

**UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E
TECNOLOGIA DA COMPUTAÇÃO**

Análise de Traços de Personalidade em
Ambientes de Participação Eletrônica

Flávio Belizário da Silva Mota

Itajubá, 13 de dezembro 2021

**UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E
TECNOLOGIA DA COMPUTAÇÃO**

Flávio Belizário da Silva Mota

**Análise de Traços de Personalidade em
Ambientes de Participação Eletrônica**

Dissertação submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de Mestre em Ciência e Tecnologia da Computação.

Área de Concentração: Matemática da Computação

Orientador: Prof^ª. Dr^ª. Melise Maria Veiga de Paula

Coorientador: Prof^ª. Dr^ª. Isabela Neves Drummond

13 de dezembro 2021

Itajubá

Flávio Belizário da Silva Mota

Análise de Traços de Personalidade em Ambientes de Participação Eletrônica

Dissertação submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de Mestre em Ciência e Tecnologia da Computação.

Trabalho aprovado. Itajubá, 13 de Dezembro de 2021:

Prof^a. Dr^a. Melise Maria Veiga de Paula
Orientador

Prof^a. Dr^a. Isabela Neves Drummond
Coorientador

Prof^a. Dr^a. Flavia Cristina Bernardini

Prof. Dr. Carlos Henrique da Silveira

Itajubá
13 de dezembro 2021

Agradecimentos

Agradeço aos meus pais, Rita e Jorge, pelo apoio incondicional e por priorizarem a educação de seus filhos sempre. A minhas irmãs, Ana e Giovanna, obrigado por serem minhas grandes amigas. Aos meus padrinhos, Carla e Leonardo, por desde cedo acreditarem em mim. Por todo amor, compreensão e paciência, sou infinitamente grato.

Agradeço minhas orientadoras Melise Paula e Isabela Drummond pela paciência, conselhos valiosos e, certamente, pela amizade. Este trabalho não seria possível sem vocês. Obrigado por me guiar por todo caminho, da graduação até aqui, me motivando e comemorando cada conquista. Vocês são minha grande inspiração. Tenho certeza que nossa parceria e amizade será duradoura.

Às professoras e professores, em especial do Instituto de Matemática e Computação da UNIFEI, que me acompanharam na jornada da graduação e da pós-graduação. Sua dedicação ímpar na tarefa de ensinar fortaleceu em mim o desejo de aprender cada vez mais e seguir a carreira acadêmica. A vocês todo respeito e admiração.

Aos meus companheiros de trabalho e pesquisa, com quem aprendi muito e se tornaram grandes amigos, em especial Jonas, Bárbara e Alerson. Agradeço também à minha equipe, Rafael, Anna, Alejandro e Djenane, pela compreensão e apoio durante o desenvolvimento desse trabalho.

A meu grande amigo Felipe, pela compreensão e por segurar minha mão nos momentos que mais precisei, sempre me motivando, apoiando e, principalmente, lembrando dos motivos pelos quais decidi seguir o caminho da pesquisa. Agradeço também a minha querida amiga Lígia, pelo carinho e apoio infinitos e pelas horas de conversas sempre muito agradáveis. Por todos os sorrisos e bons momentos que tivemos, sou imensamente grato a vocês dois.

Agradeço à Universidade Federal de Itajubá e ao POSCOMP pelo ensino de alta qualidade ofertado. Por fim, agradeço à CAPES por financiar minhas pesquisa e a todos que estiveram presentes e colaboraram durante a realização deste trabalho.

*“E se a experiência de conhecer o futuro mudasse uma pessoa?
E se evocasse um sentido de urgência, um sentido de obrigação
de agir do modo que sabia que agiria?”
(Ted Chiang, História da sua vida)*

Resumo

A Participação Eletrônica, ou e-Participação, é definida como a participação dos cidadãos na tomada de decisão junto a gestão pública utilizando as Tecnologias de Informação e Comunicação. A e-Participação é um campo de pesquisa multidisciplinar, com contribuições de áreas como ciência política, sociologia, administração, psicologia e economia, bem como contribuições de natureza mais técnica, como a computação. A área da psicologia, especificamente, permite a investigação da personalidade humana e de suas motivações, fornecendo representações estruturais dos traços de personalidade e possibilitando descrever as intenções de um indivíduo em participar desses ambientes. Sendo assim, este trabalho tem como objetivo o estudo de técnicas computacionais para identificação e análise de traços de personalidade em ambientes de Participação Eletrônica. Para isso foram empregados algoritmos de Aprendizado de Máquina, em específico modelos de regressão, e a teoria dos Cinco Grandes Fatores de personalidade, também conhecido como Big Five. Na análise dos resultados foi constatado que o modelo de Floresta Aleatória obteve o melhor desempenho, com um erro absoluto médio igual a 0,02619. Além disso, no contexto da ferramenta analisada, o traço de personalidade que mais se destaca é o de Abertura à Experiência, acompanhado pelos traços de Responsabilidade e Cordialidade. Já os traços de Extroversão e Neuroticismo aparecem com pontuações mais baixas.

Palavras-chaves: Participação Eletrônica. Aprendizado de Máquina. Traços de Personalidade.

Abstract

Electronic Participation, or e-Participation, is defined as citizens' participation in decision-making from public administration using Information and Communications Technologies. The e-Participation is a multidisciplinary research field, with contributions from areas such as political science, sociology, administration, psychology, and economics, as well as contributions of a more technical nature, such as computer science. The field of psychology, specifically, allows the investigation of human personality and its motivations, providing structural representations of personality traits and making it possible to describe an individual's intentions to participate in these environments. Thus, this work aims to study computational techniques for identifying and analyzing personality traits in Electronic Participation environments. To aim this, Machine Learning algorithms, specifically regression models, and the Five-Factor Model, also known as the Big Five, were used. Analyzing the results, it was found that the Random Forest model had the best performance, with a mean absolute error of 0.02619. In addition, in the context of the tool analyzed, the personality trait that stands out the most is Openness, accompanied by Conscientiousness and Agreeableness. Extraversion and Neuroticism traits appear with lower scores.

Key-words: Electronic Participation. Machine Learning. Personality traits.

Lista de ilustrações

Figura 1 – E-participação	17
Figura 2 – Hierarquia clássica de aprendizado.	23
Figura 3 – Representação de um Perceptron Multicamadas com duas camadas ocultas.	24
Figura 4 – Representação da diferença vetorial de similaridade entre palavras.	31
Figura 5 – Metodologia.	34
Figura 6 – Distribuição dos traços de personalidade no conjunto de dados do Twitter.	38
Figura 7 – Distribuição do traço Abertura à Experiência - Twitter.	39
Figura 8 – Distribuição do traço Responsabilidade - Twitter.	39
Figura 9 – Distribuição do traço Extroversão - Twitter.	40
Figura 10 – Distribuição do traço Cordialidade - Twitter.	40
Figura 11 – Distribuição do traço Neuroticismo - Twitter.	41
Figura 12 – Representação vetorial de uma sentença.	42
Figura 13 – Erros de previsão do regressor Floresta Aleatória.	45
Figura 14 – Distribuição dos traços de personalidade no conjunto de dados do SoPa.	46
Figura 15 – Distribuição do traço Abertura à Experiência - SoPa.	47
Figura 16 – Distribuição do traço Responsabilidade - SoPa.	47
Figura 17 – Distribuição do traço Extroversão - SoPa.	48
Figura 18 – Distribuição do traço Cordialidade - SoPa.	48
Figura 19 – Distribuição do traço Neuroticismo - SoPa.	49
Figura 20 – Diagrama de cordas das faixas de personalidade do SoPa.	50
Figura 21 – Diagramas de corda para o traço A.	51
Figura 22 – Diagrama de corda para o traço R.	51
Figura 23 – Diagramas de corda para o traço E.	52
Figura 24 – Diagrama de corda para o traço C.	53
Figura 25 – Diagramas de corda para o traço N.	53

Lista de tabelas

Tabela 1 – Os traços de personalidade do modelo Cinco Grandes Fatores.	21
Tabela 2 – Bibliotecas empregadas no trabalho.	35
Tabela 3 – Exemplo do conjunto de dados do Twitter.	36
Tabela 4 – Exemplo do conjunto de dados do SoPa.	36
Tabela 5 – Exemplo do conjunto de dados após rotulagem.	38
Tabela 6 – Métricas dos regressores.	44
Tabela 7 – Faixas definidas.	49

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina	22
API	<i>Application Programming Interface</i>	29, 34, 35, 37
CGF	Cinco Grandes Fatores	20, 21, 29, 54
HTTP	<i>Hypertext Transfer Protocol</i>	34
ONU	Organização das Nações Unidas	16
PLN	Processamento de Linguagem Natural	28, 29
TIC	Tecnologias de Informação e Comunicação	13, 16, 54

Sumário

1	INTRODUÇÃO	13
1.1	Contextualização	13
1.2	Objetivos	14
1.3	Organização do Trabalho	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Participação Eletrônica	16
2.2	Modelo dos Cinco Grandes Fatores (<i>Big Five</i>)	19
2.3	Aprendizado de Máquina	22
2.3.1	Algoritmos de Regressão	23
2.3.1.1	Perceptron Multicamadas (MLP)	24
2.3.1.2	<i>k</i> -Vizinhos Mais Próximos (KNN)	24
2.3.1.3	Árvore de Decisão	25
2.3.1.4	Floresta Aleatória	26
2.3.2	Regressão Multirrótulo	27
2.3.3	Métricas de Erro para Regressão	27
2.4	Processamento de Linguagem Natural	28
2.4.1	PLN e Redes Sociais	28
2.4.2	<i>Word Embedding</i>	29
2.5	<i>Watson Personality Insights</i>	30
3	METODOLOGIA	33
3.1	Coleta de publicações	34
3.2	Rotulagem	37
3.2.1	Rotulagem através do <i>Watson Personality Insights</i>	37
3.2.2	O conjunto de dados rotulado	38
3.3	Vetorização	41
3.4	Regressores Multirrótulo	42
3.4.1	Geração do modelo	43
3.4.2	Identificação dos traços	43
4	RESULTADOS E DISCUSSÃO	44
4.1	Geração dos Modelos	44
4.2	Identificação dos traços do SoPa	46
4.3	Discussão dos Resultados	53

5	CONCLUSÕES	57
5.1	Contribuições	58
5.2	Limitações	58
5.3	Trabalhos Futuros	59
	ANEXOS	60
	ANEXO A – TABELA DE ASPECTOS POSITIVOS DOS TRA- ÇOS DO CGF	61
	REFERÊNCIAS	64

1 Introdução

Este Capítulo apresenta a contextualização sobre a temática abordada neste estudo, as motivações e os objetivos específicos que nortearam a pesquisa, bem como a organização do trabalho.

1.1 Contextualização

A participação, no que tange à democracia, tem como objetivo principal oferecer aos cidadãos a oportunidade de contribuir e influenciar na tomada de decisão do poder público, permitindo participar da formulação ou implementação de políticas públicas. Nesse sentido, as Tecnologias de Comunicação e Informação (TIC) podem ser utilizadas como meio para possibilitar esse tipo de participação, estabelecendo-se assim a Participação Eletrônica (e-Participação). A investigação da Participação Eletrônica a partir de diferentes teorias permite consolidar o conhecimento a respeito da área, promovendo assim o avanço dessas iniciativas e o fortalecimento do processo democrático.

Segundo [Araújo, Penteado e Santos \(2015\)](#) novas modalidades de participação surgem com o aumento do uso da Internet e, conseqüentemente, das redes sociais. Essas redes facilitam a participação dos cidadãos independentemente do tempo e do local. Em seu trabalho, [Berntzen e Johannessen \(2016\)](#) sugerem que as iniciativas de Participação Eletrônica devem se valer do uso das redes sociais em uma abordagem híbrida, na qual as redes são usadas para mobilização cidadã e as ferramentas de e-Participação para o envolvimento dos cidadãos.

Seguindo essa premissa, [Caetano, Paula e Souza \(2020\)](#) desenvolveram uma ferramenta que visa viabilizar a participação utilizando características de redes sociais. A ferramenta, chamada de Sociedade Participativa (SoPa), permite que os cidadãos discutam ideias e debatam soluções para problemas da sua comunidade. Essa interação se dá através de publicações na ferramenta.

Nesse contexto, um aspecto que pode ser considerado é a personalidade dos cidadãos envolvidos. A personalidade representa o conjunto das dimensões de variação do comportamento humano, seus diferentes traços psicológicos e preferências distintas para cada indivíduo ([ALLPORT; ODBERT, 1936](#)).

O estudo dos traços de personalidade se propõe a identificar e descrever a organização estrutural das características intrínsecas ao comportamento humano. Um dos modelos que obteve sucesso no esforço de organizar esses traços foi o modelo dos Cinco Grandes Fatores ([MCCRAE; COSTA, 1987](#)). Baseado em uma extensa revisão do uso de termos

da língua natural empregados para descrever as características dos indivíduos, esse modelo é apoiado por uma série de resultados empíricos transculturais, sendo amplamente estudado.

Considerando o aspecto linguístico no qual o modelo se fundamenta, estratégias computacionais baseadas em Processamento de Linguagem Natural e Aprendizado de Máquina podem ser empregadas para o reconhecimento automático dos traços (LIMA; CASTRO, 2014). Esse tipo de abordagem tem como benefício a análise de grandes quantidades de dados sem a necessidade da aplicação de questionários, geralmente utilizados para a detecção dos traços.

A proposta deste trabalho foi descrever um método de identificação automática de traços de personalidade através de publicações (textos) dos usuários em ambientes de Participação Eletrônica. Especificamente, foram empregadas técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina para identificar os traços de personalidade do modelo dos Cinco Grandes Fatores entre os usuários da ferramenta SoPa. Através dessa identificação, foi realizada uma análise para relacionar os traços de personalidade com as possíveis motivações em participar desses ambientes.

As contribuições deste trabalho estão relacionadas com a investigação de abordagens que visam a ampliação e consolidação do conhecimento da área de Participação Eletrônica, bem como o estudo de técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina aplicados ao contexto do reconhecimento de traços de personalidade.

1.2 Objetivos

Os objetivos específicos desta pesquisa podem ser definidos como:

- Identificar na literatura as interações entre a Participação Eletrônica, as teorias de personalidade e o Aprendizado de Máquina;
- Apresentar e descrever um método para o reconhecimento automático de traços em ambientes de Participação Eletrônica;
- Apresentar e discutir os resultados obtidos com a aplicação do método.

1.3 Organização do Trabalho

Este trabalho está organizado em cinco capítulos. Este primeiro capítulo apresenta a Introdução. No Capítulo 2, são apresentados os conceitos necessários para o entendimento da proposta. O Capítulo 3 descreve os materiais e o método utilizados na condução da pesquisa. No Capítulo 4, são apresentados os resultados obtidos e as análises a respeito. Por fim, o Capítulo 5 aborda as considerações finais, realçando as contribuições, as limitações do trabalho e as oportunidades para trabalhos futuros.

2 Fundamentação Teórica

Neste Capítulo, são apresentados conceitos sobre Participação Eletrônica (Seção 2.1), o modelo dos Cinco Grandes Fatores de personalidade (Seção 2.2), algoritmos de Aprendizado de Máquina para Regressão, as abordagens multirrótulo e métricas de validação (Seção 3.4), conceitos do Processamento de Linguagem Natural e a estratégia de representação de *Word Embedding* (Seção 2.4), e, por fim o serviço da IBM *Watson Personality Insights* (Seção 2.5).

2.1 Participação Eletrônica

A Participação Eletrônica, ou e-Participação, é definida como o uso de tecnologias inovadoras por parte das organizações governamentais visando fornecer aos cidadãos acesso às informações sobre políticas públicas e a possibilidade de opinar sobre elas (MACINTOSH, 2004). Pode ser entendida também como parte da e-democracia, que se caracteriza pelo uso das TIC (Tecnologias de Informação e Comunicação) para apoiar os processos democráticos de tomada de decisão. Sendo assim, a Participação Eletrônica visa garantir que os cidadãos participem da tomada de decisão, especialmente na concepção e avaliação de políticas públicas (QUINTERO-ANGULO; SÁNCHEZ-TORRES; CARDONA-ROMÁN, 2020).

A pesquisa sobre Governo Eletrônico da ONU (Organização das Nações Unidas) de 2020 descreve a e-Participação como uma interseção entre algumas áreas da governança e administração pública (ONU, 2020). Na representação ilustrada na Figura 1, a interseção entre o Governo Eletrônico (e-Governo) e a Participação destaca três aspectos principais da e-Participação: a tomada de decisão, a consulta e a provisão de informação. A relação da Transparência com a e-Participação se dá em função da provisão de informações, inclusive por meio dos Dados Abertos Governamentais. Já a interseção entre a Inclusão e a e-Participação se dá em relação à consulta e a tomada de decisão.

Segundo Mambrey (2004), os avanços das TIC e a sua utilização bem-sucedida em diversas áreas pavimentaram o caminho para que a democracia também pudesse aproveitar de benefícios similares. De acordo com Panopoulou, Tambouris e Tarabanis (2014), utilizar a internet para interagir com entidades governamentais e aumentar a participação política surte um efeito positivo nos cidadãos. Dessa forma, a Participação Eletrônica pode ser vista como um meio para restaurar a confiança política e a satisfação dos cidadãos nos processos democráticos.

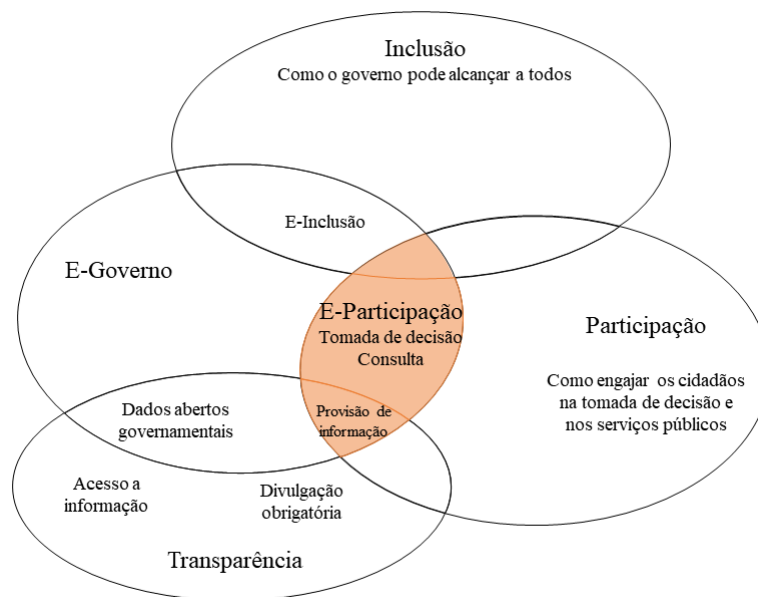


Figura 1 – E-participação

Fonte: [ONU \(2020\)](#) (adaptado).

As iniciativas de e-Participação traduzem-se em ferramentas computacionais que têm como objetivo possibilitar o processo da participação. No trabalho de [Mota et al. \(2020\)](#), os autores apresentam uma proposta de um esquema taxonômico de classificação dessas ferramentas, fornecendo um instrumento de representação do conhecimento na área. Um aspecto importante dessa classificação é em relação à abordagem, que indica a origem da iniciativa, podendo ser vertical, aquela que parte do governo para o cidadão, ou horizontal, iniciativa que parte dos cidadãos podendo ser direcionada para o governo ou ser utilizada para pressionar o poder público.

Em [Ribeiro, Cunha e Barbosa \(2018\)](#), uma análise descritiva a respeito da e-Participação no contexto brasileiro foi conduzida levando em consideração duas pesquisas realizadas pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br). Os resultados apontaram que governo e cidadãos estão em ambientes virtuais, porém as oportunidades de participação, colaboração e interação cidadã são poucas. Da perspectiva do governo, os autores argumentaram que, mesmo que possuindo páginas *web*, grande parte das entidades governamentais não estava disponível para receber sugestões ou comentários. Da perspectiva dos cidadãos, foi reportado que o uso dessas ferramentas estava fortemente influenciado pela idade e nível de instrução, o que torna a participação nesses ambientes muito menor em comparação às outras atividades na internet, como o uso de redes sociais.

As redes sociais podem ser definidas como ferramentas da Internet que permitem que usuários criem, editem, avaliem e vinculem conteúdos e também se conectem a outros

criadores de conteúdo facilmente (KAPLAN; HAENLEIN, 2010). De acordo com Bertot, Jaeger e Hansen (2012), as entidades governamentais podem se beneficiar do uso das redes sociais sob três aspectos:

- Participação democrática e engajamento: as redes sociais podem ser empregadas para envolver o público na construção de políticas governamentais, fomentar o diálogo participativo e permitir que mais grupos de cidadãos tenham representação na discussão, desenvolvimento e implementação de políticas públicas;
- Coprodução de serviços públicos: governo e público desenvolvem, projetam e fornecem serviços governamentais em conjunto usando as redes sociais, visando melhorar a qualidade do serviço;
- Soluções de *crowdsourcing* e inovações: as entidades governamentais buscam inovação através da exploração do conhecimento público por meio da interação nas redes sociais, visando desenvolver soluções inovadoras para problemas da sociedade.

O trabalho de Rodrigues, Gomes e Borges (2020) analisa o potencial da Participação Eletrônica através de redes sociais de 276 cidades brasileiras com mais de 100000 habitantes, utilizando especificamente o Facebook. Foram levados em consideração aspectos como número de publicações, comentários e compartilhamentos, tamanho da população, renda *per capita* e nível de instrução. Os resultados apontam que as entidades governamentais municipais no Brasil fazem uso das redes sociais, mas há casos onde esse uso é muito básico, podendo ser visto como uma relação unidirecional entre governo e cidadão. Por outro lado, alguns municípios geram um nível alto de interação com os cidadãos, utilizando as redes sociais para ofertar serviços, além de realizar publicações relacionadas aos aspectos culturais, instruções legais e consultas públicas.

Utilizando as mesmas estratégias e tecnologias empregadas em redes sociais, Caetano, Paula e Souza (2020) desenvolveram um artefato de abordagem horizontal chamado Sociedade Participativa (SoPa). A ferramenta tem como objetivo a coleta de informações relevantes, como as opiniões dos cidadãos sobre os problemas da sua cidade e propostas de solução para esses problemas. Tais informações podem ser utilizadas de diversas formas na tomada de decisão do setor público, podendo proporcionar melhorias no meio ambiente local, no convívio e na integração da sociedade, valendo-se das características de uma rede social.

Além disso, o SoPa também é empregado para auxílio da elaboração e revisão do Plano Diretor Municipal (PDM) de alguns municípios. O PDM é um instrumento legal que define a atuação do poder público e da iniciativa privada na oferta de serviços públicos essenciais, bem como quais são as normas para a construção nos espaços urbanos e ru-

rais, buscando assegurar melhores condições de vida para toda a população (SANTORO; CYMBALISTA; CARDOSO, 2005).

Uma problemática recorrente apresentada nos trabalhos citados anteriormente e em outros estudos sobre Participação Eletrônica é o engajamento cidadão (PADEKHENE; THINYANE; MACHIRI, 2017; NO; MOOK; SCHUGURENSKY, 2016). O engajamento cidadão pode ser entendido como as ações que os cidadãos realizam a fim de encontrar e resolver problemas em comum nas comunidades às quais pertencem (ZUKIN et al., 2006). De acordo com Alharbi, Kang e Hawryszkiewicz (2015), o baixo nível de engajamento dos cidadãos em ferramentas on-line é um desafio, uma vez que os benefícios da e-Participação não podem ser atingidos até que os cidadãos realmente se envolvam e façam uso das ferramentas. Hassan e Hamari (2020) argumentam que a e-Participação, mesmo que facilitada por meios digitais relativamente mais acessíveis que os físicos, muitas vezes, representa uma atividade que pode ser vista como algo fora das tarefas comuns do dia a dia, onde os efeitos do envolvimento e do tempo investido são geralmente invisíveis ou levam muito tempo para se materializar.

A pesquisa na área da e-Participação é um campo interdisciplinar. É possível perceber contribuições de áreas como ciência política, sociologia, administração, psicologia e economia, bem como contribuições de natureza mais técnica, como a computação (MEDAGLIA, 2012). Dada a interdisciplinaridade da pesquisa, Cortés-Cediel, Gil e Cantador (2018) apontam que nas áreas da psicologia e *marketing*, a representação e identificação da personalidade humana e suas motivações foram intensivamente investigadas, possibilitando a criação de teorias e modelos que se propõem a identificar, descrever e promover as intenções de um indivíduo para realizar determinada ação ou seguir um comportamento específico. Dessa forma, é possível utilizar o conhecimento gerado por essas áreas para compreender os indivíduos e promover o engajamento dos usuários. Sendo assim, o modelo dos Cinco Grandes Fatores, apresentado na Seção 2.2, pode ser utilizado na identificação dos traços de personalidade dos indivíduos.

2.2 Modelo dos Cinco Grandes Fatores (*Big Five*)

Segundo Rebollo e Harris (2006), a personalidade se refere a padrões de comportamento e atitudes que são típicas de um indivíduo, de forma que os traços de personalidade são diferentes de um indivíduo para outro. Durante a maior parte do século 20, os estudos sobre personalidade debateram a respeito de sua estrutura, tentando encontrar quais são as diferenças individuais duradouras que permitem descrever as características distintas de uma pessoa e como elas são organizadas (COSTA; MCCRAE, 2008). A partir da década de 1980, notou-se que cinco fatores são suficientes para abranger os termos descritivos de um traço de personalidade em função de línguas naturais como inglês e alemão, e que

esses mesmos cinco fatores foram encontrados, integralmente ou em partes, na maioria das medidas de diferenças entre indivíduos (DIGMAN, 1990). A essa abordagem se dá o nome de Cinco Grandes Fatores.

O modelo de CGF (Cinco Grandes Fatores) de personalidade, também conhecido como *Big Five*, é considerado uma teoria explicativa e preditiva da personalidade humana e de suas relações com a conduta (SILVA; NAKANO, 2011). Os cinco fatores, ou traços, em questão são: Abertura à Experiência(A), Responsabilidade(R), Extroversão(E), Cordialidade(C) e Neuroticismo(N). Um indivíduo recebe uma pontuação em uma escala contínua para cada um dos traços, podendo ter pontuações altas ou baixas a depender do traço (BENET-MARTÍNEZ; JOHN, 1998).

O traço de Abertura à Experiência(A) é entendido como a tendência que os indivíduos têm de buscar novas atividades e interesses, caracterizando os comportamentos exploratórios e o nível de importância que é atribuído às novas experiências. Pessoas com pontuações altas para esse traço tendem a ser criativas, curiosas, tolerantes e apresentam uma capacidade para reavaliar valores sociais (LOPES; SALOVEY; STRAUS, 2003). Já valores baixos para o traço indicam pessoas com um leque limitado de interesses, essencialmente conservadoras. Geralmente, está relacionado com indivíduos mais religiosos e que são aversos às mudanças profundas e radicais (COSTA; MCCRAE, 1988).

O fator Responsabilidade(R) representa o grau de organização, controle e motivação para alcançar um objetivo. Pontuações altas para esse traço indicam indivíduos mais decididos, capazes de seguir normas e padrões de comportamento, cautelosos e ponderados. Tendem a ser pessoas metódicas e com forte sentido de dever. Por outro lado, uma pontuação baixa indica indivíduos desorganizados, impulsivos, sem respeito pelas obrigações pessoais e sociais (TRENTINI et al., 2009).

A Extroversão(E) está ligada a quantidade e intensidade das interações interpessoais, afetividade positiva e sociabilidade. Indivíduos com pontuações altas tendem a ser mais afetuosos, sociáveis, assertivos e ativos. Geralmente, se arriscam mais, são despreocupados e otimistas. De forma oposta, valores mais baixos para o traço indicam indivíduos introvertidos, socialmente reservados e distantes (exceto com amigos íntimos) (LOPES et al., 2014).

O traço de Cordialidade(C) refere-se à qualidade das relações do indivíduo com outros. Pontuações altas revelam pessoas sinceras, generosas, empáticas, capazes de perceber e interpretar bem suas próprias emoções e as daqueles a sua volta. Já valores baixos refletem indivíduos egocêntricos, sem escrúpulos e manipuladores (TRENTINI et al., 2009).

Por fim, o Neuroticismo(N) está relacionado com a instabilidade emocional e com a tendência do indivíduo experimentar estados emocionais negativos. Valores altos para

esse traço caracterizam pessoas ansiosas, preocupadas, que vivenciam mudanças de humor e desconforto ao estarem próximas de outras pessoas. Pontuações mais baixas indicam indivíduos mais estáveis, que tendem a responder aos estímulos emocionais de maneira controlada. Geralmente, são equilibrados, calmos e controlados (HUTZ et al., 1998). A Tabela 1 sumariza os traços de personalidade em termos de pontuações altas e baixas.

Tabela 1 – Os traços de personalidade do modelo Cinco Grandes Fatores.

Traço de Personalidade	Pontuação alta	Pontuação baixa
Abertura à Experiência(A)	Imaginativos	Convencionais
Responsabilidade(R)	Organizados	Espontâneos
Extroversão(E)	Sociáveis	Reservados
Cordialidade(C)	Confiáveis	Competitivos
Neuroticismo(N)	Preocupados	Estáveis

Segundo [García \(2006\)](#), os traços de personalidade no modelo CGF devem ser avaliados em conjunto para se obter um perfil suficiente e adequado de um indivíduo. Essa forma de analisar o perfil difere de outras teorias de traços como, por exemplo, o Indicador Tipológico de Myers-Briggs, que considera os traços de personalidade como pares de motivações/preferências dicotômicos ([MYERS; MYERS, 1995](#)).

O modelo Cinco Grandes Fatores é criticado por alguns autores por ser fruto de metodologias empiristas, e não de uma teoria propriamente dita, que justifique os cinco fatores como uma composição total da personalidade ([GOLDBERG, 1993](#)). Entretanto, o modelo é um dos mais estudados na área, encontrando suporte em estudos transculturais que corroboram com sua universalidade ([CUPERMAN; ICKES, 2009](#); [SILVA et al., 2017](#)). Segundo [McCrae e Costa \(1997\)](#), tal universalidade se deve ao fato de que as tendências inatas dos indivíduos, ou seja, os traços com base biológica, são um conjunto de características comuns à espécie humana, e que interagem com o meio social para orientar o comportamento e produzir ações, representando uma consequência psicológica fruto das experiências humanas compartilhadas pela vida em grupo.

De acordo com [John, Angleitner e Ostendorf \(1988\)](#), a base para o modelo CGF foram os estudos de termos para definir os traços em linguagem natural. Isso se deve ao fato de que a personalidade é definida por termos simples, por exemplo, amigável, sensível e pontual. Esses são termos através dos quais os indivíduos entendem a si mesmos e aos outros. A investigação da personalidade tenta explicar os fenômenos referenciados por esses termos e as maneiras como são usados na vida cotidiana. Reforçando a premissa da universalidade do modelo, estudos conduzidos em alemão ([BORKENAU; OSTENDORF, 1990](#)), japonês ([BOND; FORGAS, 1984](#)), chinês ([YANG; BOND, 1990](#)), português ([PRIMI, 2010](#)) e espanhol ([BERNARDOS, 1992](#)) foram capazes de identificar estruturas de personalidade similares em todas as línguas. O trabalho de [McCrae \(1990\)](#) foi um dos primeiros a apresentar uma organização dos termos que definem as

características positivas dos traços de personalidade, como mostrado no Anexo A

Assim como a maioria das teorias de personalidade, o modelo de Cinco Grandes Fatores se vale de questionários cujas respostas são dadas em escalas, possibilitando a medida quantificada dos traços definidos na teoria para um indivíduo (GOLDBERG; MCREYNOLDS, 1971). O questionário mais comumente aplicado é o *Revised NEO Personality Inventory (NEO-PI-R)* (COSTA; MCCRAE, 2008), cuja versão brasileira foi desenvolvida por Flores-Mendoza (2007).

Entretanto, considerando os aspectos linguísticos em que o modelo se baseia, notou-se que as medidas feitas pelos questionários podem também ser obtidas através da análise do discurso de um indivíduo. O trabalho de Mairesse et al. (2007) analisa a correlação entre a linguagem e os traços de personalidade através do estudo de marcadores linguísticos do discurso escrito e oral. Os resultados obtidos demonstraram-se estatisticamente significativos. Os autores argumentam que, utilizando esse tipo de análise da linguagem, é possível estabelecer modelos computacionais capazes de realizar reconhecimento automático de traços de personalidade.

2.3 Aprendizado de Máquina

O Aprendizado de Máquina (AM) é uma das áreas da Inteligência Artificial que engloba a utilização de técnicas computacionais capazes de melhorar o desempenho na realização de alguma tarefa, através da indução de uma hipótese a partir de experiências anteriores. Segundo Mitchell (1997), o AM tem por interesse a construção de programas de computadores que se aperfeiçoam automaticamente. Mais especificamente, os algoritmos de Aprendizado de Máquina são projetados para aprender a induzir uma função ou hipótese capaz de resolver um problema, obtendo conclusões genéricas de um conjunto particular de exemplos. De acordo com Faceli et al. (2021), os algoritmos de AM podem ser divididos em função da tarefa de aprendizado, podendo ser uma tarefa preditiva ou descritiva.

Em tarefas preditivas, o aprendizado baseia-se na existência de um supervisor externo que conhece a saída desejada para um exemplo do conjunto de dados, e informa ao algoritmo quão bem ele está atuando. Nesse modelo de aprendizado, os algoritmos aprendem ao serem apresentados a dados de treinamento previamente classificados, ou seja, que contém os valores de entrada e seus respectivos resultados esperados, geralmente chamados de rótulos (COPPIN, 2013). Pela existência de um supervisor externo, esses algoritmos seguem o paradigma do aprendizado supervisionado.

Para as tarefas descritivas, os algoritmos de AM buscam encontrar padrões no conjunto de dados sem o auxílio de um supervisor externo, uma vez que o objetivo não é a predição de um valor específico já conhecido (JAIN; MURTY; FLYNN, 1999). Por

atuarem “sem supervisão”, esses algoritmos seguem o paradigma do aprendizado não supervisionado. A Figura 2 apresenta a hierarquia clássica das categorias do aprendizado e suas tarefas.

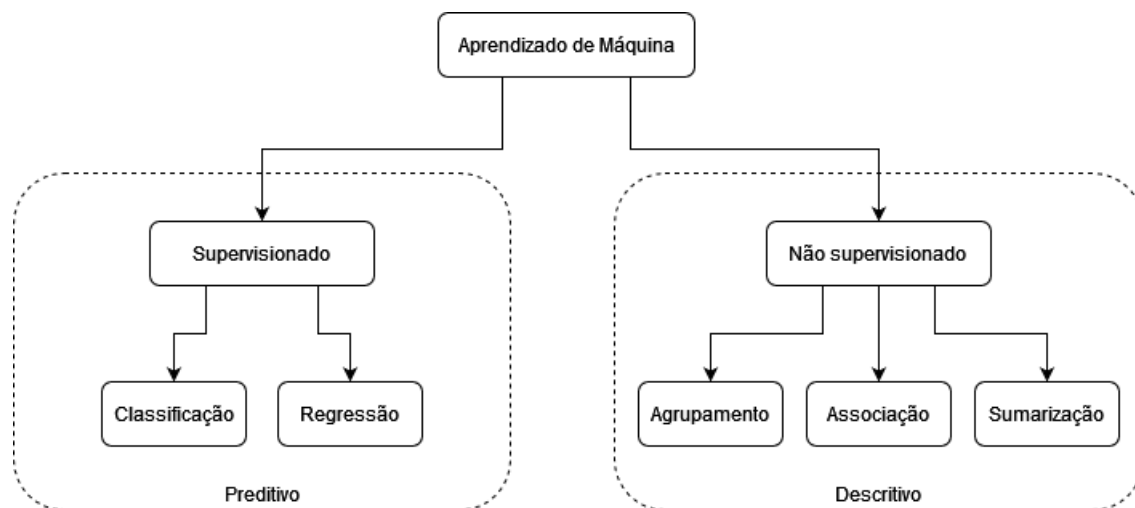


Figura 2 – Hierarquia clássica de aprendizado.

Fonte: [Faceli et al. \(2021\)](#).

As tarefas preditivas são divididas em classificação e regressão. Essa divisão se dá em função do valor do rótulo a ser encontrado, sendo que para valores discretos é empregada a classificação. Já para valores contínuos, é empregada a regressão. Para as tarefas descritivas, a divisão geralmente é feita em: agrupamento, que realiza a divisão dos dados em função de sua similaridade; associação, que procura padrões frequentes entre os atributos do conjunto de dados; e por fim sumarização, que busca encontrar uma descrição compacta para um conjunto de dados ([FACELI et al., 2021](#)).

Alguns trabalhos propõem uma revisão da hierarquia clássica do aprendizado, adicionando tarefas como o aprendizado semissupervisionado ([ZHU; GOLDBERG, 2009](#)), aprendizado ativo ([SETTLES, 2009](#)) e aprendizado por reforço ([SUTTON; BARTO et al., 1998](#)).

2.3.1 Algoritmos de Regressão

De forma abrangente, algoritmos de regressão são métodos estatísticos que permitem estimar a relação entre um valor de saída e as variáveis que influenciam essa saída ([YILDIZ; BILBAO; SPROUL, 2017](#)). Os modelos de aprendizado supervisionado podem se comportar como classificadores ou regressores dada a estratégia empregada no processo de aprendizagem. Nesta Seção serão apresentadas as estratégias para regressão dos algoritmos Perceptron Multicamadas (MLP, do inglês *Multilayer Perceptron*), k -Vizinhos Mais Próximos (KNN, do inglês *k-Nearest Neighbour*), Árvore de Decisão e Floresta Aleatória.

2.3.1.1 Perceptron Multicamadas (MLP)

As redes neurais artificiais do tipo Perceptron Multicamadas (MLP) se baseiam no Perceptron proposto por [Rosenblatt \(1958\)](#). Essa arquitetura apresenta uma ou mais camadas intermediárias (ocultas) de neurônios conectadas entre si e a informação sempre flui dos neurônios de entrada para os neurônios de saída, não existindo retroalimentação. Os neurônios implementam funções de ativação que fornecem a resposta do neurônio a uma entrada. O algoritmo de *back-propagation* ([WERBOS, 1974](#); [RUMELHART](#); [FELDMAN, 1987](#)) é a regra de aprendizagem mais comumente aplicada no processo de aprendizado do modelo. O MLP é conhecido também por ser um aproximador de funções ([FUNAHASHI, 1989](#)).

De forma simplificada, durante a etapa de aprendizagem, um conjunto de dados é apresentado como um vetor de entrada para a rede, e uma função de erro é definida. Cada conexão da rede possui um peso, que é atualizado iterativamente até que o vetor de saída encontrado pela rede seja igual ao valor da saída real do conjunto de treinamento ([GARRIDO; OÑA; OÑA, 2014](#)). Cada neurônio na camada de saída da rede representa uma classe do conjunto de dados, sendo que para problemas de regressão a função de ativação desse neurônio é uma função identidade ([FACELI et al., 2021](#)). Os termos vetores de entrada e saída referem-se às entradas e saídas do MLP, como mostrado na Figura 3.

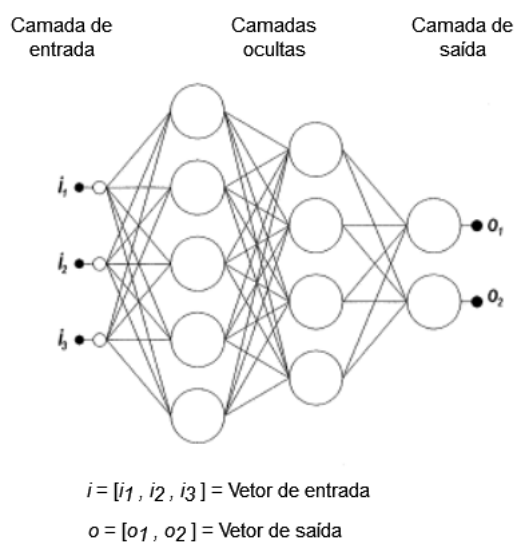


Figura 3 – Representação de um Perceptron Multicamadas com duas camadas ocultas.

Fonte: [Gardner e Dorling \(1998\)](#) (adaptado).

2.3.1.2 k -Vizinhos Mais Próximos (KNN)

O algoritmo KNN se baseia na representação das amostras a serem classificadas como pontos, definidos pelos atributos, em um espaço de entrada. Utilizando uma métrica

nesse espaço, é possível calcular a distância entre dois pontos, sendo a distância euclidiana a métrica mais comumente empregada. O algoritmo primeiramente memoriza todas as amostras já rotuladas. Para classificar um novo exemplo, computa a distância entre esse ponto e os k pontos vizinhos e encontra aqueles mais próximos. Então, cada ponto vizinho mais próximo fornece uma classe e o agregado dessas previsões é utilizado para definir o rótulo da nova amostra (COVER; HART, 1967). Segundo Faceli et al. (2021), em problemas de regressão essa agregação é feita utilizando a média ponderada, definida na Equação 2.1.

$$y_t = \frac{\sum_{i=1}^k w_i y_i}{\sum w_i} \quad (2.1)$$

onde y_i é o rótulo associado ao ponto x_i e w_i é definido por

$$w_i = \frac{1}{d(x_t, x_i)} \quad (2.2)$$

em que $d(x_t, x_i)$ é a métrica de distância entre os pontos.

O KNN possui como vantagens a simplicidade do funcionamento, transparência, robustez contra dados de treinamento ruidosos e facilidade de implementação. Já as desvantagens incluem a complexidade de computação, limitação de memória, baixo desempenho no tempo de execução em grandes conjuntos de treinamento e influência negativa de atributos irrelevantes (CUNNINGHAM; DELANY, 2021).

2.3.1.3 Árvore de Decisão

O algoritmo de Árvore de Decisão tem como objetivo criar um modelo que seja capaz de encontrar a classe de uma amostra aprendendo regras de decisão simples inferidas a partir dos atributos do conjunto de dados (BREIMAN et al., 1984).

A construção de uma Árvore de Decisão envolve a divisão recursiva do conjunto de treinamento, utilizando critérios de separação até que a maior parte das amostras pertencente à partição, ou nó, sejam de uma mesma classe. Esses critérios de separação indicam quão bem determinado atributo discrimina as classes, permitindo selecionar aquele que maximiza ou minimiza essa medida (RUTKOWSKI et al., 2013).

O cálculo do critério de separação em tarefas de regressão é dado pela métrica SDR (do inglês *Standard Deviation Reduction*), proposta por Breiman et al. (1984). Considerando D como o conjunto de dados e n a quantidade de amostras do conjunto, a variância da classe alvo y é dada pela Equação 2.3, onde \bar{y} representa a média dos valores

das classes.

$$sd(D, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.3)$$

Se sobre um atributo A for aplicado um teste de hipótese h_A onde, por exemplo, $A \leq a$, a métrica SDR para estimar a redução da variância em y pode ser obtida conforme a Equação 2.4, utilizando o teste h_A .

$$SDR(h_A) = sd(D, y) - \frac{n_L}{n} \times sd(D_L, y) - \frac{n_R}{n} \times sd(D_R, y) \quad (2.4)$$

Pela Equação 2.4, o conjunto de dados D é dividido em dois subconjuntos, D_L e D_R , que têm tamanhos n_L e n_R , de forma que $n = n_L + n_R$. Considerando cada atributo e cada teste possível no valor do atributo, é realizado o cálculo da variância associada ao teste. O teste que obtiver maior redução na variância é escolhido como critério de separação para o nó.

O processo de classificação de novas amostras se dá pelo percorrimto na árvore, em que a cada nó, um teste é realizado considerando os valores dos atributos. Esses testes são realizados até que um nó folha seja atingido, permitindo obter o valor da classe para a amostra (QUINLAN, 1986).

De acordo com Faceli et al. (2021), em problemas de regressão, usualmente busca-se minimizar o erro quadrático, sendo a média a constante que minimiza esse erro. Dessa forma, o valor de uma classe associada a um nó folha da Árvore de Decisão representa a média dos valores das classes utilizadas no treinamento que caíram nessa folha. Quando aplicado à tarefa de regressão, o algoritmo pode receber o nome de Árvore de Regressão.

2.3.1.4 Floresta Aleatória

O algoritmo de Floresta Aleatória é uma combinação homogênea de classificadores do tipo Árvore de Decisão, no qual várias árvores são geradas e treinadas utilizando amostragem com reposição do conjunto de dados, combinada com seleção aleatória de atributos. As previsões geradas por cada árvore são combinadas para gerar uma predição final (BREIMAN, 2001).

Segundo Faceli et al. (2021), o algoritmo utiliza dois parâmetros: o número de árvores da floresta (L) e o número de atributos a considerar para os testes de decisão em cada nó da árvore (i). São geradas L árvores utilizando uma amostra com reposição do conjunto de dados. No momento de escolher o atributo de teste para gerar a árvore, são considerados apenas i atributos, selecionados de forma aleatória. Na tarefa de classificação, o rótulo final de uma amostra é dado pela classe que foi majoritariamente prevista pelas

árvores. Já para a tarefa de regressão, o rótulo é obtido considerando a média das previsões (SVETNIK et al., 2003).

2.3.2 Regressão Multirrótulo

O aprendizado supervisionado não se limita a identificar uma única classe para os classificadores, ou um único valor de saída para os regressores. Quando a tarefa de classificação ou regressão envolve prever simultaneamente múltiplas saídas com base em uma entrada, ela é chamada de multirrótulo (XU et al., 2019). Segundo Borchani et al. (2015), as abordagens para a Regressão Multirrótulo podem ser categorizadas em duas: transformação do problema e adaptação de algoritmo.

Na transformação do problema, uma tarefa de regressão com múltiplas saídas é transformada em diversas regressões com uma única saída. Nessa abordagem é construído um modelo para cada classe e, posteriormente, as previsões obtidas individualmente são concatenadas. A vantagem dessa abordagem está na sua representação e interpretabilidade, permitindo que o conhecimento gerado por cada modelo possa ser acessado (OSOJNIK; PANOVA; DŽEROSKI, 2017). Já a abordagem de adaptação de algoritmo se baseia na ideia de prever simultaneamente todas as classes usando um único modelo, tendo como vantagem a captura das dependências e relacionamentos internos entre as classes (BREIMAN; FRIEDMAN, 1997).

2.3.3 Métricas de Erro para Regressão

Os algoritmos de aprendizado supervisionado são avaliados por meio do desempenho do modelo gerado na classificação de dados não apresentados durante a etapa de treinamento (MONARD; BARANAUSKAS, 2003). O trabalho de Botchkarev (2019) encontrou evidências na literatura que apontam o uso de três principais métricas de avaliação para a tarefa de regressão, sendo elas o erro quadrático médio (MSE, do inglês *Mean Squared Error*), a raiz quadrada do erro quadrático médio (RMSE, do inglês *Root Mean Squared Error*) e o erro absoluto médio (MAE, do inglês *Mean Absolut Error*). São métricas que calculam a distância entre a saída real e a saída encontrada pelo algoritmo, ou seja, quão bem o modelo é capaz de gerar rótulos que se aproximam dos rótulos reais das amostras. O MSE, o RMSE e o MAE são definidos pelas Equações 2.5, 2.6 e 2.7, respectivamente.

$$MSE(y, \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.5)$$

$$RMSE(y, \hat{y}_i) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.6)$$

$$MAE(y, \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.7)$$

Todas as métricas são sempre não negativas e valores mais baixos são indicativos de melhores modelos.

2.4 Processamento de Linguagem Natural

A busca pelo desenvolvimento de máquinas com capacidade de compreender e se expressar utilizando a linguagem humana levou à criação de um campo de estudo chamado Processamento de Linguagem Natural (PLN). Entendido como parte da grande área de pesquisa da Inteligência Artificial, o PLN engloba conceitos das chamadas ciências cognitivas, como as neurociências, filosofia, psicologia e linguística, bem como a computação (FERREIRA, 2019).

De acordo com Allen (1987), a linguagem deve ser considerada um dos aspectos fundamentais do comportamento humano e um componente crucial de nossas vidas, seja por sua capacidade de transmitir conhecimento a longo prazo enquanto forma escrita, ou como meio de coordenar as atividades diárias dos seres humanos enquanto forma falada. Para o autor, o desenvolvimento de estratégias computacionais na área da linguagem tem duas motivações principais: a científica, que visa obter uma melhor compreensão de como a linguagem funciona, e a técnica, que tem como objetivo a revolução da forma como computadores são utilizados, permitindo que esses sistemas computacionais possam acessar todo o conhecimento humano e, mais ainda, possibilitando que sistemas complexos sejam acessados por todos.

Segundo Hirschberg e Manning (2015), o PLN é um campo de estudo que se beneficia de quatro fatores principais: (i) grande aumento do poder de computação, (ii) disponibilidade de grandes bases de dados linguísticos, (iii) desenvolvimento de algoritmos de Aprendizado de Máquina bem-sucedidos e (iv) uma compreensão mais rica da estrutura da linguagem humana.

2.4.1 PLN e Redes Sociais

As redes sociais são uma fonte de dados rica e vasta e o Processamento de Linguagem Natural é um dos meios mais promissores para o processamento dessas informações. Farzindar e Inkpen (2015) argumentam que é um desafio científico o desenvolvimento de métodos e algoritmos capazes de extrair informações relevantes desse grande volume de dados, muitas vezes proveniente de diferentes fontes e idiomas. Exemplos do uso do PLN em redes sociais podem ser vistos em aplicações para identificação de traços depressivos (GIUNTINI et al., 2021) e risco suicida (COPPERSMITH et al., 2018), identificação de

informações falsas (MEHTA et al., 2021), monitoramento de questões de saúde pública (ABAD et al., 2021) e opiniões sobre produtos (GOZUACIK; SAKAR; OZCAN, 2021).

Um exemplo de rede social capaz de fornecer dados para o PLN é o Twitter. O Twitter é uma rede popular de *microblog* que permite que usuários publiquem e compartilhem mensagens de texto curtas, também conhecidas como *tweets*. A plataforma pode ser acessada através de diversos dispositivos, como computadores e telefones celulares (KWAK et al., 2010). Dados do relatório aos acionistas do 3º trimestre de 2021¹, apontam expressivos 211 milhões de usuários ativos diariamente na rede.

Uma das vantagens que o Twitter tem em relação a outras redes sociais é uma política mais aberta à pesquisa, o que significa que a plataforma disponibiliza recursos para que os dados gerados pelos usuários possam ser coletados e analisados. Para isso, o Twitter oferece uma interface de programação de aplicativos (do inglês *API (Application Programming Interface)*) para acessar esses dados. Essa *API* oferece acesso programático aos principais elementos do Twitter, como: *tweets*, mensagens diretas, listas e dados de usuários (Twitter, 2021). Nos casos onde o uso desses recursos se qualifica como Pesquisa Acadêmica, é possível recuperar, de forma gratuita, até 10 milhões de *tweets* por mês. Os trabalhos de Chaudhry et al. (2021), Alwabel e Zeng (2021), Lopez e Gallemore (2021) e Ceron et al. (2021) trazem aplicações do uso da *API* do Twitter como forma de coleta de dados para permitir a análise de sentimento sobre o período eleitoral dos Estados Unidos no ano de 2020, análise de aceitação de tecnologia orientada a dados e análises sobre a pandemia do novo coronavírus (SARS-CoV-2), respectivamente.

Considerando os aspectos linguísticos nos quais o modelo *Cinco Grandes Fatores* se baseia (vide Seção 2.2), os trabalhos de Qiu et al. (2012), Farnadi et al. (2016), Skowron et al. (2016), Junior e Inkpen (2017) e Gruda et al. (2021) são exemplos de estudos que empregam técnicas de PLN e Aprendizado de Máquina para o realizar reconhecimento automático de traços de personalidade utilizando como fonte de dados o Twitter.

2.4.2 *Word Embedding*

De acordo com Sebastiani (2002), a classificação de textos no contexto do Processamento de Linguagem Natural envolve uma série de etapas. Dentre essas etapas, a representação do texto é o processo de codificar cada palavra do texto como um vetor numérico, possibilitando assim a aplicação de algoritmos de Aprendizado de Máquina (LEOPOLD; KINDERMANN, 2002). Essas representações podem ser obtidas através de estratégias como a contagem do número de vezes que a palavra ocorre no texto, a frequência de sua ocorrência normalizada pelo tamanho total do texto, e a contagem normalizada pela frequência inversa da palavra no texto (IKONOMAKIS; KOTSIANTIS; TAMPAKAS, 2005). Outra estratégia empregada é conhecida como *Word Embedding*.

¹ Relatório aos acionistas - Twitter

Proposta por Mikolov et al. (2013), *Word Embedding* é uma técnica de representação vetorial numérica de palavras que permite obter representações de baixa dimensionalidade, com vetores de até 100 elementos, e que preservam as similaridades semânticas e sintáticas das palavras. Um *embedding* pode ser entendido como uma estrutura matemática que contém outra estrutura embutida. No que se refere às palavras (*word*), significa que uma palavra e seu significado estão embutidos em um vetor (GHOSH; GUNNING, 2019). Essa representação permite que operações algébricas possam ser executadas com os vetores que representam as palavras, como no exemplo onde o vetor para a palavra “Rei” (“*King*”) é subtraído pelo vetor “Homem” (“*Man*”) e somado ao vetor “Mulher” (“*Woman*”), resultando em um vetor que está próximo da representação da palavra “Rainha” (“*Queen*”) (MIKOLOV; YIH; ZWEIG, 2013).

Utilizando como base a técnica de *Word Embedding*, Pennington, Socher e Manning (2014) propuseram o GloVe (*Global Vectors*), que pode ser definido como um algoritmo de aprendizagem não supervisionado para a obtenção de representações vetoriais de palavras que utiliza estatísticas globais de coocorrência palavra-palavra agregadas de um conjunto de dados. Os autores argumentam que o modelo utiliza dos benefícios dos métodos de contagens de palavras enquanto simultaneamente consegue capturar subestruturas lineares significativas, fornecendo assim um modelo de aprendizagem de representações de palavras que supera outros modelos em analogia de palavras, similaridade de palavras e tarefas de reconhecimento de entidades nomeadas. A Figura 4 apresenta dois exemplos da similaridade que pode ser percebida entre duas palavras, quando utilizada a representação vetorial com o GloVe. É esperado que o valor da diferença entre vetores de palavras que representam, por exemplo, gênero ou superlativo, possa se manter quase o mesmo.

Uma vez que um modelo como o GloVe é treinado, as representações geradas podem ser armazenadas no formato de vetores de palavras pré-treinados, e utilizadas como dicionários para encontrar a representação vetorial de uma palavra de uma outra base de dados. Essa estratégia também é conhecida como Transferência de Aprendizado (PAN; YANG, 2010).

2.5 *Watson Personality Insights*

Considerando que a linguagem humana é capaz de refletir a personalidade de um indivíduo (Seção 2.2) e que é possível realizar o reconhecimento de traços de personalidade de forma automática através da análise do discurso escrito (Seção 2.4), a empresa IBM conduziu um conjunto de estudos para entender se as características de personalidade que são inferidas por meio de dados de mídia social podem prever o comportamento e as preferências das pessoas. Como resultado desse estudo, foi criado o serviço *Watson Personality Insights*, que é capaz de inferir características da personalidade com base em

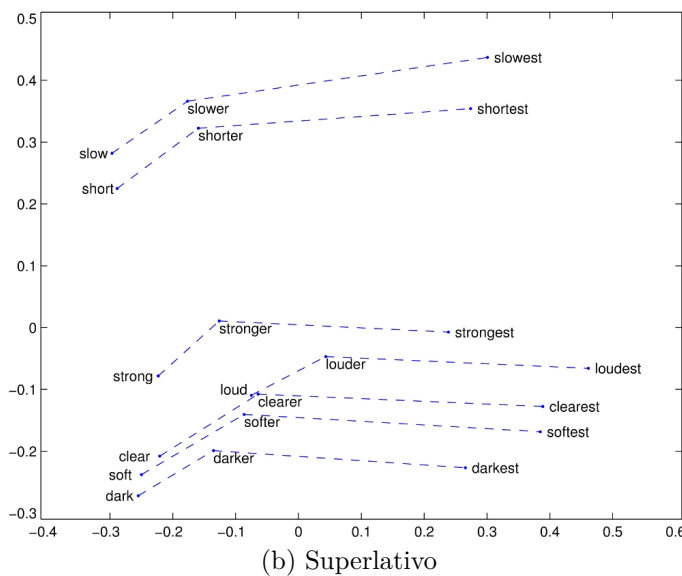
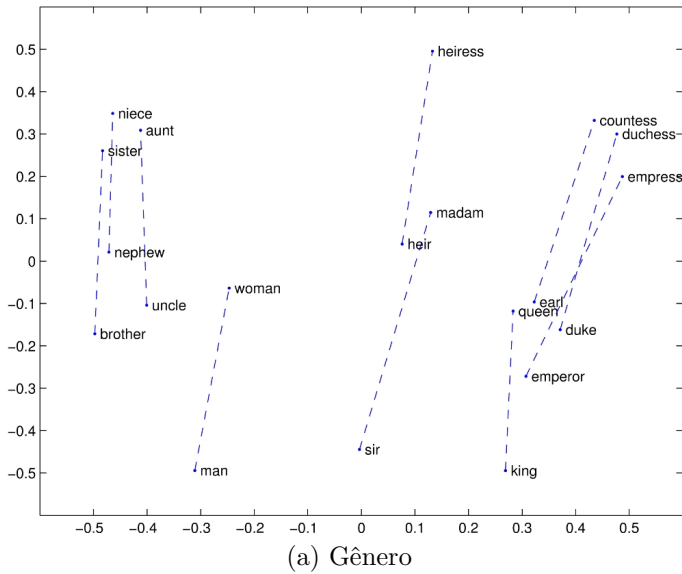


Figura 4 – Representação da diferença vetorial de similaridade entre palavras.

Fonte: [Pennington, Socher e Manning \(2014\)](#).

dados textuais, utilizando uma abordagem de vocabulário aberto ([IBM, 2020a](#)).

O funcionamento do serviço consiste na conversão de um texto de entrada em *tokens* (palavras) para obter uma representação em um espaço n -dimensional. Então, a técnica GloVe (Seção 2.4) é utilizada para obter uma representação vetorial das palavras no texto. Em seguida, essa representação é apresentada a um algoritmo de Aprendizado de Máquina que infere um perfil de personalidade com os traços do modelo Cinco Grandes Fatores, o modelo de Necessidades ([ARMSTRONG et al., 2014](#)) e Valores ([SCHWARTZ, 2006](#)). Os cinco traços de personalidade retornados pelo serviço têm valores no intervalo entre 0 e 1. O *Personality Insights* consegue processar textos em Árabe, Inglês, Japonês, Coreano e Espanhol, e que contenham mais de 100 palavras.

Para treinar o algoritmo, o serviço utilizou pontuações de pesquisas que foram conduzidas entre milhares de usuários juntamente com dados de suas publicações do Twitter. A empresa IBM considera o erro absoluto médio (MAE) como critério de precisão do serviço, indicando que ele possui um valor médio de 0,12 (IBM, 2020a). O uso do *Watson Personality Insights* pode ser visto nos trabalhos de Junior e Inkpen (2017), Gallo et al. (2020) e Aguiar, Araújo e Costa (2020).

3 Metodologia

Este Capítulo descreve o método empregado para permitir a descoberta e análise de Traços de Personalidade em um ambiente de Participação Eletrônica. Alguns trabalhos encontrados na literatura apresentam estudos cujo objetivo é realizar a predição das personalidades dos usuários no contexto de redes sociais (LIMA; CASTRO, 2014; AZUCAR; MARENGO; SETTANNI, 2018; LIMAS; PRIMI; CARVALHO, 2018; MOTA; PAULA; DRUMMOND, 2021). Apesar das diferentes abordagens empregadas, esses trabalhos se baseiam na coleta automática de dados, principalmente os que estão em formato textual. Esse tipo de dado é capaz de fornecer informações muito ricas a respeito do comportamento humano e suas interações sociais (BARBIER; LIU, 2011). Considerando a contribuição da análise de linguagens naturais para a criação do modelo dos Cinco Grandes Fatores (apresentada na Seção 2.2), é possível aplicar técnicas computacionais capazes de prever a personalidade de um indivíduo através da análise das palavras utilizadas por ele para se expressar, como pode ser visto em Golbeck, Robles e Turner (2011), Quercia et al. (2011) e Sumner et al. (2012).

Neste trabalho, as etapas executadas para possibilitar essa análise podem ser definidas em dois processos abrangentes: geração do modelo, no qual os algoritmos de Aprendizado de Máquina são treinados e avaliados, e identificação dos traços, onde o modelo gerado é empregado para encontrar os traços, como apresentado na Figura 5. Cada etapa é descrita em detalhes da Seção 3.1 à Seção 3.4.

Em linhas gerais, foram coletados os textos de publicações da rede social Twitter para a criação de uma base de dados, que contém publicações de usuários que estão de alguma forma envolvidos com iniciativas de Participação Eletrônica. O conjunto de dados coletado foi rotulado utilizando o serviço *Watson Personality Insights*, que é capaz de retornar os traços de personalidade dos indivíduos com base no modelo dos Grandes Cinco Fatores. Foi gerada a representação vetorial dessas publicações, aplicando o método GloVe. Então, foram treinados modelos de regressão para realizar a tarefa de predição desses traços, assim como é feito pelo serviço de rotulagem. Posteriormente, esses modelos foram empregados para identificar os traços dos usuários na base de dados gerada a partir da coleta de publicações da ferramenta SoPa. Por fim, utilizando a estrutura do modelo dos Grandes Cinco Fatores, foi realizada a análise dos traços encontrados nesse ambiente.

No que diz respeito à implementação, foi empregada a linguagem de programação Python e, conseqüentemente, diversas bibliotecas da linguagem. A Tabela 2 apresenta de forma sumarizada cada biblioteca e seu propósito.

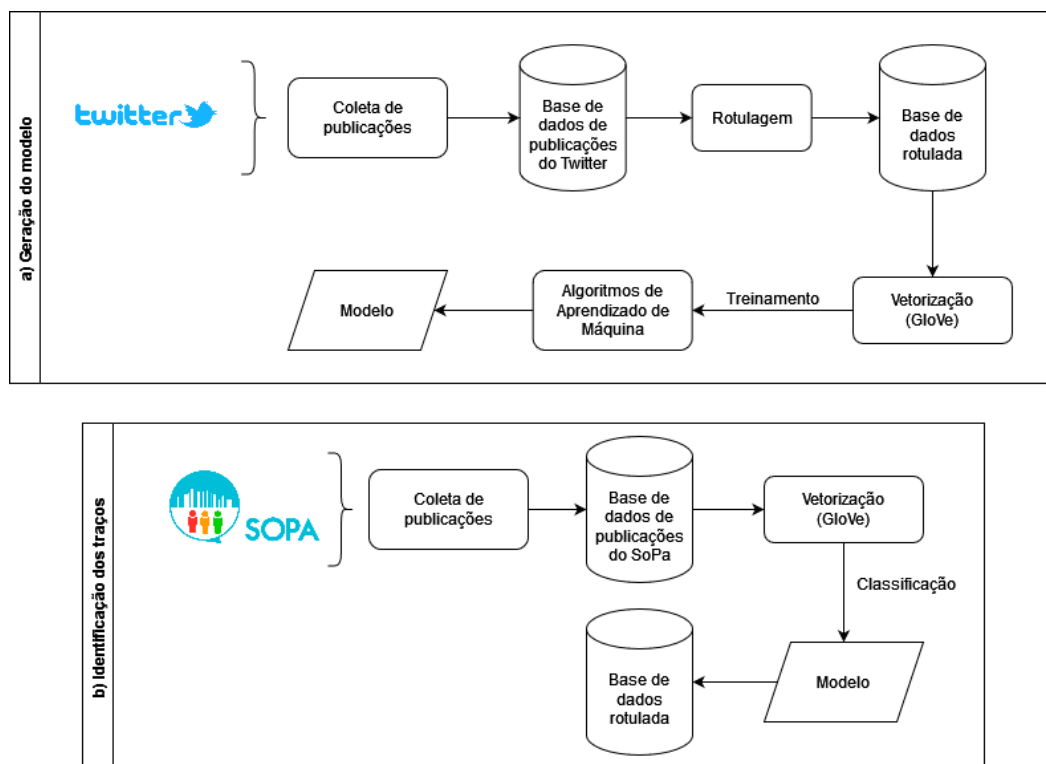


Figura 5 – Metodologia.

3.1 Coleta de publicações

A coleta de dados do Twitter foi feita utilizando a *API (Application Programming Interface)* da rede social. Essa *API* permite recuperar e analisar dados da rede, além de possibilitar a publicação de conteúdo. Dentre as informações possíveis de serem recuperadas estão os *tweets* (publicações), dados dos usuários, mensagens diretas, listas, tópicos em alta, mídias (vídeo e fotos) e localização. Para obter acesso aos recursos é necessário realizar uma solicitação à plataforma, informando a intenção ou propósito para uso das *APIs* do Twitter, quais dados serão coletados, detalhes sobre as análises que serão conduzidas, os métodos ou técnicas empregados para isso, além de informar se e como essas informações serão divulgadas.

A fim de utilizar os recursos disponibilizados pela *API* do Twitter, foi empregada a biblioteca Tweepy (ROESSLEIN, 2020). Essa biblioteca implementa métodos que são abstrações das requisições *HTTP (Hypertext Transfer Protocol)* aceitas pela *API*, facilitando o processo de desenvolvimento de programas que dependem desses recursos.

A abordagem para a coleta dos dados se baseou em encontrar usuários que estivessem relacionados com iniciativas de Participação Eletrônica ou Governo Eletrônico. Para isso, foram selecionados os usuários que são seguidores do perfil da ferramenta SoPa e do perfil de uma outra ferramenta de e-Participação mais conhecida comercialmente, chamada Colab (Colab, 2021). Nessa etapa, o método *followers_ids* da Tweepy foi utilizado

Tabela 2 – Bibliotecas empregadas no trabalho.

Biblioteca	Propósito	Métodos/Classes/Módulos
tweepy	Coleta de publicações do Twitter	<i>followers_ids()</i> <i>user_timeline()</i>
google.cloud	Utilização da API Translation do pacote de serviços do Google Cloud Translation	<i>translate()</i>
demoji	Tratar os emojis das publicações	<i>replace()</i>
ibm_watson	Utilização do serviço Watson Personality Insights	<i>PersonalityInsightsV3</i> <i>profile()</i>
NLTK	<i>Tokenização de textos</i>	<i>tokenize()</i>
gensim	Geração da representação vetorial das sentenças	<i>KeyedVectors()</i>
scikit-learn	Implementação dos algoritmos de Aprendizado de Máquina, métodos de amostragem e validação e métricas	<i>MLPRegressor</i> <i>KNeighborsRegressor</i> <i>DecisionTreeRegressor</i> <i>RandomForestRegressor</i> <i>MultiOutputRegressor</i> <i>train_test_split()</i> <i>cross_validate()</i> <i>mean_absolute_error()</i> <i>mean_squared_error()</i>
pickle	Persistir e carregar um modelo treinado	<i>dump()</i> <i>load()</i>

para recuperar os usuários seguidores dos dois perfis e o método *user_timeline* foi empregado para recuperar as publicações desses usuários. A API do Twitter permite recuperar 200 publicações de cada usuário e não é possível coletar dados de contas privadas. Ao final do processo, foram encontradas 459 contas de usuários com até 200 publicações em cada. A Tabela 3 traz um exemplo desse conjunto de dados.

Os usuários do SOPA, ao se cadastrarem, concordam com um termo de uso no qual estão descritos quais são os dados coletados pela ferramenta e quais são os usos desses dados¹. Sendo assim, as publicações dos usuários foram coletadas na base de dados da ferramenta, através de *scripts* desenvolvidos na linguagem SQL. No SoPa, um usuário pode publicar um problema para ser debatido e também escrever comentários sobre os problemas publicados por outros usuários. Sendo assim, os dados coletados são provenientes das duas tabelas do banco de dados que armazenam essas informações. Considerando que cada usuário tem um identificador único (id) na base de dados, esse foi o atributo utilizado para agregar todas as publicações de um determinado usuário. Por fim, obteve-se um conjunto de dados com 110 usuários. Um exemplo do conjunto de dados gerado é apresentado na Tabela 4.

¹ <<https://www.sopaneiru.com/privacidade>>

Tabela 3 – Exemplo do conjunto de dados do Twitter.

posts
<p>As Prefeituras precisam redobrar os cuidados com a limpeza das ambulâncias. Esses veículos passam o dia inteiro circulando pelas cidades, entrando nos hospitais, caso não haja limpeza diária, tornam-se focos de transmissão de doenças...</p> <p>NO MANGUE: O lixo continua sendo uma enorme ameaça ao manguezal do litoral norte de Pernambuco. Todos esses dejetos entre eles as garrafas de bebidas, copos plásticos e outros objetos são facilmente encontrados dentro dos mangues..</p>
<p>@metro_rio carro 6026 sentido Botafogo sem ar condicionado. Pessoas passando mal. O tratamento é desumano... @Ordem_Publica @MCrivella, mais uma vez, os ônibus parado de forma errada e nada é feito pela @Prefeitura_Rio esperando a @Ordem_Publica, mas simplesmente fecham a reclamação do @Central_1746, sem nem aparecer por lá, gostaria de saber pq não vão lá aos sábados quando reclamo dos ônibus da igreja no mesmo lugar? Ali tem um hidratante e um ponto de táxi...</p>
<p>A bicicleta sendo legal! Com ela: menos estresse, mais saúde, mais barato para o seu bolso e para o da coletividade...</p> <p>Começando o ano de 2019 por aqui com nosso grande pensador Zygmunt Bauman: interpretações do livro Confiança e Medo na Cidade. Investimentos públicos de infraestrutura em áreas urbanas valorizam propriedades privadas. É justo que esse valor agregado retorne ao poder público, para uso coletivo, ao invés de servir ao enriquecimento daqueles já privilegiados com a propriedade de imóveis.</p>

Tabela 4 – Exemplo do conjunto de dados do SoPa.

posts
<p>Rua na saída do SIND-UFLA, é uma rua com maior fluxo de carro e uma inclinação alta, acredito que a sinalização de uma placa pare poderia ser modificada para a rua perpendicular....</p>
<p>Criação de espaços públicos urbanos de qualidade, como por exemplo a Dique 1. Atualmente ela prioriza o fluxo de automóveis não dando a devida atenção ao o pedestre e ciclista. Para essa região seria necessário uma revitalização criando espaços seguros para caminha e atividades físicas ao ar livre...</p>
<p>Pressão de loteamentos próximos as duas unidades de conservação, horto e a reserva biológica. Estas duas áreas possuem grande relevância ambiental para o município e região, pois abriga grande diversidade da fauna e flora...</p>

O serviço *Watson Personality Insights* impõe uma limitação de que os textos estejam em Inglês para que seja possível retornar os traços de personalidade de um indivíduo. Portanto, antes de utilizar o serviço para rotular as bases geradas, é necessário realizar a tradução. Para isso, foi empregada a API Translation do pacote de serviços do Goo-

gle Cloud Translation. Essa API emprega a tradução automática neural do Google e dá suporte para textos em mais de cem idiomas (Google, 2021a). Uma desvantagem do uso dessa solução é que existe um valor cobrado pelas traduções quando o limite de 1.000.000 de caracteres é atingido. Para realizar a tradução, foi empregado o módulo *translate* da biblioteca *google.cloud* (Google, 2021b).

Por fim, foram removidos das publicações *links* e *emojis*. Os *emojis* são usados, geralmente, em mensagens eletrônicas e redes sociais e representam expressões faciais, objetos, lugares, animais e tipos de clima, através de uma linguagem não textual. Para isso, foram utilizadas expressões regulares e a biblioteca *demoji*. Ainda que a biblioteca permita a conversão dos *emojis* em um texto equivalente, essa abordagem não foi empregada neste trabalho.

3.2 Rotulagem

A etapa de rotulagem é responsável por atribuir os valores dos cinco traços de personalidade do modelo dos Cinco Grandes Fatores para cada amostra no conjunto de dados. Para isso, foi empregado o serviço *Watson Personality Insights*. Através desse serviço é possível obter a pontuação de cada traço informando um texto, como apresentado na Seção 2.5.

3.2.1 Rotulagem através do *Watson Personality Insights*

Neste trabalho, o uso do serviço para obter os traços de personalidade foi limitado por dois fatores: o idioma e a quantidade de palavras em cada publicação. O modelo do *Personality Insights* só consegue processar textos em Árabe, Inglês, Japonês, Coreano e Espanhol, e somente aqueles com mais de 100 palavras. Em relação aos dados do Twitter, as publicações coletadas atingiram mais de 100 palavras, porém parte delas foram publicadas em língua diferente do Inglês. Já para os dados do SoPa, uma parte inexpressiva do conjunto de dados atingiu mais de 100 palavras e todas as publicações estavam em Português. Sendo assim, somente a base de publicações do Twitter foi utilizada na etapa de rotulagem. Dessa forma, criou-se uma base rotulada, empregada no treinamento e geração de um modelo de regressão capaz de identificar os traços.

Para realizar a rotulagem foi empregada a biblioteca *ibm_watson*, o módulo *PersonalityInsightsV3* e o método *profile* para recuperação dos valores de cada traço de personalidade. Esses valores são retornados em uma escala contínua entre 0 e 1. Um exemplo do conjunto de dados do Twitter após a rotulagem é apresentado na Tabela 5.

Tabela 5 – Exemplo do conjunto de dados após rotulagem.

usuario	posts	A	R	E	C	N
94538417	0.79	0.64	0.55	0.73	0.50
12654438	0.83	0.63	0.55	0.65	0.39
30689304	0.77	0.64	0.58	0.70	0.56

3.2.2 O conjunto de dados rotulado

Os traços de personalidade, entendidos como as classes alvo de um problema de regressão, são definidos pelas letras A (Abertura à Experiência), R (Responsabilidade), E (Extroversão), C (Cordialidade) e N (Neuroticismo). A Figura 6 apresenta os gráficos *boxplot* para uma visão geral da distribuição dos valores dos traços de personalidade da base, obtidos através do serviço *Watson Personality Insights*. Para permitir a análise de cada traço individualmente, foram gerados os gráficos de histograma e *boxplot* referentes a cada um deles.

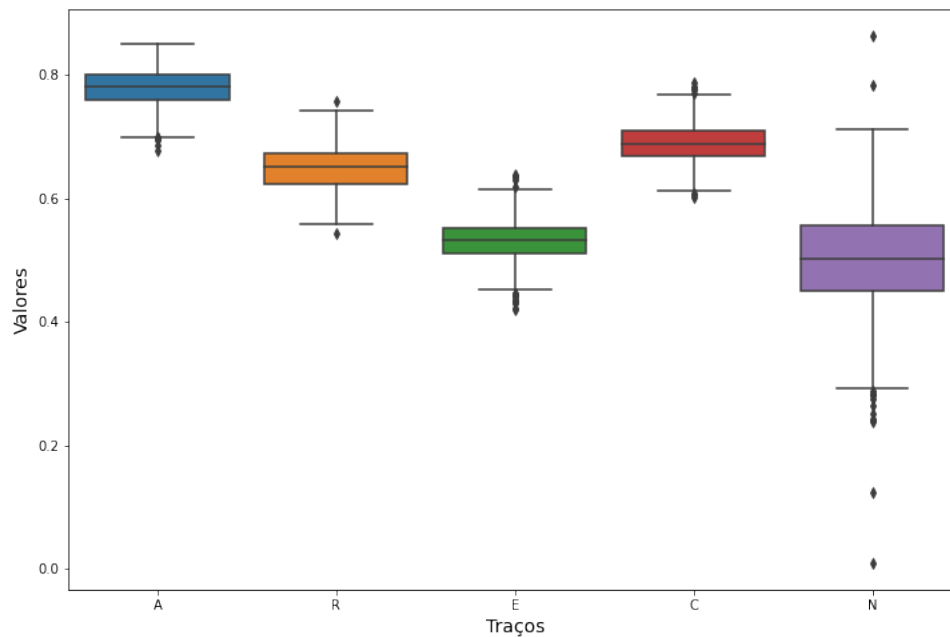


Figura 6 – Distribuição dos traços de personalidade no conjunto de dados do Twitter.

Com valores de média e mediana próximos (0,77 e 0,78, respectivamente), a classe A apresenta os maiores valores obtidos, indicando altas pontuações para o traço de Abertura à Experiência, como mostrado nos gráficos da Figura 7. Para 50% dos indivíduos as pontuações obtidas foram entre os valores 0,75 e 0,79. É possível notar também que os *outliers* dessa classe encontram-se abaixo do mínimo definido pelo quartil, mas nenhum deles fica abaixo de 0,67.

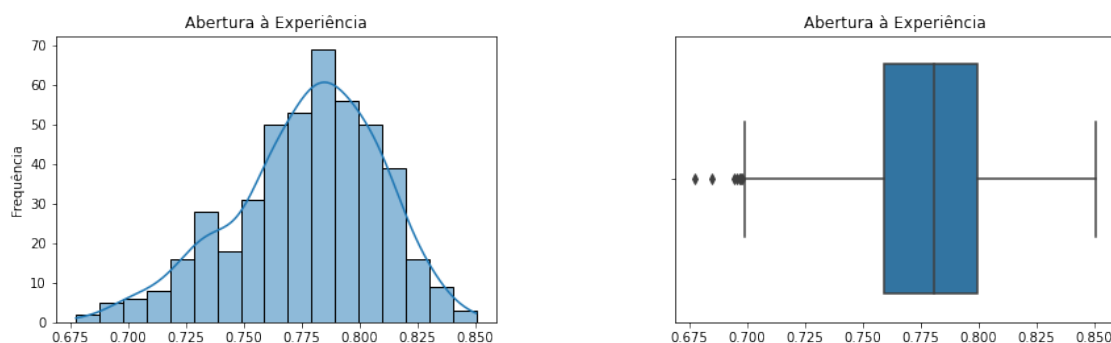


Figura 7 – Distribuição do traço Abertura à Experiência - Twitter.

Para a classe R foram encontrados os valores de 0,65 para a mediana e de 0,64 para a média, sendo que 50% dos indivíduos pontuaram entre 0,62 e 0,67. Não foram registradas pontuações abaixo de 0,50, o que possibilita afirmar que o traço se mostra forte entre os indivíduos desse conjunto de dados, ainda que possua valor médio abaixo dos encontrados para as classes A e C. A distribuição mostra-se quase simétrica, inclusive com relação aos *outliers*, como mostra a Figura 8.

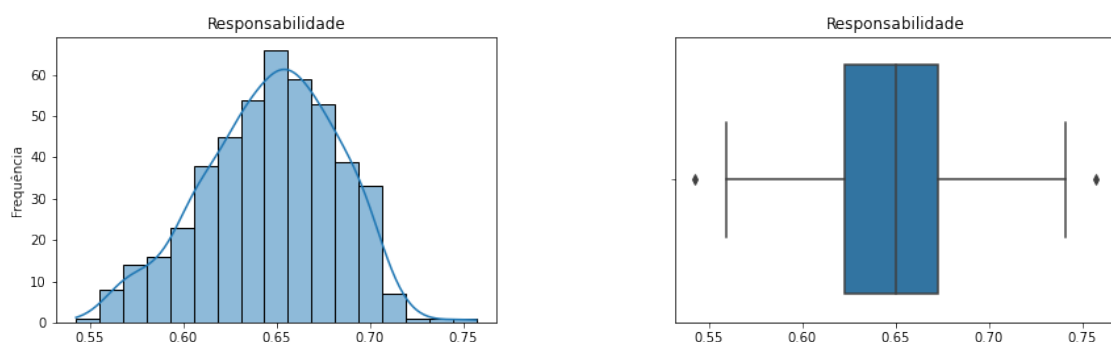


Figura 8 – Distribuição do traço Responsabilidade - Twitter.

Através da Figura 9, é possível notar que a classe E tem pontuação média de 0,53, que é idêntica à mediana, indicando uma distribuição simétrica. Ainda que identificado um valor máximo de 0,63, a distribuição mostra que 50% do conjunto está entre os valores 0,51 e 0,55, e que 25% dele se encontra abaixo de 0,51, indicando a presença de indivíduos com pontuação baixa para o traço.

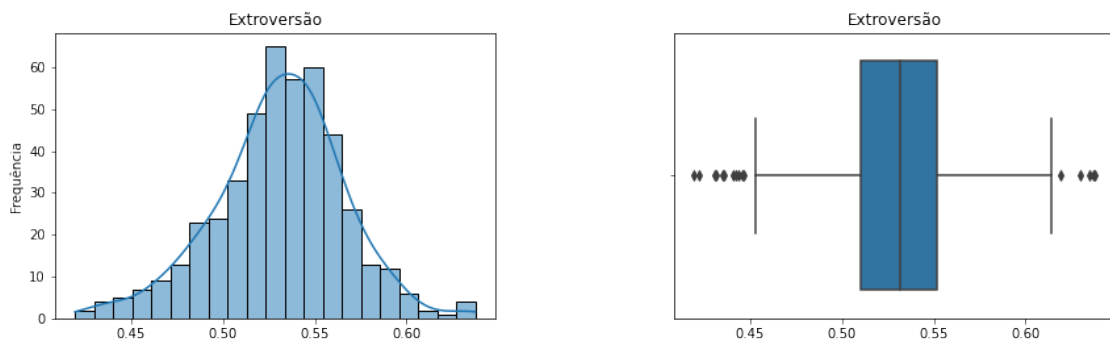


Figura 9 – Distribuição do traço Extroversão - Twitter.

A classe C possui valor médio de 0,69 e mediana de 0,68, contendo mais *outliers* acima do máximo do *boxplot*, como mostra a Figura 10. As pontuações obtidas por essa classe podem ser entendidas como altas, uma vez que nenhuma ocorrência encontra-se abaixo do valor 0,60, o valor máximo encontrado é de 0,78 e o intervalo de 0,66 a 0,70 representa 50% da distribuição.

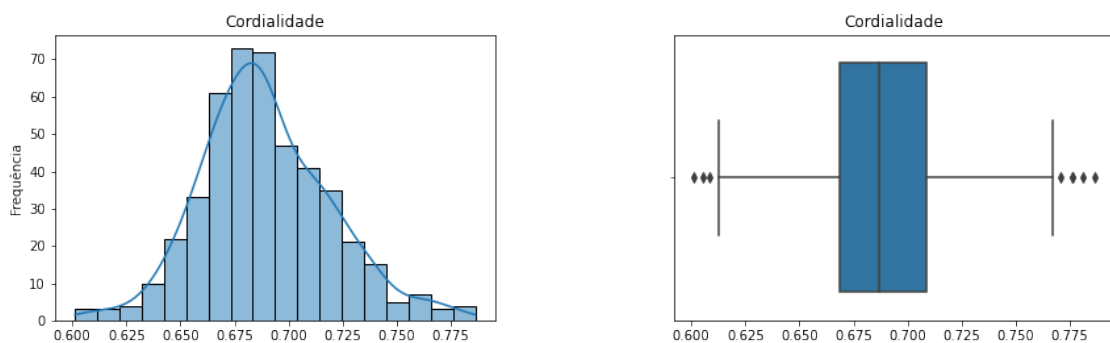


Figura 10 – Distribuição do traço Cordialidade - Twitter.

A classe N, apresentada na Figura 11, tem distribuição mais dispersa, com *outliers* distantes dos pontos mínimos e máximos, indicando uma alta variabilidade nos valores obtidos para o traço de Neuroticismo. Isso demonstra que, para esse conjunto de dados, existem tanto indivíduos que são estáveis emocionalmente (pontuação baixa), quanto indivíduos mais propensos às emoções negativas. Ainda assim, a análise dos quartis permite identificar que 50% dos indivíduos obtiveram pontuação entre 0,44 e 0,55.

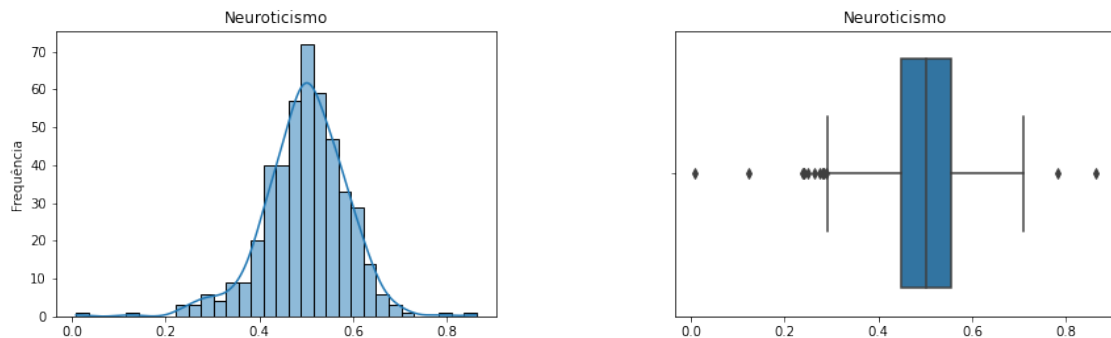


Figura 11 – Distribuição do traço Neuroticismo - Twitter.

3.3 Vetorização

A etapa de vetorização consiste na transformação das publicações em vetores equivalentes, utilizando o método GloVe. Como apresentado na Seção 2.4, esse método consiste de um dicionário de palavras que possuem uma representação vetorial. Essa estratégia é empregada internamente pelo serviço *Watson Personality Insights* para extração de atributos dos textos informados. Sendo assim, optou-se por empregar a mesma estratégia no processo de extração de atributos nos conjuntos de dados após a tradução. Para obter a representação de uma sentença em formato vetorial, foi empregada a técnica do valor médio, na qual o resultado obtido é a média dos vetores que representam individualmente as palavras da publicação.

O módulo *tokenize* da biblioteca *NLTK* (BIRD; KLEIN; LOPER, 2009) foi empregado para a etapa de *tokenização* das publicações. Já o módulo *KeyedVectors* da biblioteca *gensim* (ŘEHŮŘEK; SOJKA, 2010) foi utilizado para a geração da representação vetorial das publicações, baseando-se no dicionário pré-treinado de 6 bilhões de palavras do GloVe (PENNINGTON; SOCHER; MANNING, 2014). Ao fim dessa etapa, os textos das publicações foram transformados em vetores de 100 posições. É importante destacar que, tanto para a base de dados do Twitter, quanto para a base de dados do Sopa, não houve perda de palavras no processo de transformação, ou seja, foi possível encontrar uma representação vetorial para todas as palavras dos textos das bases. A Figura 12 apresenta uma sentença representada no formato vetorizado.

```

[-0.07265686  0.20927382  0.38044903 -0.23544374 -0.12299725  0.22936444
-0.1320063   0.18276326 -0.07632607 -0.03546263  0.16790645  0.02444898
 0.20900209  0.01227809  0.12679099 -0.25009942  0.15196443  0.16024108
-0.36628231  0.26510382  0.22777963  0.04840916  0.1219237  -0.0477
 0.1980648   0.02990444 -0.12934059 -0.43802974  0.1359292  -0.19079785
-0.10297208  0.46601087  0.01410181  0.04753721  0.0024511  0.27140212
-0.04495474  0.22446808  0.03546988 -0.1745626  -0.32244238 -0.19302505
 0.07394803 -0.36596373 -0.17768301  0.0019871  0.16619562 -0.27947202
-0.0479622  -0.68790549  0.01814309 -0.0493883  0.12156504  0.96011448
-0.25871822 -2.18410063  0.0186183  -0.0974308  1.37639868  0.43179139
-0.10368808  0.77449733 -0.31053978  0.00721369  0.65717155  0.04774292
 0.44639614  0.26967543  0.09824674 -0.14925072  0.01576478 -0.22172084
-0.04505716 -0.33590621  0.13752908  0.08587325 -0.06492657 -0.05107902
-0.70369637 -0.02743753  0.461853  -0.00237142 -0.44391763  0.10030984
-1.22967708 -0.1660274  0.01830303 -0.15473725 -0.1936543  -0.31204093
-0.06075453 -0.0817024  -0.03385366  0.07370418 -0.47663456 -0.04437754
-0.20964561 -0.28903788  0.40832505  0.24296118]

```

Figura 12 – Representação vetorial de uma sentença.

3.4 Regressores Multirrótulo

Os algoritmos de Aprendizado de Máquina utilizados nesta etapa foram regressores multirrótulo, uma vez que o intuito é prever os cinco traços de personalidade de um indivíduo e cada traço possui um valor no intervalo contínuo entre 0 e 1. Todos os modelos empregados são implementados pela biblioteca *scikit-learn* (PEDREGOSA et al., 2011).

O modelo Perceptron de Multicamadas (MLP) utilizado foi o *MLPRegressor*. Essa implementação faz o treinamento usando retropropagação sem função de ativação na camada de saída. Sendo assim, o erro quadrado é empregado como a função de perda e a saída é um conjunto de valores contínuos.

A implementação do KNN empregada foi o *KNeighborsRegressor*, na qual a média dos rótulos dos vizinhos mais próximos de uma entrada de dados define o rótulo dessa entrada. O valor de k utilizado foi 5, o padrão da biblioteca.

Para os modelos Árvore de Decisão e Floresta Aleatória, as implementações empregadas foram *DecisionTreeRegressor* e *RandomForestRegressor*, respectivamente. Ambas implementações empregam como critério de seleção de atributos (parâmetro *criterion*) o erro quadrático médio, que corresponde à redução da variância para medir a qualidade da divisão de um nó da árvore.

Como forma de permitir a regressão multirrótulos, foi empregada a estratégia *MultiOutputRegressor*, que consiste em treinar um regressor por classe. Dessa maneira, cada classe é representada por exatamente um regressor, sendo possível obter informações através de seu regressor correspondente.

3.4.1 Geração do modelo

Definidos os algoritmos empregados, foram treinados modelos com a base de dados do Twitter, como indicado na Seção 3.2. Na etapa de treinamento, a base foi dividida em duas partes, sendo 80% dos dados para treinamento e 20% para teste. Para isso, foi utilizada a função *train_test_split*.

Foram empregadas as métricas de validação MSE, RMSE e MAE, através das funções *mean_absolute_error* e *mean_squared_error*, sendo a última utilizada com o parâmetro *squared* configurado para falso, indicando o cálculo do RMSE. Os modelos foram validados utilizando validação cruzada com 10 pastas, através da função *cross_validate*.

3.4.2 Identificação dos traços

Após a etapa de treinamento, os modelos gerados foram persistidos através da biblioteca *pickle*. Posteriormente, o modelo com melhor desempenho foi aplicado para prever os traços de personalidade da base de dados do SoPa. Os resultados gerados na etapa de classificação foram analisados em função da distribuição dos valores dos traços encontrados. Essa análise é apresentada e discutida em detalhes no Capítulo 4.

4 Resultados e Discussão

Neste Capítulo são apresentados os resultados obtidos na etapa de treinamento e geração dos modelos, bem como a identificação e análise dos traços de personalidade na ferramenta SoPa. A partir das métricas utilizadas para avaliação dos regressores, aquele que apresentou melhor desempenho foi escolhido para a tarefa de rotulagem dos dados do SoPa, como apresentado na Seção 4.1. A Seção 4.2 aborda os resultados encontrados a partir da aplicação do modelo. Por fim, a Seção 4.3 apresenta a discussão dos resultados.

4.1 Geração dos Modelos

Nesta etapa, o conjunto de dados do Twitter, que foi rotulado através do serviço *Watson Personality Insights*, foi empregado para o treinamento e validação dos regressores escolhidos, utilizando as implementações dos algoritmos apresentados na Seção 3.4.

Como medida de desempenho e critério de comparação dos regressores, as métricas MSE, RMSE e MAE foram adotadas. Para essas métricas, quanto mais próximo do valor 0, melhor o modelo se comporta. A Tabela 6 apresenta o resultado de cada medida para os regressores.

Tabela 6 – Métricas dos regressores.

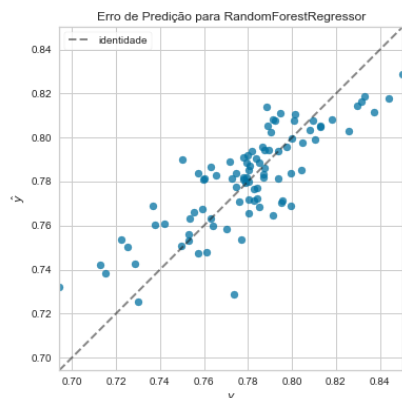
Regressor	MSE	RMSE	MAE
MLP	0,00530	0,07131	0,05386
KNN	0,00192	0,03701	0,02791
Árvore de Decisão	0,00335	0,04854	0,03690
Floresta Aleatória	0,00176	0,03504	0,02619

Quando analisadas em conjunto, essas métricas permitem indicar os modelos que sofrem grande penalização por gerar previsões muito distantes dos valores reais (MSE e RMSE) e possibilitam interpretar o valor do erro na mesma unidade dos valores do problema (MAE).

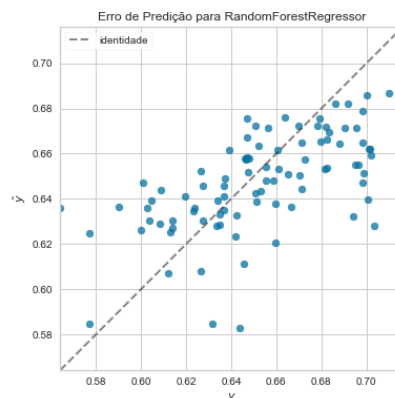
Conforme mostra a Tabela 6, o regressor com melhor desempenho foi a Floresta Aleatória. Isso pode ser justificado pelo fato do algoritmo se basear em uma estratégia de combinação de regressores homogêneos, o que, geralmente, produz resultados mais precisos (KUNCHEVA, 2014).

Através dos gráficos de erro de predição apresentados na Figura 13, é possível observar o comportamento do regressor Floresta Aleatória nos dados de teste. Nesses gráficos, o eixo rotulado como y representa o valor real da classe. O eixo \hat{y} representa o valor previsto pelo modelo. Um modelo sem erros geraria um gráfico onde todos os

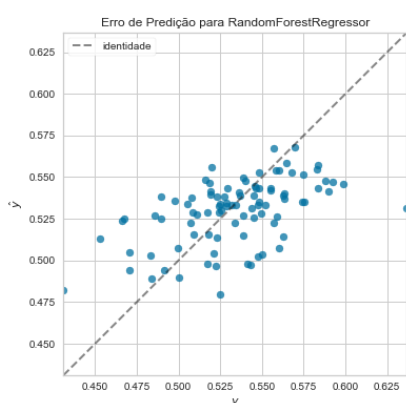
pontos estão sobre a reta identidade. Sendo assim, quanto mais próximo o ponto estiver da identidade, menor o erro do modelo para aquela previsão.



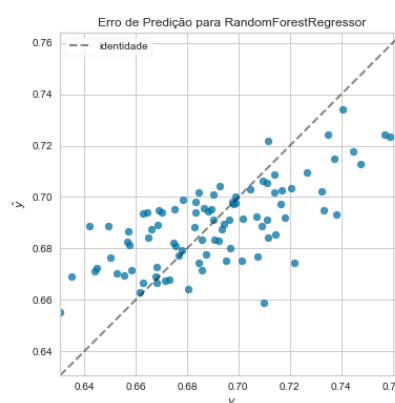
(a) *Abertura à Experiência*



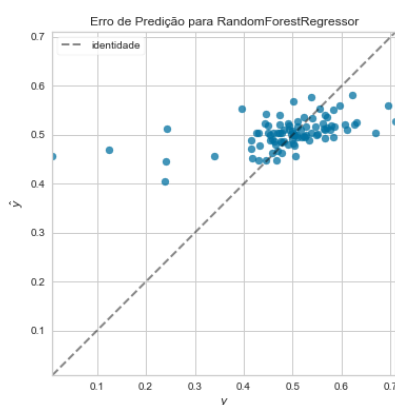
(b) *Responsabilidade*



(c) *Extroversão*



(d) *Cordialidade*



(e) *Neuroticismo*

Figura 13 – Erros de predição do regressor Floresta Aleatória.

A dispersão dos pontos nos gráficos mostra que, apesar do erro existir em todas as classes, os valores tendem a estar distribuídos próximos a identidade, com destaque para a classe A. Os resultados obtidos são condizentes com aqueles encontrados nos trabalhos de [Farnadi et al. \(2016\)](#), [Skowron et al. \(2016\)](#) e [Golbeck, Robles e Turner \(2011\)](#).

4.2 Identificação dos traços do SoPa

Considerando os resultados apresentados na Seção 4.1, o modelo produzido pelo algoritmo Floresta Aleatória foi empregado na tarefa de rotulagem do conjunto de dados do SoPa. A distribuição dos valores gerados para cada traço de personalidade é apresentada no *boxplot* da Figura 14.

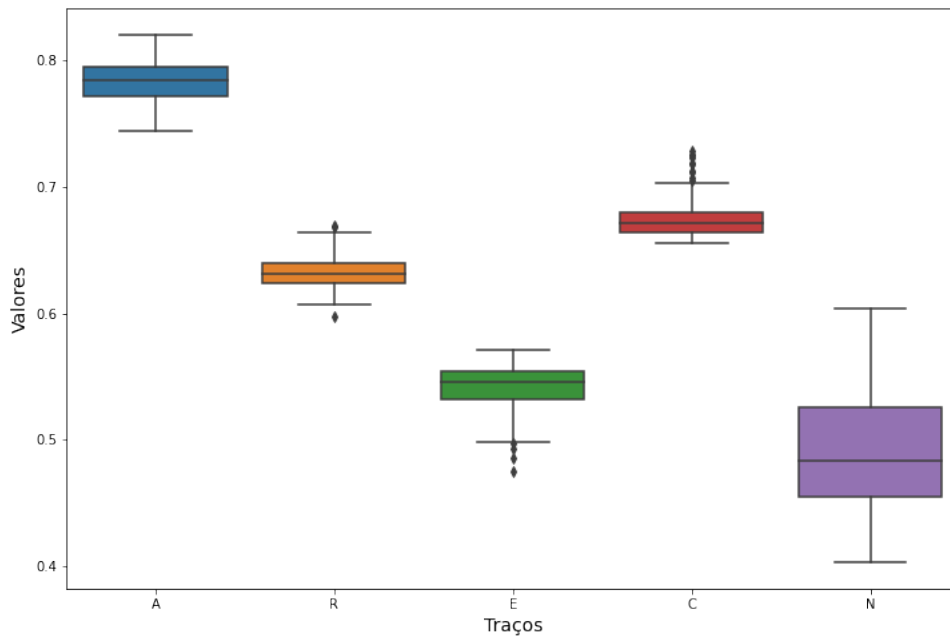


Figura 14 – Distribuição dos traços de personalidade no conjunto de dados do SoPa.

Como apresentado na Figura 15, o traço de Abertura à Experiência possui 50% de sua distribuição localizada no intervalo de valores entre 0,77 e 0,79, um valor máximo de 0,82 e mínimo de 0,74, representando a classe com as pontuações mais altas. Dessa forma, é possível afirmar que os usuários do SoPa tendem a ter características mais proeminentes para esse traço, sendo classificados como indivíduos mais criativos, com interesses amplos, intrigado por novas ideias e como explorá-las.

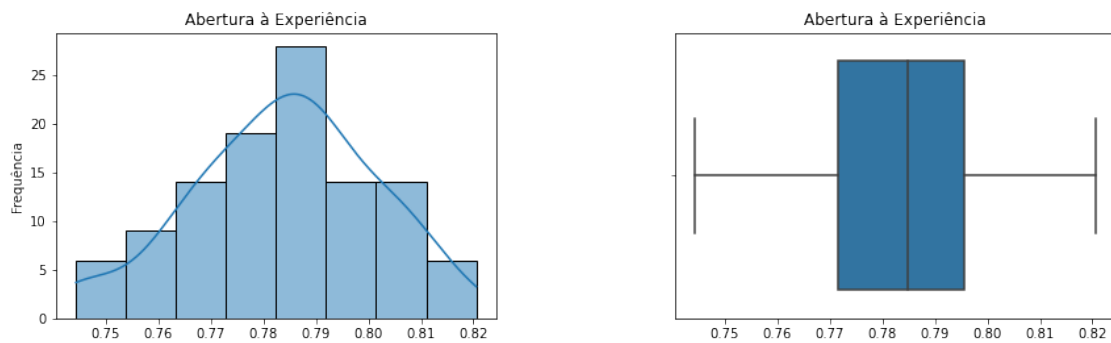


Figura 15 – Distribuição do traço Abertura à Experiência - SoPa.

O traço Responsabilidade possui uma distribuição de valores no intervalo de 0,59 a 0,67, sendo que 50% dos indivíduos receberam pontuações no intervalo de 0,62 a 0,64, conforme mostrado pela Figura 16. É possível verificar que os usuários da ferramenta obtiveram pontuações acima de 0,50, porém não altas o suficiente para permitir afirmar que o traço se destaca. Esses usuários, podem ser definidos como pessoas organizadas, cautelosas e auto-disciplinadas.

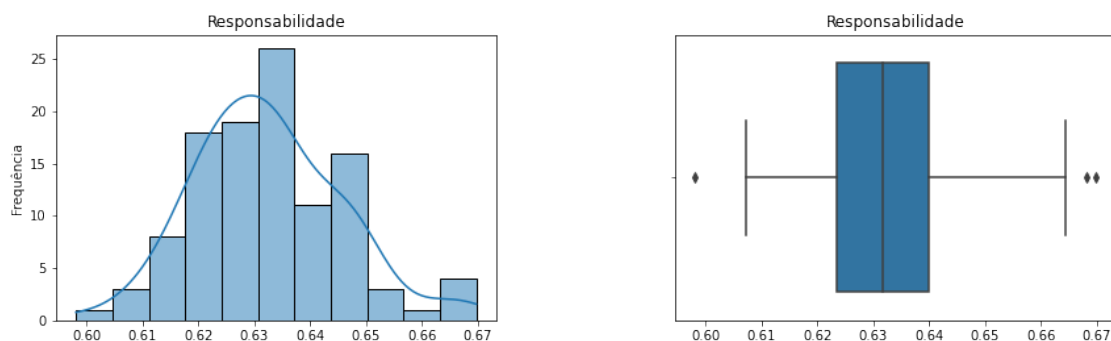


Figura 16 – Distribuição do traço Responsabilidade - SoPa.

Com uma distribuição assimétrica a direita, para o traço Extroversão foram obtidos valores no intervalo entre 0,47 e 0,57, 50% dos indivíduos receberam uma pontuação entre 0,53 e 0,54 e os *outliers* encontram-se todos abaixo do mínimo, como é apresentado na Figura 17. A maior parte dos usuários pontuou acima do valor 0,50, o que indica a presença do traço em menor intensidade, indicando pessoas sociáveis e comunicativas. Ainda que em menor quantidade, os usuários que pontuaram abaixo de 0,50 representam indivíduos mais reservados e sérios.

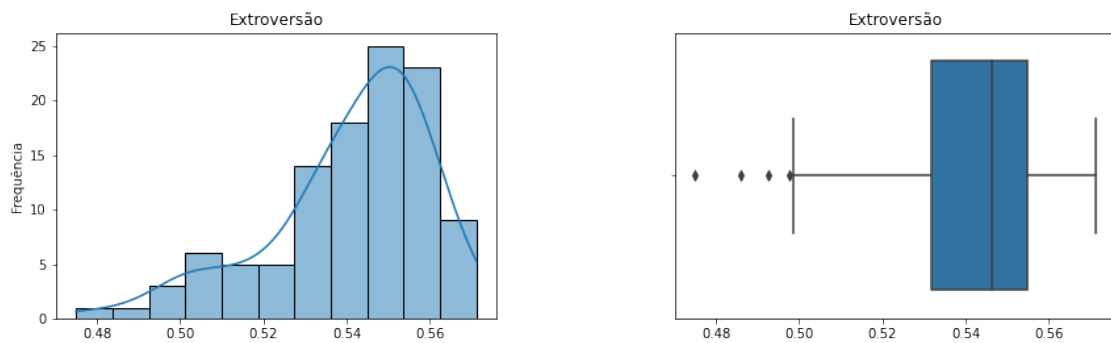


Figura 17 – Distribuição do traço Extroversão - SoPa.

Os resultados para o traço Cordialidade, apresentados na Figura 18, indicam uma distribuição assimétrica a esquerda, no intervalo de 0,65 a 0,72, com muitos *outliers* acima do limite superior. É possível notar que 50% da distribuição está localizada no intervalo de valores de 0,66 a 0,68. Para esse traço, os usuários do SoPa podem ser caracterizados como pessoas altruístas, empáticas e confiáveis.

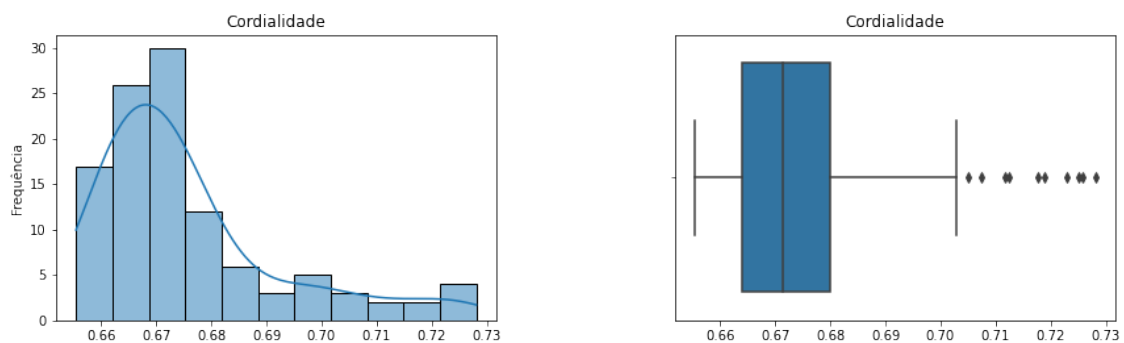


Figura 18 – Distribuição do traço Cordialidade - SoPa.

Por fim, o traço Neuroticismo apresenta uma distribuição no intervalo de valores de 0,40 a 0,60, sendo que 50% dos indivíduos receberam pontuações entre 0,47 e 0,52, conforme mostra a Figura 19. Para esse traço, os usuários com pontuações abaixo de 0,50 podem ser considerados indivíduos confiantes e controlados. Por outro lado, as pontuações acima de 0,50 indicam pessoas com tendência a demonstrarem melancolia e preocupação. Vale destacar que as pontuações não foram altas o suficiente para considerar o traço proeminente.

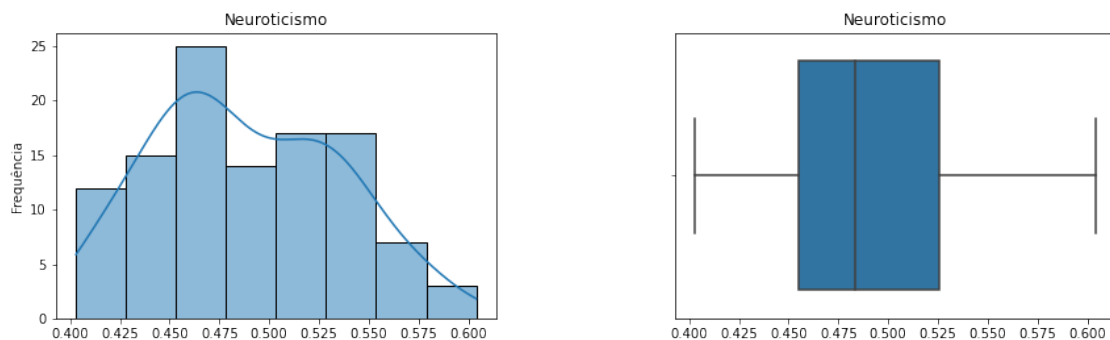


Figura 19 – Distribuição do traço Neuroticismo - SoPa.

A fim de possibilitar uma análise de como as classes se relacionam, foram estabelecidas faixas de intervalos de valores, como apresentado na Tabela 7. A faixa 1 compreende o intervalo de valores abaixo de 0,49, representando as pontuações que definem a presença fraca do traço de personalidade. A faixa 2 foi definida como o intervalo de valores entre 0,50 e 0,74, podendo ser entendida como as pontuações que caracterizam uma tendência mais alta do que a média para um traço. Já a faixa 3, definida como o intervalo de valores entre 0,75 e 1, corresponde às pontuações que caracterizam a presença muito forte do traço de personalidade, possibilitando discernir de forma clara e fácil as características que o definem.

Tabela 7 – Faixas definidas.

Faixas	Intervalo	Significado
1	0 , 0,49	Traço fraco
2	0,50 , 0,74	Traço acima da média
3	0,75 , 1,0	Traço facilmente discernível

Para realizar a análise das faixas encontradas, foi empregada a técnica de visualização do diagrama de cordas (*chord diagram*). Segundo [Koochaksaraei et al. \(2017\)](#), o diagrama de cordas é uma forma de ilustrar os relacionamentos entre elementos do conjunto de dados, no qual cada elemento é um nó e a relação entre eles é representado por uma aresta. O diagrama recebe esse nome uma vez que a coleção das arestas dentro da ilustração se assemelha à cordas. A Figura 20 apresenta a visualização gerada com os dados do SoPa.

A documentação do serviço *Watson Personality Insights*, apresenta uma forma de interpretar as relações entre os traços, descrevendo as combinações de características da personalidade ([IBM, 2020b](#)). Essa interpretação foi utilizada para analisar as relações entre as faixas obtidas.

Através da análise do gráfico é possível notar que foram obtidas 8 faixas. Isso significa que para alguns traços de personalidade as pontuações obtidas destacam-se em

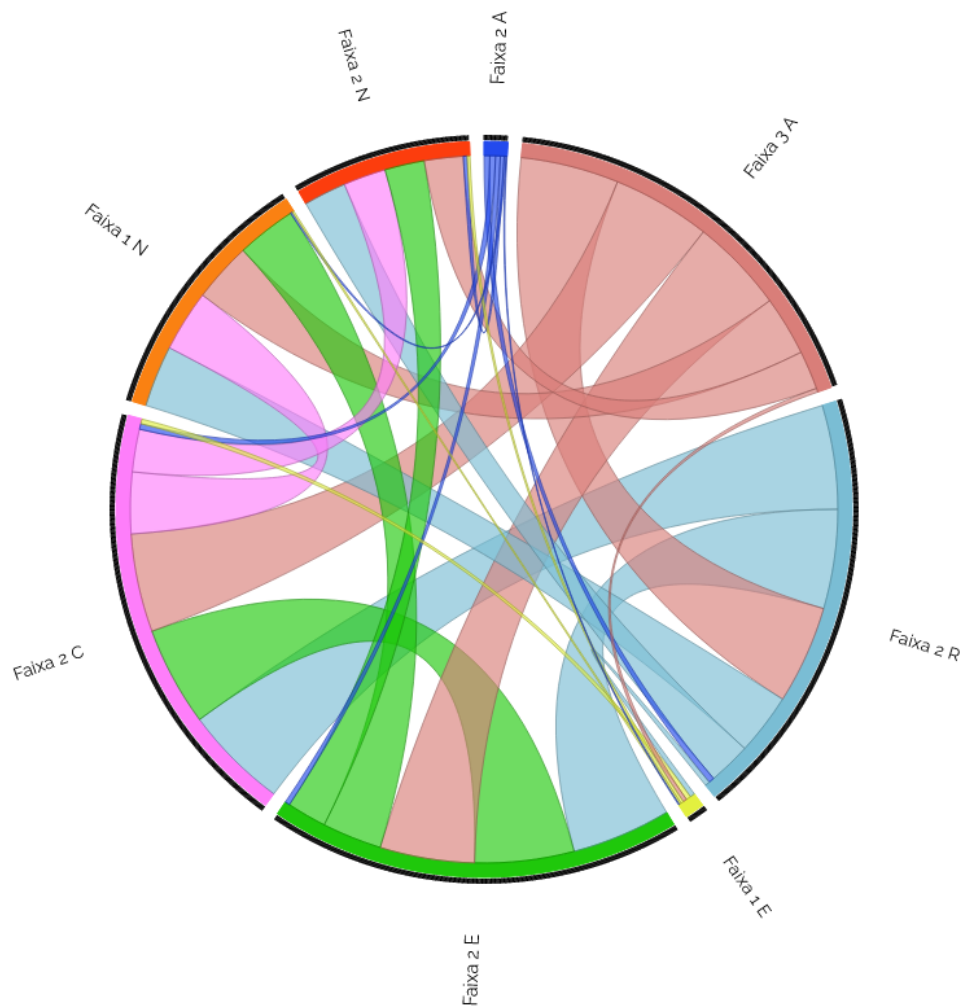


Figura 20 – Diagrama de cordas das faixas de personalidade do SoPa.

um intervalo de valores.

Conforme apresentado na Figura 21, é possível observar que para o traço de Abertura à Experiência as pontuações estão nas faixas 2 e 3, sendo inclusive o único traço que se encontra na terceira faixa. Assim, é possível afirmar que os indivíduos do conjunto possuem a presença do traço de Abertura à Experiência acima da média e muito forte, sendo a maior parte pertencente à faixa de valores que permite discernir o traço facilmente.

A faixa 3 se relaciona com as faixas 2 de Responsabilidade, Cordialidade e Extroversão, indicando indivíduos idealistas, diplomáticos, analíticos, perceptivos, eloquentes e inquisitivos. Em relação ao traço de Extroversão, o traço de Abertura se relaciona com a faixa 1 em menor quantidade, evidenciando a existência de indivíduos introspectivos e contempladores. Por fim, o relacionamento com o traço de Neuroticismo reflete indivíduos criativos, intelectuais, e inventivos.

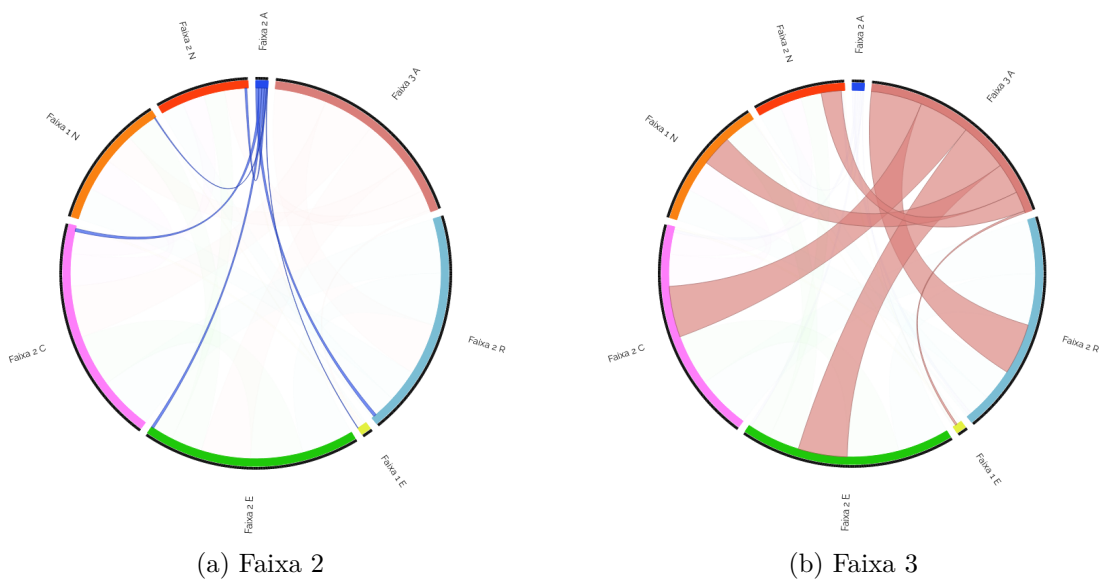


Figura 21 – Diagramas de corda para o traço A.

O traço de Responsabilidade apresentou valores apenas na faixa 2, o que indica a existência de indivíduos organizados e cuidadosos. Através das conexões mostradas na Figura 22, nota-se que o traço está relacionado com todas as demais faixas dos outros traços, sendo em maior quantidade com as faixas 3 de A, 2 de E e 2 de C. Isso indica que esses indivíduos tendem a ser perfeccionistas, determinados, competitivos, educados e atenciosos. Em relação ao traço N, as ocorrências indicam pessoas consistentes, auto-disciplinadas e lógicas. Vale destacar a relação com a faixa 1 do traço E que, ainda que em menor quantidade, permite identificar indivíduos cautelosos e pontuais.

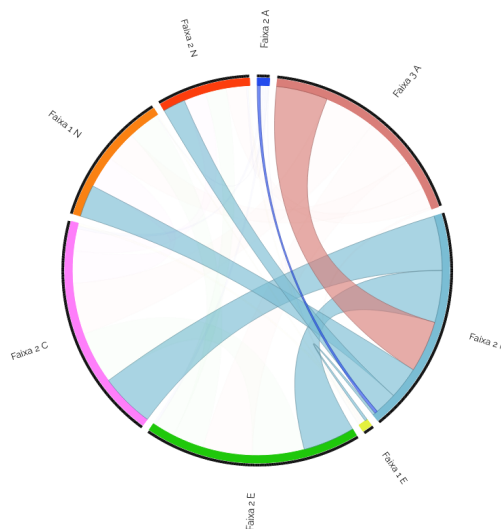


Figura 22 – Diagrama de corda para o traço R.

Analisando o traço de Extroversão, é possível notar que grande parte dos indivíduos do conjunto apresentam uma tendência à socialização, enquanto uma porção menor pode

ser entendida como mais reservados, uma vez que as pontuações para o traço se encontram nas faixas 1 e 2, como mostrado na Figura 23.

A maior parte dos indivíduos desse traço está na faixa 2, que se relaciona em maior quantidade com as faixas 3 de A, 2 de R e 3 de C, indicando que se tratam de pessoas comunicativas, espontâneas e persistentes. A relação com a faixa 1 do traço N permite identificar a presença de indivíduos confiantes e desinibidos, ao passo que as conexões em menor quantidade com a faixa 2, caracterizam pessoas com tendências mais explosivas e prolixas. A faixa 1 da classe E, que identifica a presença fraca do traço, ainda que em menor quantidade, permite identificar indivíduos introspectivos, cautelosos, íntegros, humildes, reservados e imparciais.

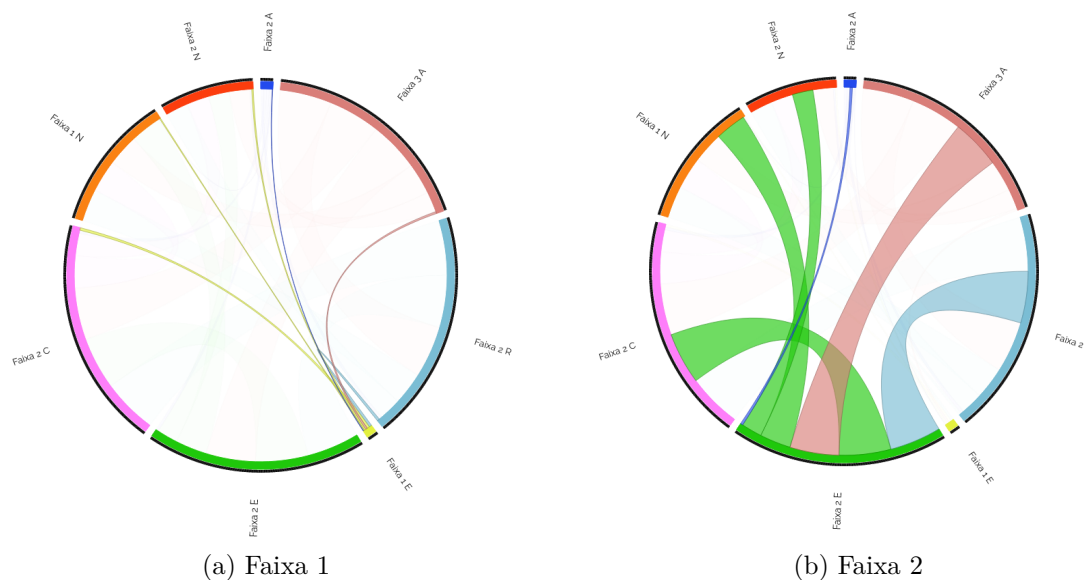


Figura 23 – Diagramas de corda para o traço E.

Contido unicamente na faixa 2, o traço de Cordialidade possui valores que permitem identificar indivíduos com tendência a serem mais compassivos e cooperativos. Conforme apresentado na Figura 24, as maiores relações do traço são com as faixas 3 de A, 2 de R e 2 de E, já discutidas anteriormente. A maior proporção da relação com a classe N está na faixa 1, caracterizando a presença de indivíduos generosos e tolerantes. Já a relação com a faixa 2 de N, indica pessoas sentimentais e afetuosas.

Por fim, o traço de Neuroticismo aparece em duas faixas, 1 e 2, indicando que entre os indivíduos existem aqueles que são emocionalmente estáveis (pontuações baixas) e aqueles que apresentam tendência a sentir emoções negativas (pontuações altas). A faixa 1 ocorre em maior quantidade e se relaciona com todas as demais faixas de outras classes, indicando pessoas sinceras, criativas, racionais, pacientes e realistas. Os relacionamentos da faixa 2, por sua vez, caracterizam indivíduos reservados, tensos, prolixos, explosivos e emotivos. A Figura 25 apresenta os diagramas do traço N.

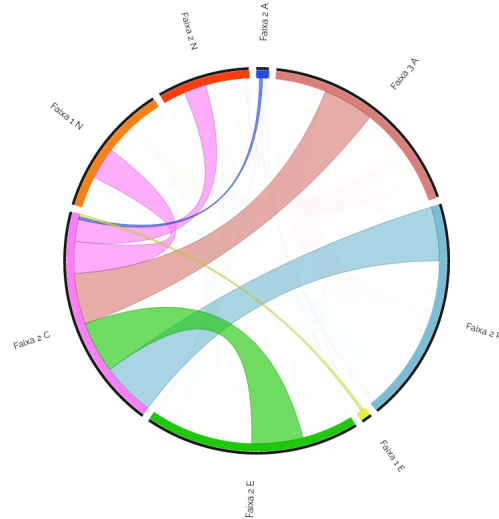
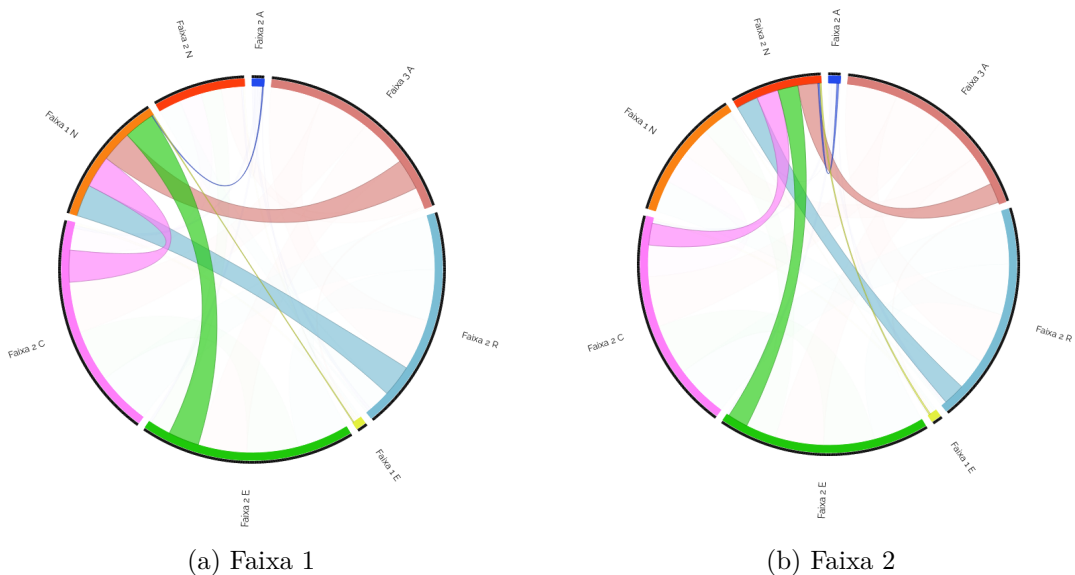


Figura 24 – Diagrama de corda para o traço C.



(a) Faixa 1

(b) Faixa 2

Figura 25 – Diagramas de corda para o traço N.

4.3 Discussão dos Resultados

Nesta Seção serão analisados os traços identificados no contexto da ferramenta SoPa e como eles podem estar relacionados com a área da Participação Eletrônica. Para isso, são levadas em consideração as características que descrevem cada traço segundo o modelo de Cinco Grandes Fatores de personalidade e evidências já encontradas na literatura.

O traço com as maiores pontuações no contexto do SoPa foi o de Abertura à Experiência. Esse traço descreve o quanto uma pessoa está interessada em novas ideias, novas experiências e estímulos intelectuais. Além disso, pontuações altas para o traço indicam pessoas que não se sentem ofendidas ao lidar com opiniões contrárias às suas, e que

sentem inclinação para se engajar em tarefas menos convencionais, simplesmente porque tais tarefas representam ideias não convencionais (GERBER et al., 2011). Considerando o fato de que o SoPa é uma ferramenta que se vale das TICs e que emprega estratégias de rede social, no intuito de promover a interação entre usuários para o debate e resolução de problemas sociais, é justificável a forte presença desse traço entre os indivíduos desse estudo.

Considerado por alguns autores como um traço cuja relação com a participação não é claramente definida (HA; KIM; JO, 2013; WEINSCHENK, 2017), o traço de Responsabilidade descreve pessoas organizadas, obedientes, capazes de seguir normas e determinadas em alcançar realizações pessoais. A pontuação obtida pelos usuários do SoPa para esse traço pode ser considerada alta, ainda que não suficiente para afirmar que é um traço de destaque. Isso pode ser justificado pelo fato de que os dados coletados nesse estudo correspondem ao período em que a ferramenta estava sendo utilizada nos seguintes contextos:

- Revisão do Plano Diretor de Pouso Alegre - MG (2018 – 2021)
- Elaboração do Plano de Meio Ambiente de Pouso Alegre - MG (2020 – 2021)
- Elaboração do Plano de Resíduos Sólidos de Pouso Alegre - MG (2020 – 2021)
- Revisão do Plano Diretor de Elói Mendes - MG (2020 – 2021)
- Elaboração do Plano de Saneamento de Elói Mendes - MG (2020 – 2021)
- Elaboração do Plano de Mobilidade Urbana de Elói Mendes - MG (2020 – 2021)

O uso da ferramenta para essa finalidade pode implicar em indivíduos que participaram ativamente por considerar a tarefa um dever cívico, algo que o modelo CGF descreve como característica do traço de Responsabilidade. Além disso, o trabalho de Mondak e Halperin (2008) estabelece uma relação positiva entre o traço de Responsabilidade e a tendência dos cidadãos em comparecer às reuniões do governo local, um resultado que pode corroborar com a pontuação obtida pelos indivíduos desse estudo, uma vez que o SoPa foi empregado como ferramenta de apoio nos processos de revisão e elaboração de diversos Planos de Desenvolvimento Urbano, que ocorreram no contexto da pandemia do novo coronavírus de 2020, no qual encontros presenciais não eram permitidos.

A Extroversão define pessoas que são assertivas, sociáveis, confiantes e enérgicas e que, portanto, têm maior probabilidade de comparecer a festas e alcançar cargos de liderança (MCCRAE; COSTA, 1997). Os resultados obtidos demonstram que o SoPa possui indivíduos cujo traço de Extroversão não apresenta valores tão altos, indicando usuários que podem ser considerados mais reservados e sérios. Esse fato pode ser justificado

pela relação entre as pontuações obtidas pelas outras classes, uma vez que pontuações altas para o traço de Extroversão combinadas às pontuações altas dos outros traços, tende a indicar pessoas dramáticas, explosivas, prolixas, extravagantes e voláteis. Ser “o centro das atenções” não é algo esperado em um ambiente que tem como objetivo a colaboração entre indivíduos para debater e solucionar problemas.

Segundo [Mondak \(2010\)](#), pessoas extrovertidas têm a tendência de se engajar mais em tarefas de participação que envolvam a interação social, nas quais suas ideias sejam compartilhadas e tenham visibilidade em um ambiente público. Sendo assim, as pontuações identificadas no SoPa podem ser justificadas pelo fato de que o ambiente proporciona um tipo de interação virtual, o que pode levar indivíduos muito extrovertidos a considerar que suas opiniões não são ouvidas ou vistas. Outro fator que pode ser levado em consideração é que o SoPa se trata de um ambiente virtual ainda de pouco alcance. Talvez, em uma rede social popular, essas pessoas estariam mais propensas a opinar.

O traço de Cordialidade descreve a tendência da preocupação pró-social e comunitária para com o outro. Entre os usuários do SoPa, as pontuações obtidas permitem afirmar que o traço tem presença significativa, indicando indivíduos altruístas, confiáveis e modestos. Ainda que visto como um traço que busca evitar conflitos ([BLAIS; ST-VINCENT, 2011](#)), a Cordialidade faz com que os indivíduos tenham a tendência de serem cooperativos, confiáveis e prestativos, o que surte um efeito positivo quando se trata de atividades que visam o bem estar comunitário ([ROCCAS et al., 2002](#)). Sendo assim, as pontuações obtidas pelos usuários do SoPa para esse traço podem ser justificadas levando em consideração o propósito da ferramenta. Seu uso visa o debate e a proposta de solução de problemas que afetam diretamente uma comunidade, seja de forma espontânea ou motivados por uma entidade governamental. O senso de colaboração para alcançar o bem estar comunitário pode ser uma das motivações dos usuários.

O Neuroticismo caracteriza a estabilidade emocional de um indivíduo. Nas pontuações obtidas pelos usuários do SoPa, é possível notar a presença de valores que denotam tanto a presença fraca do traço, quanto a forte, ainda que não tão expressiva. Aquelas pessoas com o traço fraco tendem a ser autoconfiantes, calmas e não se aborrecem facilmente. Por outro lado, os indivíduos com traço forte podem sofrer de ansiedade, depressão e outras emoções negativas.

Pessoas com alta estabilidade emocional, ou seja, com pontuação baixa para o traço Neuroticismo, tendem a ser mais confiantes e, portanto, mais dispostas a defenderem suas opiniões e a lidar melhor com a ansiedade, o estresse e os conflitos que essa exposição pode gerar ([GERBER et al., 2011](#)). Esse fato pode justificar a presença mais fraca do traço entre os usuários do SoPa, uma vez que na ferramenta é preciso apresentar sua ideia aos outros usuários e estar disposto a debatê-la.

No que diz respeito às pontuações mais altas para o Neuroticismo, [Mondak \(2010\)](#)

argumenta que os indivíduos que participam se comportam como hipocondríacos. Assim como os hipocondríacos tendem a se preocupar exageradamente com sua saúde, as pessoas com baixa estabilidade emocional tendem a participar para expressar sua preocupação excessiva com problemas políticos ou sociais. Nos casos em que os usuários do SoPa pontuaram para uma presença do traço Neuroticismo, talvez não seja adequado considerar que são indivíduos exageradamente preocupados com problemas, uma vez que os valores não são altos. Segundo [Gerber et al. \(2011\)](#), a identificação de uma correlação para pontuações altas desse traço e a participação se mostrou inconclusiva.

Dentre os trabalhos que buscam encontrar uma relação entre os traços de personalidade e a participação, o estudo conduzido por [Weinschenk \(2017\)](#) inclui o Brasil e países da América Latina na análise dessas relações. O autor definiu sete atos políticos a serem analisados, entre eles a frequência em que o indivíduo tentou ajudar a resolver um problema em sua comunidade ou bairro, tarefa que pode ser realizada utilizando o SoPa como meio para que o ato político aconteça. Considerando os resultados obtidos para o Brasil, o autor observou que os traços de Abertura à Experiência e Responsabilidade tem uma relação positiva com o ato de resolver um problema local. Para os traços de Extroversão e Cordialidade a relação foi neutra. Já para o traço de Neuroticismo, a relação foi negativa. Essa análise vai ao encontro dos resultados obtidos neste trabalho, em especial no que se refere aos traços de Abertura à Experiência e Responsabilidade.

5 Conclusões

Esta dissertação de mestrado abordou a análise de traços de personalidade em ambientes de Participação Eletrônica. A identificação desses traços permite descrever as características proeminentes dos indivíduos nesses ambientes, bem como teorizar sobre as motivações que levam à participação. Para isso, foi apresentado um método de reconhecimento automático de traços de personalidade, empregando técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina.

Inicialmente, foi elaborada uma base de dados com publicações da rede social Twitter, considerando uma amostra de usuários que estavam envolvidos com a questão da Participação. Para obter a pontuação dos traços de personalidade definidos no modelo dos Cinco Grandes Fatores, foi utilizado o serviço da IBM *Watson Personality Insights*. Com os rótulos obtidos nessa etapa, a técnica de representação vetorial GloVe foi empregada para permitir que os textos fossem utilizados no treinamento dos regressores MLP, KNN, Árvore de Decisão e Floresta Aleatória.

Os modelos foram avaliados através das métricas MSE, RMSE e MAE, sendo o algoritmo Floresta Aleatória o de melhor desempenho, obtendo os valores de 0,00176, 0,03504 e 0,02619 para cada métrica, respectivamente. Esse modelo foi aplicado na tarefa de identificação dos traços dos usuários da ferramenta SoPa.

Através dos resultados obtidos foi possível notar a presença mais proeminente do traço de Abertura à Experiência. Os traços de Responsabilidade e Cordialidade podem ser considerados traços fortes, ainda que com pontuações menores em relação à Abertura à Experiência. Extroversão e Neuroticismo são os traços com as menores pontuações.

Esses resultados indicam que os dados analisados nesse trabalho caracterizam, de forma geral, os usuários do SoPa como indivíduos analíticos, perceptivos, articulados, diplomáticos, criativos e introspectivos. Quando analisados em função do propósito da ferramenta, a presença desses traços permite refletir sobre as motivações que levam um indivíduo a se engajar em ambientes de Participação Eletrônica sob um ponto de vista psicológico.

Para concluir essa dissertação, são listadas nas Seções 5.1 e 5.2 as contribuições e limitações desse trabalho, respectivamente, bem como trabalhos futuros que podem ser desenvolvidos.

5.1 Contribuições

A principal contribuição desse trabalho é a descrição de um método para reconhecimento automático e análise de traços de personalidade em ambientes de Participação Eletrônica.

No que diz respeito a área da Inteligência Artificial, o trabalho traz contribuições em relação à aplicação de algoritmos de Aprendizado de Máquina em tarefas de regressão multirrótulo para as quais o objetivo final possui um grau mais elevado de subjetividade, dada a proposta de identificação de traços de personalidade por meio desses modelos.

Com relação à pesquisa em Participação Eletrônica, a principal contribuição desse trabalho reside na investigação de uma abordagem para identificar e descrever aspectos cognitivos e comportamentais dos indivíduos envolvidos em ambientes de e-Participação sob a ótica das estruturas de personalidade. Isso possibilita que, combinado às abordagens provenientes de áreas como *marketing* e Fatores Humanos em Sistemas Computacionais (IHC), possam ser definidas estratégias no intuito de aumentar o engajamento cidadão, problemática apresentada na Seção 2.1.

Além disso, pode-se destacar que as análises desse trabalho foram realizadas com dados coletados em um contexto de utilização real da ferramenta SoPa, sem que houvesse incentivo direcionado para participar desse estudo.

A pesquisa sobre o reconhecimento automático de personalidades, necessária para a elaboração dessa dissertação, resultou na apresentação e publicação de um artigo na *18th International Conference on Information Technology-New Generations*. Intitulado *Combined Classification Models Applied to People Personality Identification*, o estudo conduzido avaliou o emprego da combinação de classificadores para o reconhecimento de traços de personalidade na teoria de Indicadores Tipológicos de Myers-Briggs (MBTI).

5.2 Limitações

O trabalho descreve um método para identificação de traços de personalidade em ambientes de Participação Eletrônica de acordo com o modelo Cinco Grandes Fatores. Ainda que amplamente estudado, o modelo não é consenso entre os pesquisadores das estruturas de personalidade. Sendo assim, a análise dos traços de personalidades apresentada se limita ao espectro de tendências comportamentais que o modelo engloba. Embora os resultados relatados sejam consistentes com descobertas de estudos anteriores e, portanto, corroborem com a análise das relações entre personalidade e participação apresentadas, não é objetivo do trabalho estabelecer um paradigma para essa análise.

Uma outra limitação do estudo se refere ao conjunto de dados. A quantidade de palavras utilizadas pelos usuários do SoPa no intervalo de tempo contemplado nesse estudo

foi pequena, fator que limitou inclusive a utilização do serviço *Watson Personality Insights* para identificação dos traços. Em contrapartida, tal limitação motivou a exploração dos algoritmos de regressão no treinamento de um modelo que fosse capaz de rotular a base de dados, possibilitando a identificação dos traços de personalidade.

Não faz parte do escopo deste trabalho analisar as implicações do emprego de técnicas computacionais para o reconhecimento de traços de personalidade através da coleta automática de dados. Entretanto, dada a vigência da Lei Geral de Proteção de Dados (LGPD) (Brasil, 2018) e a recente aprovação do projeto de lei que regulamenta uso da Inteligência Artificial pela Câmara Federal¹, esses aspectos devem ser levados em consideração em trabalhos futuros.

5.3 Trabalhos Futuros

As discussões, análises e limitações apresentadas ao longo desse trabalho permitem identificar oportunidades de trabalhos futuros. Uma dessas oportunidades é a validação das pontuações dos traços obtidas pelo modelo gerado junto às amostras de usuários utilizando outras abordagens, como questionários, o que permitirá ajustar o modelo e gerar predições com maior acurácia. Esse processo pode ser repetido à medida que novos usuários são identificados, melhorando de forma contínua o desempenho do modelo.

Outra proposta de evolução do trabalho é o estudo da variação dos hiperparâmetros dos algoritmos de Aprendizado de Máquina empregados, buscando encontrar configurações que otimizem o desempenho dos modelos. Além disso, o estudo da combinação dos diferentes regressores pode apresentar resultados ainda melhores.

A utilização de informações da interação do usuário com a ferramenta como, por exemplo, tempo de uso, quantidade de acessos, categoria de problema de maior interesse, dentre outros, combinadas com os textos publicados é uma oportunidade de identificar outros fatores que podem influenciar a personalidade.

Outra oportunidade identificada é a implementação de uma abordagem que utilize a representação das palavras na língua portuguesa, possibilitando a identificação da influência do idioma na detecção dos traços de personalidade.

Além disso, o conhecimento gerado pelo método apresentado pode ser empregado em conjunto com outras estratégias, como o *marketing* digital, *design* e até mesmo gamificação, para o estudo e definição de abordagens que permitam aumentar o engajamento cidadão nesses ambientes.

Por fim, o método descrito e os principais resultados desta pesquisa, detalhados nos Capítulos 3 e 4, serão objeto de futuras publicações.

¹ Agência Câmara de Notícias

Anexos

ANEXO A – Tabela de aspectos positivos dos traços do CGF

Este anexo reproduz a tabela apresentada por [McCrae e Costa \(1997\)](#) como um primeiro esforço de organização dos termos em língua natural que definem os aspectos positivos dos traços.

Table 1
Examples of Adjectives, Q-Sort Items, and Questionnaire Scales Defining the Five Factors

Factor		Factor definers		
Name	Number	Adjectives ^a	Q-sort items ^b	Scales ^c
Extraversion (E)	I	Active	Talkative	Warmth
		Assertive	Skilled in play, humor	Gregariousness
		Energetic	Rapid personal tempo	Assertiveness
		Enthusiastic	Facially, gesturally expressive	Activity
		Outgoing	Behaves assertively	Excitement Seeking
Agreeableness (A)	II	Talkative	Gregarious	Positive Emotions
		Appreciative	Not critical, skeptical	Trust
		Forgiving	Behaves in giving way	Straightforwardness
		Generous	Sympathetic, considerate	Altruism
		Kind	Arouses liking	Compliance
Conscientiousness (C)	III	Sympathetic	Warm, compassionate	Modesty
		Trusting	Basically trustful	Tender-Mindedness
		Efficient	Dependable, responsible	Competence
		Organized	Productive	Order
		Planful	Able to delay gratification	Dutifulness
		Reliable	Not self-indulgent	Achievement Striving
		Responsible	Behaves ethically	Self-Discipline
		Thorough	Has high aspiration level	Deliberation

Neuroticism (N)	-IV	Anxious Self-pitying Tense Touchy Unstable Worrying	Thin-skinned Brittle ego defenses Self-defeating Basically anxious Concerned with adequacy Fluctuating moods	Anxiety Hostility Depression Self-Consciousness Impulsiveness Vulnerability
Openness (O)	V	Artistic Curious Imaginative Insightful Original Wide interests	Wide range of interests Introspective Unusual thought processes Values intellectual matters Judges in unconventional terms Aesthetically reactive	Fantasy Aesthetics Feelings Actions Ideas Values

a. Adjective Check List items defining the factor in a study of 280 men and women rated by 10 psychologists serving as observers during an assessment weekend at the Institute of Personality Assessment and Research (John, 1989a).

b. California Q-Sct items from self-sorts by 403 men and women in the Baltimore Longitudinal Study of Aging (McCrae, Costa, & Busch, 1986).

c. Revised NEO Personality Inventory facet scales from self-reports by 1,539 adult men and women (Costa, McCrae, & Dye, 1991).

Referências

- ABAD, Z. S. H.; KLINE, A.; SULTANA, M.; NOAEEN, M.; NURMAMBETOVA, E.; LUCINI, F.; AL-JEFRI, M.; LEE, J. Digital public health surveillance: a systematic scoping review. *npj Digital Medicine*, v. 4, n. 1, p. 41, dez. 2021. ISSN 2398-6352. Disponível em: <<http://www.nature.com/articles/s41746-021-00407-6>>. 29
- AGUIAR, J. J. B.; ARAÚJO, J. M. F. R. d.; COSTA, E. d. B. Estudo Comparativo de Abordagens para Sistemas de Recomendação baseados em Personalidade com uso do serviço IBM Watson Personality Insights. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, n. 40, p. 73–88, dez. 2020. ISSN 16469895. Disponível em: <http://scielo.pt/scielo.php?script=sci_arttext&pid=S1646-98952020000500073&lng=pt&nrm=iso&tlng=pt>. 32
- ALHARBI, A.; KANG, K.; HAWRYSZKIEWYCZ, I. The Influence of Trust and subjective Norms on Citizens' Intentions to Engage in E-participation on E-government Websites. In: *ACIS 2015 Proceedings*. Adelaide: [s.n.], 2015. 19
- ALLEN, J. *Natural language understanding*. Menlo Park, Calif: Benjamin/Cummings Pub. Co, 1987. (Benjamin/Cummings series in computer science). ISBN 978-0-8053-0330-8. 28
- ALLPORT, G. W.; ODBERT, H. S. Trait-names: A psycho-lexical study. *Psychological Monographs*, v. 47, n. 1, p. i–171, 1936. ISSN 0096-9753. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/h0093360>>. 13
- ALWABEL, A. S. A.; ZENG, X.-J. Data-driven modeling of technology acceptance: A machine learning perspective. *Expert Systems with Applications*, v. 185, p. 115584, dez. 2021. ISSN 09574174. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417421009866>>. 29
- ARAÚJO, R. d. P. A.; PENTEADO, C. L. C.; SANTOS, M. B. P. d. Democracia digital e experiências de e-participação: webativismo e políticas públicas. *História, Ciências, Saúde-Manguinhos*, v. 22, n. suppl, p. 1597–1619, dez. 2015. ISSN 0104-5970. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-59702015001001597&lng=pt&tlng=pt>. 13
- ARMSTRONG, G.; ADAM, S.; DENIZE, S.; KOTLER, P. *Principles of marketing*. [S.l.]: Pearson Australia, 2014. 31
- AZUCAR, D.; MARENGO, D.; SETTANNI, M. Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, v. 124, p. 150–159, abr. 2018. ISSN 01918869. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0191886917307328>>. 33
- BARBIER, G.; LIU, H. Data mining in social media. In: _____. Boston, MA: Springer US, 2011. p. 327–352. ISBN 978-1-4419-8462-3. Disponível em: <https://doi.org/10.1007/978-1-4419-8462-3_12>. 33

BENET-MARTÍNEZ, V.; JOHN, O. P. Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, v. 75, n. 3, p. 729–750, 1998. ISSN 1939-1315, 0022-3514. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.75.3.729>>. 20

BERNARDOS, M. S. La estructura de personalidad: El enfoque léxico y los «cinco grandes». *Estudios de psicología*, Taylor & Francis, v. 13, n. 47, p. 73–87, 1992. 21

BERNTZEN, L.; JOHANNESSEN, M. R. The Role of Citizen Participation in Municipal Smart City Projects: Lessons Learned from Norway. In: GIL-GARCIA, J. R.; PARDO, T. A.; NAM, T. (Ed.). *Smarter as the New Urban Agenda*. Cham: Springer International Publishing, 2016. v. 11, p. 299–314. ISBN 978-3-319-17619-2 978-3-319-17620-8. Series Title: Public Administration and Information Technology. Disponível em: <http://link.springer.com/10.1007/978-3-319-17620-8_16>. 13

BERTOT, J. C.; JAEGER, P. T.; HANSEN, D. The impact of polices on government social media usage: Issues, challenges, and recommendations. *Government Information Quarterly*, v. 29, n. 1, p. 30–40, jan. 2012. ISSN 0740624X. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0740624X11000992>>. 18

BIRD, S.; KLEIN, E.; LOPER, E. *Natural language processing with Python: analyzing text with the natural language toolkit*. [S.l.]: "O'Reilly Media, Inc.", 2009. 41

BLAIS, A.; ST-VINCENT, S. L. Personality traits, political attitudes and the propensity to vote: personality traits, political attitudes and the propensity to vote. *European Journal of Political Research*, v. 50, n. 3, p. 395–417, maio 2011. ISSN 03044130. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1475-6765.2010.01935.x>>. 55

BOND, M. H.; FORGAS, J. P. Linking Person Perception to Behavior Intention Across Cultures: The Role of Cultural Collectivism. *Journal of Cross-Cultural Psychology*, v. 15, n. 3, p. 337–352, set. 1984. ISSN 0022-0221, 1552-5422. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0022002184015003006>>. 21

BORCHANI, H.; VARANDO, G.; BIELZA, C.; LARRAÑAGA, P. A survey on multi-output regression: Multi-output regression survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 5, n. 5, p. 216–233, set. 2015. ISSN 19424787. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/widm.1157>>. 27

BORKENAU, P.; OSTENDORF, F. Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, v. 11, n. 5, p. 515–524, jan. 1990. ISSN 01918869. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/019188699090065Y>>. 21

BOTCHKAREV, A. A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, v. 14, p. 045–076, 2019. ISSN 1555-1237. Publisher: Informing Science Institute. Disponível em: <<http://dx.doi.org/10.28945/4184>>. 27

Brasil. *Lei nº 13.709, de 14 de agosto de 2018*. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm>. 59

- BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. ISSN 08856125. Disponível em: <<http://link.springer.com/10.1023/A:1010933404324>>. 26
- BREIMAN, L.; FRIEDMAN, J. H. Predicting Multivariate Responses in Multiple Linear Regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 59, n. 1, p. 3–54, fev. 1997. ISSN 1369-7412, 1467-9868. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00054>>. 27
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. *Classification And Regression Trees*. 1. ed. Routledge, 1984. ISBN 978-1-315-13947-0. Disponível em: <<https://www.taylorfrancis.com/books/9781351460491>>. 25
- CAETANO, B.; PAULA, M.; SOUZA, J. D. SoPa: A Social Media for a Participatory Society. *IEEE Access*, v. 8, p. 70627–70639, 2020. ISSN 2169-3536. Disponível em: <<https://ieeexplore.ieee.org/document/9060917/>>. 13, 18
- CERON, W.; SANSEVERINO, G. G.; LIMA-SANTOS, M. de; QUILES, M. G. COVID-19 fake news diffusion across Latin America. *Social Network Analysis and Mining*, v. 11, n. 1, p. 47, dez. 2021. ISSN 1869-5450, 1869-5469. Disponível em: <<https://link.springer.com/10.1007/s13278-021-00753-z>>. 29
- CHAUDHRY, H. N.; JAVED, Y.; KULSOOM, F.; MEHMOOD, Z.; KHAN, Z. I.; SHOAI, U.; JANJUA, S. H. Sentiment Analysis of before and after Elections: Twitter Data of U.S. Election 2020. *Electronics*, v. 10, n. 17, p. 2082, ago. 2021. ISSN 2079-9292. Disponível em: <<https://www.mdpi.com/2079-9292/10/17/2082>>. 29
- Colab. *Colab - Fazer a cidade melhor está na sua mão*. 2021. Acessado em: 29/09/2021. Disponível em: <<https://www.colab.re/>>. 34
- COPPERSMITH, G.; LEARY, R.; CRUTCHLEY, P.; FINE, A. Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical Informatics Insights*, v. 10, p. 117822261879286, jan. 2018. ISSN 1178-2226, 1178-2226. Disponível em: <<http://journals.sagepub.com/doi/10.1177/1178222618792860>>. 28
- COPPIN, B. *Inteligência artificial*. [S.l.]: Editora LTC, 2013. 22
- CORTÉS-CEDIÉL, M. E.; GIL, O.; CANTADOR, I. Defining the engagement life cycle in e-participation. In: *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. Delft The Netherlands: ACM, 2018. p. 1–2. ISBN 978-1-4503-6526-0. Disponível em: <<https://dl.acm.org/doi/10.1145/3209281.3209393>>. 19
- COSTA, P. T.; MCCRAE, R. R. From catalog to classification: Murray’s needs and the five-factor model. *Journal of Personality and Social Psychology*, v. 55, n. 2, p. 258–265, 1988. ISSN 1939-1315, 0022-3514. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.55.2.258>>. 20
- COSTA, P. T.; MCCRAE, R. R. The Revised NEO Personality Inventory (NEO-PI-R). In: *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing*. 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, 2008. p. 179–198. ISBN 978-1-4129-4652-0 978-1-84920-047-9. Disponível em: <http://sk.sagepub.com/reference/hdbk_personalitytheory2/n9.xml>. 19, 22

- COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, jan. 1967. ISSN 0018-9448, 1557-9654. Disponível em: <<http://ieeexplore.ieee.org/document/1053964/>>. 25
- CUNNINGHAM, P.; DELANY, S. J. k-Nearest Neighbour Classifiers - A Tutorial. *ACM Computing Surveys*, v. 54, n. 6, p. 1–25, jul. 2021. ISSN 0360-0300, 1557-7341. Disponível em: <<https://dl.acm.org/doi/10.1145/3459665>>. 25
- CUPERMAN, R.; ICKES, W. Big Five predictors of behavior and perceptions in initial dyadic interactions: Personality similarity helps extraverts and introverts, but hurts “disagreeables”. *Journal of Personality and Social Psychology*, v. 97, n. 4, p. 667–684, 2009. ISSN 1939-1315, 0022-3514. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/a0015741>>. 21
- DIGMAN, J. M. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, v. 41, n. 1, p. 417–440, jan. 1990. ISSN 0066-4308, 1545-2085. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.ps.41.020190.002221>>. 20
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. d.; ALMEIDA, T. A. d. *Inteligência Artificial: uma abordagem de aprendizado de máquina*. 2. ed. Rio de Janeiro: Editora LTC, 2021. 22, 23, 24, 25, 26
- FARNADI, G.; SITARAMAN, G.; SUSHMITA, S.; CELLI, F.; KOSINSKI, M.; STILLWELL, D.; DAVALOS, S.; MOENS, M.; COCK, M. D. Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, v. 26, n. 2-3, p. 109–142, jun. 2016. ISSN 0924-1868, 1573-1391. Disponível em: <<http://link.springer.com/10.1007/s11257-016-9171-0>>. 29, 46
- FARZINDAR, A.; INKPEN, D. Natural Language Processing for Social Media. *Synthesis Lectures on Human Language Technologies*, v. 8, n. 2, p. 1–166, ago. 2015. ISSN 1947-4040, 1947-4059. Disponível em: <<http://www.morganclaypool.com/doi/10.2200/S00659ED1V01Y201508HLT030>>. 28
- FERREIRA, M. *Para conhecer: linguística computacional*. [S.l.: s.n.], 2019. OCLC: 1135409755. ISBN 978-85-520-0152-2. 28
- FLORES-MENDOZA, C. E. *Inventário de Personalidade NEO Revisado NEO PI-R - Manual*. São Paulo: Vetor, 2007. OCLC: 246923948. ISBN 978-85-7585-220-0. 22
- FUNAHASHI, K.-I. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, v. 2, n. 3, p. 183–192, jan. 1989. ISSN 08936080. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0893608089900038>>. 24
- GALLO, F. R.; SIMARI, G. I.; MARTINEZ, M. V.; FALAPPA, M. A. Predicting user reactions to Twitter feed content based on personality type and social cues. *Future Generation Computer Systems*, v. 110, p. 918–930, set. 2020. ISSN 0167739X. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167739X19304091>>. 32
- GARCÍA, L. F. Teorias psicométricas da personalidade. In: *Introdução à psicologia das diferenças individuais*. Porto Alegre: Artmed, 2006. p. 219–242. ISBN 978-85-363-1418-1. 21

- GARDNER, M.; DORLING, S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, v. 32, n. 14-15, p. 2627–2636, ago. 1998. ISSN 13522310. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1352231097004470>>. 24
- GARRIDO, C.; OÑA, R. de; OÑA, J. de. Neural networks for analyzing service quality in public transportation. *Expert Systems with Applications*, v. 41, n. 15, p. 6830–6838, nov. 2014. ISSN 09574174. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417414002681>>. 24
- GERBER, A. S.; HUBER, G. A.; DOHERTY, D.; DOWLING, C. M.; RASO, C.; HA, S. E. Personality Traits and Participation in Political Processes. *The Journal of Politics*, v. 73, n. 3, p. 692–706, jul. 2011. ISSN 0022-3816, 1468-2508. Disponível em: <<https://www.journals.uchicago.edu/doi/10.1017/S0022381611000399>>. 54, 55, 56
- GHOSH, S.; GUNNING, D. *Natural Language Processing Fundamentals*. [s.n.], 2019. OCLC: 1104232719. ISBN 978-1-78995-404-3. Disponível em: <<http://sbiproxy.uqac.ca/login?url=https://international.scholarvox.com/book/88867574>>. 30
- GIUNTINI, F. T.; MORAES, K. L. de; CAZZOLATO, M. T.; KIRCHNER, L. de F.; REIS, M. de J. D. D.; TRAINA, A. J. M.; CAMPBELL, A. T.; UEYAMA, J. Modeling and Assessing the Temporal Behavior of Emotional and Depressive User Interactions on Social Networks. *IEEE Access*, v. 9, p. 93182–93194, 2021. ISSN 2169-3536. Disponível em: <<https://ieeexplore.ieee.org/document/9462934/>>. 28
- GOLBECK, J.; ROBLES, C.; TURNER, K. Predicting personality with social media. In: *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*. Vancouver, BC, Canada: ACM Press, 2011. p. 253. ISBN 978-1-4503-0268-5. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1979742.1979614>>. 33, 46
- GOLDBERG, L. R. The structure of phenotypic personality traits. *American Psychologist*, v. 48, n. 1, p. 26–34, 1993. ISSN 1935-990X, 0003-066X. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.48.1.26>>. 21
- GOLDBERG, L. R.; MCREYNOLDS, P. A historical survey of personality scales and inventories. In: *Advances in psychological assessment, Volume 2*. Palo Alto, CA: Science and Behavior Books, 1971. 22
- Google. *Cloud Translation | Google Cloud*. 2021. Disponível em: <<https://cloud.google.com/translate?hl=pt-br>>. 37
- Google. *Python client library | Cloud Translation | Google Cloud*. 2021. Disponível em: <<https://cloud.google.com/translate/docs/reference/libraries/v3/python?hl=pt-br>>. 37
- GOZUACIK, N.; SAKAR, C. O.; OZCAN, S. Social media-based opinion retrieval for product analysis using multi-task deep neural networks. *Expert Systems with Applications*, v. 183, p. 115388, nov. 2021. ISSN 09574174. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0957417421008125>>. 29

- GRUDA, D.; KARANATSIU, D.; MENDHEKAR, K.; GOLBECK, J.; VAKALI, A. I Alone Can Fix It: Examining interactions between narcissistic leaders and anxious followers on Twitter using a machine learning approach. *Journal of the Association for Information Science and Technology*, v. 72, n. 11, p. 1323–1336, nov. 2021. ISSN 2330-1635, 2330-1643. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/asi.24490>>. 29
- HA, S. E.; KIM, S.; JO, S. H. Personality Traits and Political Participation: Evidence from South Korea: Personality and Political Participation in Korea. *Political Psychology*, v. 34, n. 4, p. 511–532, ago. 2013. ISSN 0162895X. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/pops.12008>>. 54
- HASSAN, L.; HAMARI, J. Gameful civic engagement: A review of the literature on gamification of e-participation. *Government Information Quarterly*, v. 37, n. 3, p. 101461, jul. 2020. ISSN 0740624X. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0740624X19302606>>. 19
- HIRSCHBERG, J.; MANNING, C. D. Advances in natural language processing. *Science*, v. 349, n. 6245, p. 261–266, 2015. Publisher: American Association for the Advancement of Science. 28
- HUTZ, C. S.; NUNES, C. H.; SILVEIRA, A. D.; SERRA, J.; ANTON, M.; WIECZOREK, L. S. O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia: Reflexão e Crítica*, v. 11, n. 2, p. 395–411, 1998. ISSN 0102-7972. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-79721998000200015&lng=pt&nrm=iso&tlng=pt>. 21
- IBM. *A ciência por trás do serviço*. 2020. Disponível em: <<https://cloud.ibm.com/docs/personality-insights?topic=personality-insights-science>>. 31, 32
- IBM. *Modelos de personalidade*. 2020. Acessado em: 23/03/2021. Disponível em: <<https://cloud.ibm.com/docs/personality-insights?topic=personality-insights-models>>. 49
- IKONOMAKIS, M.; KOTSIANTIS, S. B.; TAMPAKAS, V. T. Text Classification Using Machine Learning Techniques. In: . [S.l.: s.n.], 2005. 29
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM Computing Surveys*, v. 31, n. 3, p. 264–323, set. 1999. ISSN 0360-0300, 1557-7341. Disponível em: <<https://dl.acm.org/doi/10.1145/331499.331504>>. 22
- JOHN, O. P.; ANGLEITNER, A.; OSTENDORF, F. The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, v. 2, n. 3, p. 171–203, set. 1988. ISSN 0890-2070, 1099-0984. Disponível em: <<http://journals.sagepub.com/doi/10.1002/per.2410020302>>. 21
- JUNIOR, R. A. P.; INKPEN, D. Using Cognitive Computing to Get Insights on Personality Traits from Twitter Messages. In: MOUHOU, M.; LANGLAIS, P. (Ed.). *Advances in Artificial Intelligence*. Cham: Springer International Publishing, 2017. v. 10233, p. 278–283. ISBN 978-3-319-57350-2 978-3-319-57351-9. Series Title: Lecture Notes in Computer Science. Disponível em: <http://link.springer.com/10.1007/978-3-319-57351-9_32>. 29, 32

- KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, v. 53, n. 1, p. 59–68, jan. 2010. ISSN 00076813. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0007681309001232>>. 18
- KOOCHAKSARAEI, R. H.; MENEGHINI, I. R.; COELHO, V. N.; GUIMARÃES, F. G. A new visualization method in many-objective optimization with chord diagram and angular mapping. *Knowledge-Based Systems*, v. 138, p. 134–154, dez. 2017. ISSN 09507051. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0950705117304562>>. 49
- KUNCHEVA, L. I. *Combining pattern classifiers: methods and algorithms*. [S.l.]: John Wiley & Sons, 2014. 44
- KWAK, H.; LEE, C.; PARK, H.; MOON, S. What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World wide web - WWW '10*. Raleigh, North Carolina, USA: ACM Press, 2010. p. 591. ISBN 978-1-60558-799-8. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1772690.1772751>>. 29
- LEOPOLD, E.; KINDERMANN, J. Text Categorization with Support Vector Machines. How to Represent Texts in Input Space? *Machine Learning*, v. 46, n. 1, p. 423–444, jan. 2002. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1023/A:1012491419635>>. 29
- LIMA, A. C. E.; CASTRO, L. N. de. A multi-label, semi-supervised classification approach applied to personality prediction in social media. *Neural Networks*, v. 58, p. 122–130, out. 2014. ISSN 08936080. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0893608014001282>>. 14, 33
- LIMAS, A. F. d.; PRIMI, R.; CARVALHO, L. d. F. Avaliação da personalidade por redes sociais online: uso do Facebook na área. *Revista Sul-Americana de Psicologia*, v. 2, n. 1, nov. 2018. Section: Artículos. Disponível em: <<http://ediciones.ucsh.cl/ojs/index.php/RSAP/article/view/1732>>. 33
- LOPES, P. N.; SALOVEY, P.; STRAUS, R. Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, v. 35, n. 3, p. 641–658, ago. 2003. ISSN 01918869. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0191886902002428>>. 20
- LOPES, R.; SOUZA, L. M.; AMÂNCIO, L.; MARTINS, J.; MESQUITA, N.; SANTOS, G.; CASTRO, P. A. INTERAÇÕES INTERPESSOAIS E TRAÇOS DE PERSONALIDADE: UM ESTUDO DE VALIDADE PARA O CHECKLIST DE RELAÇÕES INTERPESSOAIS – II. *ENCICLOPÉDIA BIOSFERA*, v. 10, n. 19, nov. 2014. Disponível em: <<https://conhecer.org.br/ojs/index.php/biosfera/article/view/2550>>. 20
- LOPEZ, C. E.; GALLEMORE, C. An augmented multilingual Twitter dataset for studying the COVID-19 infodemic. *Social Network Analysis and Mining*, v. 11, n. 1, p. 102, dez. 2021. ISSN 1869-5450, 1869-5469. Disponível em: <<https://link.springer.com/10.1007/s13278-021-00825-0>>. 29
- MACINTOSH, A. Characterizing e-participation in policy-making. In: *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. [S.l.: s.n.], 2004. p. 10 pp.–. 16

- MAIRESSE, F.; WALKER, M. A.; MEHL, M. R.; MOORE, R. K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, v. 30, p. 457–500, nov. 2007. ISSN 1076-9757. Disponível em: <<https://jair.org/index.php/jair/article/view/10520>>. 22
- MAMBREY, P. Networked ICT to Foster e-Democracy? In: HUTCHISON, D.; KANADE, T.; KITTLER, J.; KLEINBERG, J. M.; MATTERN, F.; MITCHELL, J. C.; NAOR, M.; NIERSTRASZ, O.; RANGAN, C. P.; STEFFEN, B.; SUDAN, M.; TERZOPOULOS, D.; TYGAR, D.; VARDI, M. Y.; WEIKUM, G.; TRAUNMÜLLER, R. (Ed.). *Electronic Government*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. v. 3183, p. 31–35. ISBN 978-3-540-22916-2 978-3-540-30078-6. Series Title: Lecture Notes in Computer Science. Disponível em: <http://link.springer.com/10.1007/978-3-540-30078-6_6>. 16
- MCCRAE, R. R. Traits and trait names: How well is Openness represented in natural languages? *European Journal of Personality*, v. 4, n. 2, p. 119–129, jun. 1990. ISSN 0890-2070, 1099-0984. Disponível em: <<http://journals.sagepub.com/doi/10.1002/per.2410040205>>. 21
- MCCRAE, R. R.; COSTA, P. T. Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, v. 52, n. 1, p. 81–90, 1987. ISSN 1939-1315, 0022-3514. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.52.1.81>>. 13
- MCCRAE, R. R.; COSTA, P. T. Personality trait structure as a human universal. *American Psychologist*, v. 52, n. 5, p. 509–516, 1997. ISSN 1935-990X, 0003-066X. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.52.5.509>>. 21, 54, 61
- MEDAGLIA, R. eParticipation research: Moving characterization forward (2006–2011). *Government Information Quarterly*, v. 29, n. 3, p. 346–360, 2012. ISSN 0740-624X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0740624X12000378>>. 19
- MEHTA, D.; DWIVEDI, A.; PATRA, A.; KUMAR, M. A. A transformer-based architecture for fake news classification. *Social Network Analysis and Mining*, v. 11, n. 1, p. 39, dez. 2021. ISSN 1869-5450, 1869-5469. Disponível em: <<https://link.springer.com/10.1007/s13278-021-00738-y>>. 29
- MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. *Efficient Estimation of Word Representations in Vector Space*. 2013. _eprint: 1301.3781. 30
- MIKOLOV, T.; YIH, W.; ZWEIG, G. Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics, 2013. p. 746–751. Disponível em: <<https://aclanthology.org/N13-1090>>. 30
- MITCHELL, T. *Machine Learning*. [S.l.]: McGraw Hill, 1997. 22
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos Sobre Aprendizado de Máquina. In: *Sistemas Inteligentes Fundamentos e Aplicações*. 1. ed. Barueri-SP: Manole Ltda, 2003. p. 89–114. ISBN 85-204-168. 27

- MONDAK, J. J. *Personality and the Foundations of Political Behavior*. Cambridge: Cambridge University Press, 2010. ISBN 978-0-511-76151-5. Disponível em: <<http://ebooks.cambridge.org/ref/id/CBO9780511761515>>. 55
- MONDAK, J. J.; HALPERIN, K. D. A Framework for the Study of Personality and Political Behaviour. *British Journal of Political Science*, v. 38, n. 2, p. 335–362, abr. 2008. ISSN 0007-1234, 1469-2112. Disponível em: <https://www.cambridge.org/core/product/identifier/S0007123408000173/type/journal_article>. 54
- MOTA, F.; PAULA, M.; DRUMMOND, I. Combined Classification Models Applied to People Personality Identification. In: LATIFI, S. (Ed.). *ITNG 2021 18th International Conference on Information Technology-New Generations*. Cham: Springer International Publishing, 2021. v. 1346, p. 457–462. ISBN 978-3-030-70415-5 978-3-030-70416-2. Series Title: Advances in Intelligent Systems and Computing. Disponível em: <https://link.springer.com/10.1007/978-3-030-70416-2_59>. 33
- MOTA, F. B. d. S.; PAULA, J. H. R.; CAETANO, B. P.; PAULA, M. M. V. d.; SOUZA, J. M. d. TAPE: taxonomia de aplicações em participação eletrônica. *Revista Tecnologia e Sociedade*, v. 16, n. 42, p. 207, jun. 2020. ISSN 1984-3526, 1809-0044. Disponível em: <<https://periodicos.utfpr.edu.br/rts/article/view/10455>>. 17
- MYERS, I. B.; MYERS, P. B. *Gifts differing: understanding personality type*. Palo Alto, Calif: Davies-Black Pub, 1995. ISBN 978-0-89106-074-1. 21
- NO, W.; MOOK, L.; SCHUGURENSKY, D. Examining ideation processes in online invited spaces. In: *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*. Shanghai China: ACM, 2016. p. 67–75. ISBN 978-1-4503-4339-8. Disponível em: <<https://dl.acm.org/doi/10.1145/2912160.2912197>>. 19
- ONU. *United Nations E-Government Survey 2020: Digital government in the decade of action for sustainable development*. New York: United Nations, 2020. (United Nations e-government survey, 2020). ISBN 978-92-1-123210-3. 16, 17
- OSOJNIK, A.; PANOV, P.; DŽEROSKI, S. Multi-label classification via multi-target regression on data streams. *Machine Learning*, v. 106, n. 6, p. 745–770, jun. 2017. ISSN 0885-6125, 1573-0565. Disponível em: <<http://link.springer.com/10.1007/s10994-016-5613-5>>. 27
- PADE-KHENE, C.; THINYANE, H.; MACHIRI, M. Building foundations before technology: An operation model for digital citizen engagement in resource constrained contexts. In: . [S.l.: s.n.], 2017. 19
- PAN, S. J.; YANG, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, out. 2010. ISSN 1041-4347. Disponível em: <<http://ieeexplore.ieee.org/document/5288526/>>. 30
- PANOPOULOU, E.; TAMBOURIS, E.; TARABANIS, K. Success factors in designing eParticipation initiatives. *Information and Organization*, v. 24, n. 4, p. 195–213, out. 2014. ISSN 14717727. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S147177271400027X>>. 16

- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. 42
- PENNINGTON, J.; SOCHER, R.; MANNING, C. D. GloVe: Global Vectors for Word Representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*. [s.n.], 2014. p. 1532–1543. Disponível em: <<http://www.aclweb.org/anthology/D14-1162>>. 30, 31, 41
- PRIMI, R. Avaliação psicológica no Brasil: fundamentos, situação atual e direções para o futuro. *Psicologia: Teoria e Pesquisa*, v. 26, n. spe, p. 25–35, 2010. ISSN 0102-3772. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-37722010000500003&lng=pt&tlng=pt>. 21
- QIU, L.; LIN, H.; RAMSAY, J.; YANG, F. You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, v. 46, n. 6, p. 710–718, dez. 2012. ISSN 00926566. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S009265661200133X>>. 29
- QUERCIA, D.; KOSINSKI, M.; STILLWELL, D.; CROWCROFT, J. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In: *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*. Boston, MA, USA: IEEE, 2011. p. 180–185. ISBN 978-1-4577-1931-8 978-0-7695-4578-3. Disponível em: <<http://ieeexplore.ieee.org/document/6113111/>>. 33
- QUINLAN, J. R. Induction of decision trees. *Machine Learning*, v. 1, n. 1, p. 81–106, mar. 1986. ISSN 0885-6125, 1573-0565. Disponível em: <<http://link.springer.com/10.1007/BF00116251>>. 26
- QUINTERO-ANGULO, R. A. D.; SÁNCHEZ-TORRES, J. M.; CARDONA-ROMÁN, D. M. Problem areas in e-participation: A systematic review. In: *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*. New York, NY, USA: Association for Computing Machinery, 2020. (ICEGOV 2020), p. 544–550. ISBN 9781450376747. Disponível em: <<https://doi.org/10.1145/3428502.3428584>>. 16
- REBOLLO, I.; HARRIS, J. R. Genes, ambiente e personalidade. In: *Introdução à psicologia das diferenças individuais*. [S.l.: s.n.], 2006. p. 300–322. 19
- ŘEHŮŘEK, R.; SOJKA, P. Software Framework for Topic Modelling with Large Corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, 2010. p. 45–50. <<http://is.muni.cz/publication/884893/en>>. 41
- RIBEIRO, M. M.; CUNHA, M. A.; BARBOSA, A. F. E-participation, social media and digital gap: challenges in the brazilian context. In: *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. Delft The Netherlands: ACM, 2018. p. 1–9. ISBN 978-1-4503-6526-0. Disponível em: <<https://dl.acm.org/doi/10.1145/3209281.3209373>>. 17

- ROCCAS, S.; SAGIV, L.; SCHWARTZ, S. H.; KNAFO, A. The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin*, v. 28, n. 6, p. 789–801, jun. 2002. ISSN 0146-1672, 1552-7433. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0146167202289008>>. 55
- RODRIGUES, J. d. L.; GOMES, S. C.; BORGES, F. Q. Electronic participation via Facebook: the potential from Brazilian municipalities. *Transforming Government: People, Process and Policy*, ahead-of-print, n. ahead-of-print, jul. 2020. ISSN 1750-6166, 1750-6166. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/TG-06-2019-0049/full/html>>. 18
- ROESSLEIN, J. Tweepy: Twitter for python! URL: <https://github.com/tweepy/tweepy>, 2020. 34
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386–408, 1958. ISSN 1939-1471, 0033-295X. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042519>>. 24
- RUMELHART, D. E. M.; FELDMAN, J. J. L. *Parallel Distributed Processing Explorations in the Microstructure of Cognition: Foundations*. A Bradford Book, 1987. OCLC: 1191080283. ISBN 978-0-262-29140-8. Disponível em: <<http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9780262291408>>. 24
- RUTKOWSKI, L.; PIETRUCZUK, L.; DUDA, P.; JAWORSKI, M. Decision Trees for Mining Data Streams Based on the McDiarmid's Bound. *IEEE Transactions on Knowledge and Data Engineering*, v. 25, n. 6, p. 1272–1279, jun. 2013. ISSN 1041-4347. Disponível em: <<http://ieeexplore.ieee.org/document/6171195/>>. 25
- SANTORO, P. F.; CYMBALISTA, R.; CARDOSO, P. d. M. *Plano Diretor: participar é um direito!* 1. ed. Brasil, São Paulo: Editora Instituto Pólis, 2005. 19
- SCHWARTZ, S. H. Les valeurs de base de la personne : théorie, mesures et applications. *Revue française de sociologie*, v. 47, n. 4, p. 929, 2006. ISSN 0035-2969, 1958-5691. Disponível em: <<http://www.cairn.info/revue-francaise-de-sociologie-1-2006-4-page-929.htm>>. 31
- SEBASTIANI, F. Machine learning in automated text categorization. *ACM Computing Surveys*, v. 34, n. 1, p. 1–47, mar. 2002. ISSN 0360-0300, 1557-7341. Disponível em: <<https://dl.acm.org/doi/10.1145/505282.505283>>. 29
- SETTLES, B. *Active Learning Literature Survey*. [S.l.], 2009. 23
- SILVA, I. B.; NAKANO, T. d. C. Modelo dos cinco grandes fatores da personalidade: análise de pesquisas. *Avaliação Psicológica*, scielopepsic, v. 10, p. 51 – 62, 04 2011. ISSN 1677-0471. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712011000100006&nrm=iso>. 20
- SILVA, R. S.; SCHLOTTFELDT, C. G.; ROZENBERG, M. P.; SANTOS, M. T.; LELÉ, J. Replicabilidade do Modelo dos Cinco Grandes Fatores em medidas da personalidade. *Mosaico: Estudos em Psicologia*, v. 1, n. 1, mar. 2017. Disponível em: <<https://periodicos.ufmg.br/index.php/mosaico/article/view/6230>>. 21

- SKOWRON, M.; TKALČIČ, M.; FERWERDA, B.; SCHEDL, M. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. In: *Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion*. Montréal, Québec, Canada: ACM Press, 2016. p. 107–108. ISBN 978-1-4503-4144-8. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2872518.2889368>>. 29, 46
- SUMNER, C.; BYERS, A.; BOOCHEVER, R.; PARK, G. J. Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. In: *2012 11th International Conference on Machine Learning and Applications*. Boca Raton, FL, USA: IEEE, 2012. p. 386–393. ISBN 978-1-4673-4651-1 978-0-7695-4913-2. Disponível em: <<http://ieeexplore.ieee.org/document/6406767/>>. 33
- SUTTON, R. S.; BARTO, A. G. et al. *Introduction to reinforcement learning*. [S.l.]: MIT press Cambridge, 1998. v. 135. 23
- SVETNIK, V.; LIAW, A.; TONG, C.; CULBERSON, J. C.; SHERIDAN, R. P.; FEUSTON, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, v. 43, n. 6, p. 1947–1958, nov. 2003. ISSN 0095-2338. Disponível em: <<https://pubs.acs.org/doi/10.1021/ci034160g>>. 27
- TRENTINI, C. M.; HUTZ, C. S.; BANDEIRA, D. R.; TEIXEIRA, M. A. P.; GONÇALVES, M. T. A.; THOMAZONI, A. R. Correlações entre a EFN - Escala Fatorial de Neuroticismo e o IFP - Inventário Fatorial de Personalidade. *Avaliação Psicológica*, v. 8, p. 209 – 217, ago. 2009. ISSN 1677-0471. Publisher: scieloepsic. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712009000200007&nrm=iso>. 20
- Twitter. *Twitter API Documentation*. 2021. Disponível em: <<https://developer.twitter.com/en/docs/twitter-api>>. 29
- WEINSCHENK, A. C. Big Five Personality Traits, Political Participation, and Civic Engagement: Evidence from 24 Countries: Big Five Personality Traits. *Social Science Quarterly*, v. 98, n. 5, p. 1406–1421, nov. 2017. ISSN 00384941. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/ssqu.12380>>. 54, 56
- WERBOS, P. J. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Tese (Doutorado) — Harvard University, 1974. 24
- XU, D.; SHI, Y.; TSANG, I. W.; ONG, Y.; GONG, C.; SHEN, X. Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, p. 1–21, 2019. ISSN 2162-237X, 2162-2388. Disponível em: <<https://ieeexplore.ieee.org/document/8892612/>>. 27
- YANG, K.; BOND, M. H. Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology*, v. 58, n. 6, p. 1087–1095, 1990. ISSN 1939-1315, 0022-3514. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.58.6.1087>>. 21
- YILDIZ, B.; BILBAO, J.; SPROUL, A. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and*

Sustainable Energy Reviews, v. 73, p. 1104–1122, jun. 2017. ISSN 13640321. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1364032117302265>>. 23

ZHU, X.; GOLDBERG, A. B. Introduction to Semi-Supervised Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, v. 3, n. 1, p. 1–130, jan. 2009. ISSN 1939-4608, 1939-4616. Disponível em: <<http://www.morganclaypool.com/doi/abs/10.2200/S00196ED1V01Y200906AIM006>>. 23

ZUKIN, C.; KEETER, S.; ANDOLINA, M.; JENKINS, K.; CARPINI, M. X. D. *A New Engagement?* Oxford University Press, 2006. ISBN 978-0-19-518317-7. Disponível em: <<https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195183177.001.0001/acprof-9780195183177>>. 19