights* sobre as personalidades dos voluntários, 75% da base é composta por mulheres e 83% trabalham na área de tecnologia. O traço dominante mais popular foi a Afabilidade compondo 30% dos voluntários, 26% com Conscienciosidade, 26% com Abertura, 11% com Extroversão e 7% com Neuroticismo.

Devido ao desbalanceamento encontrado entre as personalidades, foram comparados os resultados dos algoritmos de AM em dois cenários: 1- os dados foram fornecidos para os algoritmos de aprendizado de máquina sem nenhum balanceamento; 2- os dados foram balanceados usando a técnica SMOTE antes de serem fornecidos para os algoritmos de aprendizado de máquina. E em ambos os cenários os algoritmos foram treinados e testados utilizando a abordagem de validação cruzada com 10-fold.

Com a técnica SMOTE para criar amostras sintéticas, pode-se observar um impacto positivo para o desempenho dos modelos. Regressão Logística e TF-IDF com *unigram* apresentou o melhor resultado, com 0.97 de AUC, 0.78 de MCC e 0.76 de *F1-score* (Tabela 3). Sem balanceamento, os algoritmos apresentaram dificuldades para diferenciar uma personalidade da outra.

Assim, em um contexto de linguagem natural sem a implementação de outras variáveis (como *emojis*, quantidade de seguidores e quantidade de pessoas que o usuário segue), ao utilizar a técnica SMOTE junto com modelos de Aprendizado de Máquina clássicos, as métricas de AUC, MCC e *F1-Score* foram superiores do que utilizar apenas

os modelos clássicos. Ademais, com os resultados promissores das métricas AUC, MCC e *FI-Score* pode-se inferir que a base de dados apresenta potencial para ser aplicada em modelos de aprendizado de máquina clássicos.

Em trabalhos futuros, mais algoritmos de classificação podem ser considerados na pesquisa como por exemplo: redes neurais, IndoBERT, XGBoost e AdaBoost, para ampliar o escopo de seleção do melhor algoritmo e obter resultados mais precisos para detecção de traços de personalidade. Outras abordagens de balanceamento multiclasse e a inclusão de testes estatísticos como Teste T, podem ser incluídos. Um outro ponto que pode ser explorado é investigar outros ajustes de parâmetros dos algoritmos, para além das configurações padrões consideradas neste presente trabalho.

## Referências

- Andrade, J. M. d. (2008). *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil*. PhD thesis, Universidade Federal de Brasília.
- Balakrishnan, V., Khan, S., Fernandez, T., and Arabnia, H. R. (2019). Cyberbullying detection on twitter using big five and dark triad features. *Personality and individual differences*, 141:252–257.
- Bowden-Green, T., Hinds, J., and Joinson, A. (2021). Understanding neuroticism and social media: A systematic review. *Personality and Individual Differences*, 168:110344.
- Buiar, J. A., Pimentel, A. R., Oliveira, L., and da Silva, Z. C. (2019). Modelo computacional para identificação de perfil de personalidade baseado em textos educacionais. *Nuevas Ideas en Informática Educativa*, 15:30–37.
- Bunker, C. J. and Kwan, V. S. (2021). Do the offline and social media big five have the same dimensional structure, mean levels, and predictive validity of social media outcomes? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 15(4).
- Cahyani, D. E. and Faishal, A. F. (2020). Classification of big five personality behavior tendencies based on study field with twitter analysis using support vector machine. In *2020 7th ICITACEE*, pages 140–145. IEEE.
- Camati, R. S. (2021). *Reconhecimento Automático de Personalidade a partir de textos: uma abordagem baseada em técnicas projetivas*. PhD thesis, PUC-PR.
- Chen, X. and Wang, N. (2020). Rumor spreading model considering rumor credibility, correlation and crowd classification based on personality. *Scientific Rep.*, 10(1):1–15.
- Christian, H., Suhartono, D., Chowanda, A., and Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data*, 8(1):1–20.
- da Silva, A. G., Loch, A. A., Leal, V. P., da Silva, P. R., Rosa, M. M., Bomfim, O. d. C., Malloy-Diniz, L. F., Schwarzbald, M. L., Diaz, A. P., and Palha, A. P. (2020). Stigma toward individuals with mental disorders among brazilian psychiatrists: a latent class analysis. *Brazilian Journal of Psychiatry*, 43:262–268.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440.

- Effrosynidis, D. and Arampatzis, A. (2021). An evaluation of feature selection methods for environmental data. *Ecological Informatics*, 61:101224.
- Ergu, İ., Işık, Z., and Yankayış, İ. (2019). Predicting personality with twitter data and machine learning models. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–5. IEEE.
- Faceli, K., Lorena, A. C., Gama, J., and Carvalho, A. C. P. d. L. F. d. (2011). *Inteligência artificial: uma abordagem de aprendizado de máquina*. Editora LTC.
- Ferreira, T. B. (2018). Raciocínio baseado em casos para apoiar a formação de grupos baseada nos traços de personalidade do modelo big five. Master's thesis, UFU.
- Flack, J. C. and D'Souza, R. M. (2014). The digital age and the future of social network science and engineering. *Proceedings of the IEEE*, 102(12):1873–1877.
- Genina, A., Gawich, M., and Hegazy, A. F. (2020). A survey for sentiment analysis and personality prediction for text analysis. In *Internet of Things—Applications and Future*, pages 347–356. Springer.
- Grandini, M., Bagli, E., and Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.
- Indira, R. and Maharani, W. (2021). Personality detection on social media twitter using long short-term memory with word2vec. In *2021 IEEE In Conf. COMNETSAT*, pages 64–69. IEEE.
- Kaushal, V. and Patwardhan, M. (2018). Emerging trends in personality identification using online social networks—a literature survey. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(2):1–30.
- Khan, E. M., Mukta, M. S. H., Ali, M. E., and Mahmud, J. (2020). Predicting users' movie preference and rating behavior from personality and values. *ACM TiiS*, 10(3):1–25.
- King, R. D., Orhobor, O. I., and Taylor, C. C. (2021). Cross-validation is safe to use. *Nature Machine Intelligence*, 3(4):276–276.
- KN, P. K. and Gavrilova, M. L. (2021). Latent personality traits assessment from social network activity using contextual language embedding. *IEEE Transactions on Computational Social Systems*, 9(2):638–649.
- Kunte, A. V. and Panicker, S. (2019). Using textual data for personality prediction: a machine learning approach. In *2019 4th In Conf. ISCON*, pages 529–533. IEEE.
- Lucky, H., Suhartono, D., et al. (2021). Towards classification of personality prediction model: A combination of bert word embedding and mlsmote. In *2021 1st ICCSAI*, volume 1, pages 346–350. IEEE.
- Majumder, N., Poria, S., Gelbukh, A., and Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2):74–79.
- Marques Junior, P. R. M. (2018). Detecção de transtorno de personalidade narcisista a partir de tweets: um estudo preliminar. Trabalho de conclusão de curso, UFF.

- Moreno, D. R. J., Gomez, J. C., Almanza-Ojeda, D.-L., and Ibarra-Manzano, M.-A. (2019). Prediction of personality traits in twitter users with latent features. In *2019 CONIELECOMP*, pages 176–181. IEEE.
- Mota, F. B. d. S. et al. (2022). Análise de traços de personalidade em ambientes de participação eletrônica. Master's thesis, Universidade Federal de Itajuba.
- Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., and Abedi, V. (2019). Artificial intelligence transforms the future of health care. *The American J. of medicine*, 132(7):795–801.
- Pratama, R. P. and Maharani, W. (2021). Predicting big five personality traits based on twitter user using random forest method. In *2021 ICoDSA*, pages 110–117. IEEE.
- Ryan, G., Katarina, P., and Suhartono, D. (2023). Mbti personality prediction using machine learning and smote for balancing data based on statement sentences. *Information*, 14(4):217.
- Salem, M. S., Ismail, S. S., and Aref, M. (2019). Personality traits for egyptian twitter users dataset. In *Conf. Software and Information Engineering*, pages 206–211.
- Sangoju, M. K., Garapati, D. S., Mudu, S. N., and Madhuri, C. R. (2022). User personality prediction in rumor propagation across digital footprints. In *2022 8th ICSSS*, pages 1–8. IEEE.
- Sekarningtyas, A. S., Ayu, M. A., and Mantoro, T. (2021). Using k-nearest neighbor algorithm for personality classification of twitter's users based on the big five theory. In *2021 IEEE 7th ICCED*, pages 1–6. IEEE.
- Smiderle, R., Rigo, S. J., Marques, L. B., Peçanha de Miranda Coelho, J. A., and Jacques, P. A. (2020). The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments*, 7(1):1–11.
- Statista (2022). Social media and user-generated content - leading countries based on number of twitter users as of january 2022. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.
- Tan, J. S., Tan, I. K., Soon, L. K., and Ong, H. F. (2022). Improved automated essay scoring using gaussian multi-class smote for dataset sampling. In *Proceedings of the 15th International Conference on Educational Data Mining*, page 647.
- Tutaysalgir, E., Karagoz, P., and Toroslu, I. H. (2019). Clustering based personality prediction on turkish tweets. In *2019 IEEE/ACM In Conf. on ASONAM*, pages 825–828. IEEE.
- Usher, J. and Dondio, P. (2020). Brexit: Psychometric profiling the political salubrious through machine learning: Predicting personality traits of boris johnson through twitter political text. In *Conf. Web Intelligence, Mining and Semantics*, pages 178–183.
- Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D., and Lestantyo, P. (2019). Cross-validation metrics for evaluating classification performance on imbalanced data. In *2019 In Conf. on computer, control, informatics and its applications (IC3INA)*, pages 14–18. IEEE.