



**UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

**MAPEANDO TRAÇOS DA PERSONALIDADE ATRAVÉS
DO RITMO DE DIGITAÇÃO**

FELIPE VALADÃO GOULART

São Cristóvão
2016



**UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

FELIPE VALADÃO GOULART

**MAPEANDO TRAÇOS DA PERSONALIDADE ATRAVÉS
DO RITMO DE DIGITAÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (PROCC) da Universidade Federal de Sergipe (UFS) como requisito para a obtenção do título de mestre em Ciência da Computação.

Orientador: Prof. Dr. Daniel Oliveira Dantas

São Cristóvão
2016

FELIPE VALADÃO GOULART

**MAPEANDO TRAÇOS DA PERSONALIDADE ATRAVÉS
DO RITMO DE DIGITAÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (PROCC) da Universidade Federal de Sergipe (UFS) como requisito para a obtenção do título de mestre em Ciência da Computação.

BANCA EXAMINADORA

Prof. Dr. Daniel Oliveira Dantas
Universidade Federal de Sergipe (UFS)

Prof. Dr. Jugurta Rosa Montalvão Filho
Universidade Federal de Sergipe (UFS)

Prof. Dra. Magali Teresinha Longhi
Universidade Federal do Rio Grande do Sul (UFRGS)

MAPEANDO TRAÇOS DA PERSONALIDADE ATRAVÉS DO RITMO DE DIGITAÇÃO

Este exemplar corresponde à redação da
Dissertação de Mestrado, sendo a defesa do
mestrando **Felipe Valadão Goulart** para ser
aprovada pela banca examinadora.

Trabalho aprovado. São Cristóvão, 15 de Agosto de 2016:

Prof. Dr. Daniel Oliveira Dantas
Orientador

Prof. Dr. Jugurta Rosa Montalvão Filho
Membro

Prof. Dra. Magali Teresinha Longhi
Membro

Resumo

A personalidade pode ser definida como um conjunto de características psicológicas capazes de determinar o padrão de pensar, agir e sentir, sendo este um fator capaz de influenciar diretamente os interesses de um indivíduo. Desde a concepção da primeira descrição de *software*, sistemas computacionais são criados com o objetivo de gerar resultados baseados nos dados de entrada dos usuários, sem se preocupar com quem ele é ou quais são seus interesses. Através do desenvolvimento de um método não intrusivo para coleta de dados do ritmo de digitação dos usuários, e a aplicação de um questionário de autoavaliação de personalidade, este trabalho tem como objetivo identificar abordagens capazes de comprovar a existência da correlação entre tais informações, visando a construção de um conhecimento base para a criação de um *framework* baseado em redes neurais para a extração de traços da personalidade a partir dos dados do ritmo de digitação.

Palavras-chave: Ritmo de digitação; Extração de traços de personalidade; Redes neurais.

Abstract

Personality can be defined as a set of psychological features that may determine the way to think, act and feel, as a factor that may directly influence an individual's interests. Since the creation of the first software's description, computer systems are created with a goal of generate results based on the user's input data, not taking into account who is the user or which are their interests. Developing a non-intrusive approach to collecting keystroke dynamics data from the users, and the use of an auto evaluation questionnaire of personality, this paper was written focused in identify approaches that can prove the correlation between these information, while building a knowledge base to develop a framework based on neural networks to extract personality traits from keystroke dynamics data.

Keywords: *Keystroke; Personality traits extraction; Neural networks.*

Eu dedico este trabalho a todos que de alguma forma me apoiaram e acreditaram em mim, familiares, amigos e professores.

Agradecimentos

Agradeço a todos que estiveram ao meu lado e de alguma forma me apoiaram durante essa jornada sombria e perturbadora. Sem as palavras de apoio e as críticas sobre minha sanidade mental, esse trabalho jamais teria sido concluído. Agradeço aos meus familiares, amigos e professores que acompanharam, e com palavras de apoio ajudaram a suportar a minha angústia se transformar em desespero. Agradeço a todos que não me deixaram desistir, mesmo quando acharam que eu deveria. Agradeço muito ao meu orientador por permitir que este trabalho fosse concluído, arriscando-se a contribuir com um trabalho já em andamento, dando continuidade ao meu sonho de conclusão do mestrado. Agradeço especialmente aos meus pais, meu irmão e a minha esposa por me suportarem no ápice do estresse, por permanecerem ao meu lado e por não me abandonarem, quando eu mesmo já havia abandonado. Sobrevivi.

Omelette du fromage

(Dexter)

Lista de tabelas

Tabela 1: Comparação de outros estudos	13
Tabela 2: Os cinco grandes traços e suas respectivas facetas.....	19
Tabela 3: Intervalos de valor utilizados na classificação dos traços de personalidade	26
Tabela 4: Experimentos realizados e suas respectivas particularidades.....	28
Tabela 5: Teste de hipótese dos experimentos na abordagem com escala Likert	50
Tabela 6: Teste de hipótese dos experimentos na abordagem de três faixas de valor.....	50
Tabela 7: Teste de hipótese dos experimentos na abordagem binária	50

Lista de ilustrações

Figura 1: Breve contextualização para o participante, sobre o objeto da ferramenta de coleta de dados	15
Figura 2: Termo de consentimento livre	15
Figura 3: Cadastro de participantes.....	16
Figura 4: Coleta de dados do ritmo de digitação	17
Figura 5: Exemplos de dados brutos obtidos a partir do ritmo de digitação, representados no formato JSON.....	18
Figura 6: Coleta de dados da personalidade (versões para o gênero masculino e feminino)	20
Figura 7: Representação das respostas de um usuário ao fim do questionário TIPI	20
Figura 8: Exibição do resultado do TIPI, em formato gráfico, ao fim do processo de coleta de dados	22
Figura 9: Breve descrição sobre cada um dos traços de personalidade analisados	22
Figura 10: Estrutura básica de uma rede neural.....	24
Figura 11: Estrutura da rede neural utilizada para mapear saídas na escala Likert	25
Figura 12: Estrutura da rede neural utilizada para mapear saídas classificadas em três faixas de valor	25
Figura 13: Estrutura da rede neural utilizada para mapear saídas binárias.....	25
Figura 14: Exemplificação do nível de confiabilidade e da região crítica do teste de hipótese ..	49

Lista de siglas e abreviações

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
CPU	Central Processing Unit
GPU	Graphics Processing Unit
IHC	Interação Humano Computador
JSON	JavaScript Object Notation
NEO-IPIP	Neuroticism-Extraversion-Openness International Personality Item Pool
RNA	Rede Neural Artificial
TIPI	Ten-Item Personality Inventory

Sumário

1	Introdução	5
1.1	Objetivos	7
1.1.1	Objetivos específicos	7
1.2	Motivação	7
1.3	Estrutura do trabalho	7
2	Descrição do problema.....	7
3	Metodologia	13
3.1	Coleta de dados	13
3.1.1	Extração do ritmo de digitação.....	16
3.1.2	Extração dos traços de personalidade	18
3.2	Pré-processamento	22
3.3	Análise.....	23
4	Experimentos e Resultados	27
4.1	Distribuição de probabilidade entre as classes.....	28
4.1.1	Likert	29
4.1.2	Três faixas.....	31
4.1.3	Binário	34
4.2	Experimento com saídas na escala Likert.....	36
4.3	Experimentos com saídas classificadas em três faixas de valor	39
4.4	Experimentos com saídas binárias	43
4.5	Análise com conhecimento a priori.....	46
4.6	Teste de hipótese.....	48
5	Conclusões	52
6	Trabalhos futuros.....	54
7	Bibliografia.....	55

1 Introdução

A computação afetiva é o ramo da computação inteligente que lida com as características da personalidade e da emoção dos usuários de sistemas computacionais, tendo como objetivo identificar, modelar e implementar a emoção humana em um formato computacional, dando ao sistema a capacidade de reagir adequadamente à personalidade do usuário. A personalidade pode ser definida, dentre outras formas, como um conjunto de características psicológicas capazes de determinar a individualidade de alguém através do padrão de pensar, agir e sentir (PERVIN, CERVONE e JOHN, 2004), possuindo direta influência no comportamento (CARVER E SCHEIER, 2000), sendo algo de conhecimento público, embora ninguém saiba descrevê-la de forma precisa (ALLPORT, 1961). A personalidade já é reconhecida como uma característica relativamente estável em um indivíduo, sujeita a mudanças, porém estável por um período de 45 anos, a partir da fase adulta (NUNES, 2010) e estudada como um fator determinante no comportamento humano, sendo capaz de influenciar diretamente os interesses de um indivíduo, tornando a identificação e a modelagem computacional da personalidade um pré-requisito para a criação novas aplicações, modelos de sistemas e ambientes virtuais personalizáveis (STATHOPOULOU et al, 2010).

Desde a concepção da primeira descrição de um *software* de computador no século XIX até os dias atuais, sistemas computacionais são desenvolvidos para desempenhar um comportamento uniforme, independente do usuário que com ele interage, preocupando-se apenas com as informações brutas representadas pelos dados de entrada. É de conhecimento comum que muitos indivíduos desejam ser tratados de uma forma única/diferenciada e personalizada, a ponto de se sentirem especiais para outras pessoas, grupos ou organizações. Ao habilitar sistemas computacionais para analisar os dados de entrada de um usuário não somente como fonte bruta de informação, mas também como um fonte de metadados capazes de permitir a identificação de indivíduos, e o agrupamento de usuários com comportamentos e interesses semelhantes, estaremos abrindo as portas para a construção de um nova classe de sistemas computacionais, voltados para a personalização da experiência do usuário, adaptando-se de acordo com as características particulares de cada indivíduo. É a partir deste enfoque que este trabalho analisou e realizou experimentos com informações obtidas a partir do ritmo de digitação de usuários de computador, em busca da correlação de tais dados com características das suas personalidades, extraídas a partir da aplicação de um questionário de autoavaliação, visando permitir que o computador enxergue o usuário além das informações por ele diretamente fornecidas.

Segundo (HEWETT et al, 1992; PREECE, 1994; AGNER, 2006) a Interação Humano Computador (IHC) é uma ciência que diz respeito ao *design*, avaliação e implementação de sistemas computacionais interativos para uso humano, através do estudo e definição de métodos e projetos de sistemas que sejam de mais fácil utilização, eficientes e que possibilitem o conforto dos usuários, podendo englobar áreas como Ciência da Computação, Inteligência Artificial e Psicologia. Tradicionalmente, a interação humano computador não apresenta nenhum tipo de troca de informações sobre a personalidade do usuário, de forma que usuários diferentes, com gostos e interesses completamente distintos são tratados de forma homogênea, e sem qualquer tipo de personalização. O problema em questão pode ser verificado na maioria dos sistemas presentes no mercado, pois, após desenvolvidos, esses sistemas sempre retornarão a mesma resposta para qualquer que seja o usuário, independente de quem ele é e/ou quais são seus interesses, levando em consideração apenas os dados brutos por ele informados. Ao

introduzir a possibilidade de personalizar um sistema baseado em características específicas do usuário, estamos permitindo que sistemas adaptativos sejam construídos, voltados para a experiência de interação do usuário, e possibilitando o desenvolvimento de novas ferramentas e pesquisas em torno da adaptação de sistemas computacionais a partir da personalidade do usuário. De acordo com o que foi exposto em (NUNES e HU, 2012; NUNES e CAZELLA, 2011), um dos problemas enfrentados pelos cientistas de Computação Afetiva, é a necessidade de criar soluções capazes de inferir a personalidade de indivíduos de forma não intrusiva e explícita. Diversas abordagens para detectar a emoção do usuário tem sido estudadas, podendo ser bem variadas e muitas vezes custosas (NAHIN et al, 2014). Procedimentos como análise de entonação vocal, sensores fisiológicos conectados à pele, análise de expressões faciais, mapeamento de gestos e poses, análise do ritmo de digitação, utilização do mouse, mineração de texto, entre outros são exemplos de abordagens já realizadas.

Como uma alternativa para a solução do referido problema, este estudo apresenta uma abordagem voltada para a análise do ritmo de digitação de usuários de computador, sendo esta uma abordagem automática, não intrusiva e com custo reduzido para aplicação, com foco na extração de características da personalidade dos usuários, cujos resultados foram comparados com os resultados da aplicação de um questionário de autoavaliação da personalidade, desenvolvido e utilizado na área de Psicologia e baseado em um modelo para representação da personalidade conhecido como *The Big Five Model*, sendo este capaz de identificar características específicas da personalidade de um indivíduo.

Análise do ritmo de digitação ou *Keystroke dynamics* é a denominação para o processo de extrair o padrão biométrico de um indivíduo utilizando a maneira e o ritmo em que ele digita caracteres em um teclado (SHEPHERD, 1995), podendo este padrão ser utilizado para autenticação, identificação e análise de características particulares do usuário. Desde a Segunda Guerra Mundial, quando as transmissões de mensagens eram realizadas pelos militares através de Código Morse, já existia a ideia de identificar/autenticar o emissor de uma mensagem através do ritmo de "digitação", como uma forma de evitar a espionagem. Nessa época, a inteligência militar identificou que um indivíduo possui um jeito único de teclar os "pontos" e "traços" de uma mensagem, criando cada um seu próprio padrão. A metodologia desenvolvida para a identificação desse padrão foi denominada "The Fist of the Sender", tendo esta sido aplicada para distinguir mensagens enviadas por aliados e/ou inimigos. Na abordagem atual, os dados brutos do ritmo de digitação (equivalentes aos "pontos" e "traços") estão disponíveis a partir de qualquer teclado convencional de computador, podendo estes serem facilmente extraídos em busca de informações sobre o *Hold Time* e o *Up Time*.

O *Hold time*, que muitas vezes é encontrado na literatura como *DWell time* ou *Down time*, representa o intervalo de tempo decorrido entre as ações de pressionar e soltar uma mesma tecla de um teclado. Já o *Up time*, também descrito como *Flight time* ou *Up-down time*, representa o tempo decorrido entre soltar a tecla atual e pressionar a próxima. A partir de um conjunto de exemplos dessas duas características, através de diferentes teclas do teclado, é possível determinar o padrão de digitar de um indivíduo, e a partir deste padrão buscar a correlação com as características de sua personalidade.

1.1 Objetivos

O principal objetivo deste trabalho é construir a base de conhecimento necessária para o desenvolvimento de um *framework* que visa simplificar a utilização de traços da personalidade na personalização de sistemas computacionais, identificando abordagens capazes de efetuar a correlação entre o ritmo de digitação de um indivíduo, e as características de sua personalidade, abrindo caminho para o desenvolvimento de novas ferramentas e pesquisas, com o intuito de popularizar a utilização de traços de personalidade em sistemas computacionais.

1.1.1 Objetivos específicos

1. Implementar uma ferramenta para extrair e armazenar as características do ritmo de digitação;
2. Adaptar o questionário de autoavaliação para ser aplicado em conjunto com a extração do ritmo de digitação;
3. Aplicar a ferramenta a um grupo de indivíduos;
4. Aplicar um questionário de autoavaliação imediatamente após a extração das características do ritmo de digitação;
5. Criar uma base de dados unificada para armazenar os dados do ritmo de digitação e o resultado do questionário de autoavaliação de cada indivíduo;
6. Avaliar *frameworks* de *machine learning* existentes e a viabilidade de utilizá-los para a construção de uma rede neural que possa classificar corretamente os dados extraídos;
7. Construir a rede neural e realizar experimentos com os dados obtidos nos passos 3 e 4;
8. Validação e análise dos resultados dos experimentos;

1.2 Motivação

Simplificar o processo necessário para utilizar traços da personalidade de usuários na personalização de sistemas computacionais, e consequentemente popularizar a construção de uma nova classe de sistemas, capazes de adaptar seu comportamento do acordo com o indivíduo que o utiliza, através da aplicação de um método de inferência de personalidade automático e não intrusivo.

1.3 Estrutura do trabalho

Além da introdução previamente apresentada, este trabalho está estruturado da seguinte forma: A seção 2 apresenta uma descrição detalhada sobre o problema abordado durante o desenvolvimento da pesquisa. A seção 3 refere-se à metodologia, que destrinchará as particularidades sobre cada uma das abordagens utilizadas, os experimentos construídos e como todo o trabalho foi desenvolvido. Na seção 4 são apresentados os experimentos realizados e seus respectivos resultados, e por fim as seções 5 e 6 apresentam a conclusão e os trabalhos futuros.

2 Descrição do problema

A personalização de sistemas computacionais tem se tornado um tema recorrente nos últimos anos, acompanhando o crescimento da área de Computação Afetiva, que desde a sua formalização no século XX através da publicação "*Affective Computing*" (PICARD, 1995), vem discutindo conceitos da computação que estão relacionados, tem origem ou são influenciados pelas emoções.

A demanda por dar ao computador a capacidade de identificar, interpretar e responder adequadamente a um usuário de acordo com suas características individuais tem se destacado como um importante passo na evolução da interação humano computador, assim como destacam (KHAN et al, 2008; NAHIN et al, 2014; SOLANKI e SHUKLA, 2014) ao afirmarem que as emoções possuem um papel significativo neste processo. Diversos estudos têm sido realizados na pretensão de mapear e identificar os traços da personalidade, emoção ou estado emocional de um indivíduo através de diferentes abordagens, principalmente focadas no desenvolvimento de formas não intrusivas de obter tais informações. Nessa seção são apresentados alguns trabalhos relacionados ao tema, como forma de descrever o que vem sendo desenvolvido nessa área de estudo, e como ela tem evoluído.

A descrição desses estudos é iniciada com o trabalho de (GILL e OBERLANDER, 2003), que através da análise/julgamento realizada(o) por 30 participantes da pesquisa, buscou identificar a possibilidade de mensurar traços específicos da personalidade, como extroversão e neuroticismo, a partir de mensagens de e-mail, obtendo taxas de acerto relativamente altas na identificação de características da extroversão. Três anos depois, (GOSLING, MHEL e PENNEBAKER, 2006) utilizaram gravadores para analisar sons ambientes do cotidiano de 96 indivíduos por dois dias, na tentativa de determinar como a personalidade se manifesta no cotidiano das pessoas.

(KHAN et al, 2008) desenvolveu uma ferramenta com o intuito de extrair os traços de personalidade de indivíduos, através do ritmo de digitação e utilização do mouse, tendo como base para seus experimentos, teorias sobre a correlação entre a extroversão e a velocidade de movimentação do homem, e também estudos que identificaram que extrovertidos interagem com a interface de um sistema de forma mais veloz que um introvertido. Como resultado de seus estudos, (KHAN et al, 2008) concluiu ser possível mensurar a personalidade de um usuário através do seu modo de usar o teclado e o mouse. Nesse estudo (KHAN et al, 2008) aplicou uma versão reduzida do NEO-IPIP (Neuroticism-Extraversion-Openness International Inventory Item Pool) como forma de autoavaliação dos participantes, sendo este questionário composto por 120 questões, o que segundo o estudo demandava entre 15 e 25 minutos do participante para responder completamente o questionário. Por ter utilizado o coeficiente de correlação de Pearson como abordagem para correlacionar as informações extraídas, as taxas de sucesso desse estudo é expressa em valores que variam entre -1 e 1, onde o valor central 0 indica ausência de correlação e os extremos -1 e 1 indicam correlação perfeita (negativa e positiva, respectivamente). Os resultados obtidos por (KHAN et al, 2008) para os cinco traços de personalidade mapeados variaram entre -0.4 e -0.56, o que de acordo com a escala de Pearson¹ representam respectivamente correlações "fraca" e "moderada".

(KHANNA e SASIKUMAR, 2010) Propuseram uma abordagem para detectar as emoções de usuários de computador a partir de dados obtidos da utilização de um teclado convencional, através do estudo de características mais específicas, como velocidade de digitação, moda, desvio padrão, quantidade de *backspaces*, entre outras, obtidas da análise do *hold time* e *up time* de cada indivíduo. O estudo por eles realizado contou com 300 participantes (45% mulheres e 55% homens), com idades entre 21 anos e 41+ anos, e obteve taxas de sucesso na identificação de emoções positivas e negativas de até 88,88% e 89,02% respectivamente.

¹ Os valores obtidos através da correlação de Pearson podem indicar correlação nos seguintes níveis: nula, ínfima, fraca, moderada, forte e perfeita.

A utilização dos dados do ritmo de digitação como fonte de informação para detecção do estado emocional também foi tema do estudo realizado por (EPP, LIPPOLD e MANDRYK, 2011), que contando com apenas 12 participantes (10 homens e 2 mulheres), mapeou 15 estados emocionais distintos com taxas de sucesso que variaram entre 77,40% e 87,80% . Já (COSTA e PORTO, 2011) apresentaram como resultado final de seu trabalho, uma ferramenta denominada PersonaliKey v1.0 cujo objetivo é auxiliar no processo de extração de personalidade de usuários, através do ritmo de digitação quando aplicada em conjunto com o questionário NEO-IPIP. Por se tratar de uma abordagem semelhante, os resultados obtidos por (COSTA e PORTO, 2011) foram utilizados como parâmetro de comparação em alguns dos experimentos realizados.

Em (ARYA et al, 2012) é descrita uma abordagem para prever emoção dos usuário de redes sociais a partir do conteúdo por eles produzido, utilizando técnicas de análise e mineração de texto. Seus experimentos foram realizados com aproximadamente 50 pessoas próximas aos autores, das quais eles afirmam ter conseguido obter similaridades notáveis em relação aos respectivos traços de personalidade.

(ZIMMERMANN et al, 2013) apresentam uma abordagem semelhante a (KHAN et al, 2008) ao utilizar simultaneamente dados extraídos do teclado e do mouse, embora os estudos de (ZIMMERMANN et al, 2013) não foquem diretamente na extração de personalidade, e sim na mensuração do humor de forma mais abrangente, com o intuito de melhorar a interação humano computador, sem se preocupar se o humor atual é oriundo de uma características permanentes da personalidade ou algo momentâneo ligado ao estado emocional do indivíduo.

No mesmo ano, (KOLAKOWSKA et al, 2013) desenvolveram um estudo focado na influência do estado emocional de programadores na produtividade e na qualidade do código desenvolvido, analisando estados emocionais como: (i) Excitação; (ii) Empoderamento; (iii) Engajamento; (iv) Desencorajamento; (v) Desgosto; (vi) Frustração; (vii) Tédio; (viii) Surpresa; Afirmando, ao fim de sua pesquisa, que o efeito da emoção sobre a qualidade do trabalho é inquestionável.

Em apoio à utilização de *keystroke* como método de extração de dados do usuário, (NAHIN et al, 2014) justificam que o teclado de computador, mesmo sendo um equipamento barato, ainda assim permite a comunicação entre humanos e computadores. Em seu trabalho, (NAHIN et al, 2014) utilizaram-se de sete classes de emoções predefinidas (alegria, medo, raiva, tristeza, culpa, vergonha e nojo), com o intuito de detectar a variação no comportamento de usuários de computador durante o processo de transição de um estado emocional para outro. O trabalho por eles desenvolvido foi feito de forma que nenhum hardware adicional além de um teclado convencional de computador fosse necessário. Para a realização de seu estudo, (NAHIN et al, 2014) definiram duas abordagens distintas: (i) Utilização de texto fixo; (ii) Utilização de texto livre; Não tendo sido apenas analisado o ritmo de digitação do usuário, mas também uma análise do texto por ele produzido. A fase de coleta de dados foi realizada com apenas 25 voluntários, cujas idades variaram entre 15 e 40 anos, sendo aproximadamente 45% dos participantes mulheres e 55% homens. Uma etapa de pré-processamento foi realizada nos dados coletados, visando extrair 19 características do ritmo de digitação (velocidade de digitação, moda, desvio padrão, variância, entre outras), que posteriormente foram aplicadas ao *software*

de *data mining* Weka². Como resultado de seu trabalho, (NAHIN et al, 2014) obtiveram entre 60% e 87% de sucesso na identificação das referidas emoções.

SOLANKI e SHUKLA também desenvolveram sua própria abordagem no esforço de extrair o estado emocional de indivíduos utilizando-se dos dados do ritmo de digitação e sua correlação com questionários de autoavaliação. Em seu trabalho (SOLANKI e SHUKLA, 2014) tinham como objetivo identificar o estado emocional de um usuário baseado no uso de um teclado convencional de computador, focando na identificação das emoções: confiança, tristeza, felicidade, cansaço, nervosismo e raiva. Assim como (NAHIN et al, 2014), o trabalho por eles desenvolvido utilizou-se de duas abordagens distintas: (i) Extração dos dados produzidos por um indivíduo ao digitar um texto fixo por eles predefinido; (ii) Extração dos dados produzidos por um indivíduo ao digitar um texto livre; Ao fim do estudo, constatou-se que ambos os experimentos obtiveram resultados positivos na identificação das classes de emoções selecionadas, sendo o primeiro experimento (com utilização de texto fixo predefinido) mais preciso na identificação de algumas das emoções.

Diferente dos demais trabalhos, (CHUDÁ e KRÁTKY, 2014) desenvolveram uma abordagem de extração de traços da personalidade baseando-se apenas nos dados obtidos da utilização do mouse por um indivíduo em um ambiente Web, sendo o estudo por eles realizado aplicado em um site de comércio eletrônico já em funcionamento. Em outro de seus estudos, (KHAN et al, 2015) analisou 47 indivíduos com o objetivo de identificar a personalidade de programadores, através da interação com teclado e mouse e da aplicação de questionários de autoavaliação. Ao fim de seu estudo, (KHAN et al, 2015) afirmou ser possível diferenciar bons programadores de programadores não tão bons de forma objetiva através da correlação dos dados por eles produzidos.

Já o projeto *Personality Insights* da IBM permite que outras organizações utilizem a ferramenta por eles desenvolvida para extração de traços de personalidade através da utilização de técnicas de mineração de texto, utilizando o *Big Five* como modelo para representação da personalidade. A IBM possibilita que qualquer texto produzido por um usuário seja enviado para a ferramenta *Personality Insights*, para que ela retorne como resposta uma árvore contendo informações cognitivas e características sociais (incluindo os fatores do *Big Five*).

Cada uma das abordagens previamente descritas está representada na Tabela 1, visando expor as principais características analisadas por este trabalho, de forma que justifique a realização de um novo estudo além dos já existentes.

	Extração de características através do ritmo de digitação	Disponibilização das ferramentas para uso e adaptação de terceiros	Utilização do <i>Big Five Model</i>	Não intrusivo	Sem a necessita de equipamentos especiais
<i>Perception of e-mail personality at zero-acquaintance extraversion takes</i>	NÃO	NÃO	SIM	NÃO	SIM

² Ferramenta formada por um conjunto de algoritmos de aprendizagem de máquina voltados para mineração de dados, disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>

<i>care of itself; neuroticism is a worry (GILL e OBERLANDER, 2003)</i>					
<i>Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life (GOSLING, MEHL e PENNEBAKER, 2006)</i>	NÃO	NÃO	SIM	NÃO	NÃO
<i>Measuring personality from keyboard and mouse use (KHAN, I. A. et al, 2008)</i>	SIM	NÃO	SIM	SIM	SIM
<i>Biometric mouse intelligent system for student's emotional and examination process analysis (KAKLAUSKAS, KRUTINIS e SENIUT, 2009)</i>	NÃO	NÃO	NÃO	SIM	SIM
<i>Recognising emotions from keyboard stroke pattern (KHANNA, SASIKUMAR, 2010)</i>	SIM	NÃO	NÃO	SIM	SIM
<i>Identifying emotional states using keystroke dynamics (EPP, LIPPOLD e MANDRYK, 2011)</i>	SIM	NÃO	NÃO	SIM	SIM
<i>PersonaliKey (COSTA e PORTO, 2011)</i>	SIM	NÃO	SIM	SIM	SIM
<i>A Text Analysis Based Seamless Framework for Predicting Human Personality Traits from Social Networking Sites</i>	NÃO	NÃO	NÃO	SIM	SIM

(ARYA et al, 2012)					
<i>Affective computing—a rationale for measuring mood with mouse and keyboard</i> (ZIMMERMANN et al, 2013)	SIM	NÃO	NÃO	SIM	SIM
<i>Emotion recognition and its application in software engineering</i> (KOLAKOWSKA et al, 2013)	SIM	NÃO	NÃO	SIM	SIM
<i>Estimation of the User's Emotional State by Keystroke Dynamics</i> (SOLANKI e SHUKLA, 2014)	SIM	NÃO	NÃO	SIM	SIM
<i>Identifying emotion by keystroke dynamics and text pattern analysis</i> (NAHIN et al, 2014)	SIM	NÃO	NÃO	SIM	SIM
<i>Usage of computer mouse characteristics for identification in web browsing</i> (CHUDÁ, KRÁTKY, 2014)	NÃO	NÃO	NÃO	SIM	SIM
<i>Under pressure: sensing stress of computer users</i> (HERNANDEZ et al, 2014)	SIM	NÃO	NÃO	SIM	SIM
<i>Your mouse can tell about your emotions</i> (LALI, P. et al, 2014)	NÃO	NÃO	NÃO	SIM	SIM
<i>Predicting programmers' personality via interaction behaviour with keyboard and mouse</i> (KHAN et al, 2015)	SIM	NÃO	SIM	SIM	SIM

Watson (IBM)	NÃO	NÃO	SIM	SIM	SIM
Presente trabalho	SIM	SIM	SIM	SIM	SIM

Tabela 1: Comparação de outros estudos

3 Metodologia

O desenvolvimento deste trabalho teve início com uma revisão bibliográfica sobre os demais trabalhos desenvolvidos por outros pesquisadores, em busca de abordagens semelhantes e/ou técnicas que pudessem maximizar os resultados da abordagem selecionada. Partindo do objetivo de identificar a correlação entre o ritmo de digitação de um indivíduo e seus traços de personalidade, tornou-se necessária a construção de uma base de dados contendo características do ritmo de digitação de diversos indivíduos, assim como dados referentes às características da personalidade de cada um deles.

A metodologia deste trabalho está organizada da seguinte forma: (i) Coleta de dados, sendo esta etapa a responsável pela aquisição das informações base para a realização deste estudo, através da identificação da forma de digitar dos participantes, em conjunto com a aplicação de um questionário de autoavaliação; (ii) Pré-processamento, etapa onde foram realizadas equalizações e conversões dos dados obtidos, para adequá-los a cada uma das abordagens de análise; (iii) Análise, etapa que descreve as abordagens selecionadas para a realização deste trabalho.

3.1 Coleta de dados

A coleta de dados foi realizada de forma totalmente *online*, visando ser acessada pelo maior número possível de participantes, através de uma página web³ construída para extrair os dados brutos do ritmo de digitação dos participantes, a partir de um teclado convencional de computador, e em seguida realizar a aplicação do questionário de autoavaliação TIPI.

Essa página ficou disponível para a coleta de dados dos experimentos por um período de 56 dias, contando com um total de 177 participantes, sendo 56 do gênero feminino e 121 do gênero masculino, cujas idades variam entre 12 e 46 anos, tendo 24,83 anos como a idade média dos participantes. O Gráfico 1 e o Gráfico 2 apresentam respectivamente a distribuição de participantes por gênero e por idade. Todo o processo de coleta de dados realizado *online* foi dividido em três etapas, sendo (i) Contextualização, consentimento livre e cadastro do participante; (ii) Extração do ritmo de digitação; (iii) Extração dos traços de personalidade.

³ A página web utilizada no processo de coleta de dados pode ser acessada através do endereço <http://www.felipegoulart.com/personalityapi>

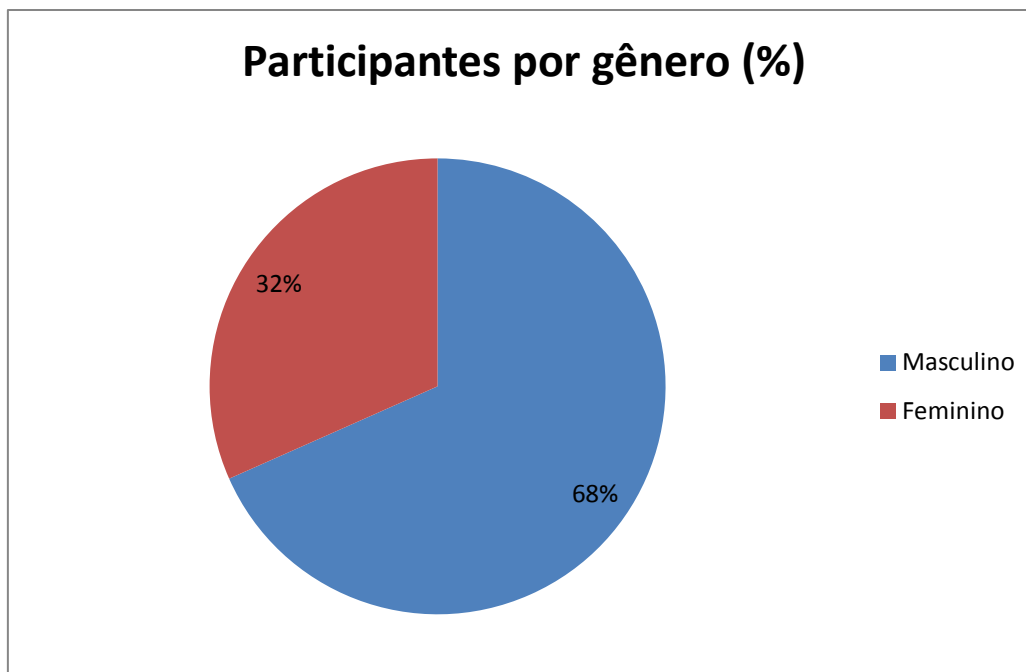


Gráfico 1: Percentual dos participantes da pesquisa por gênero

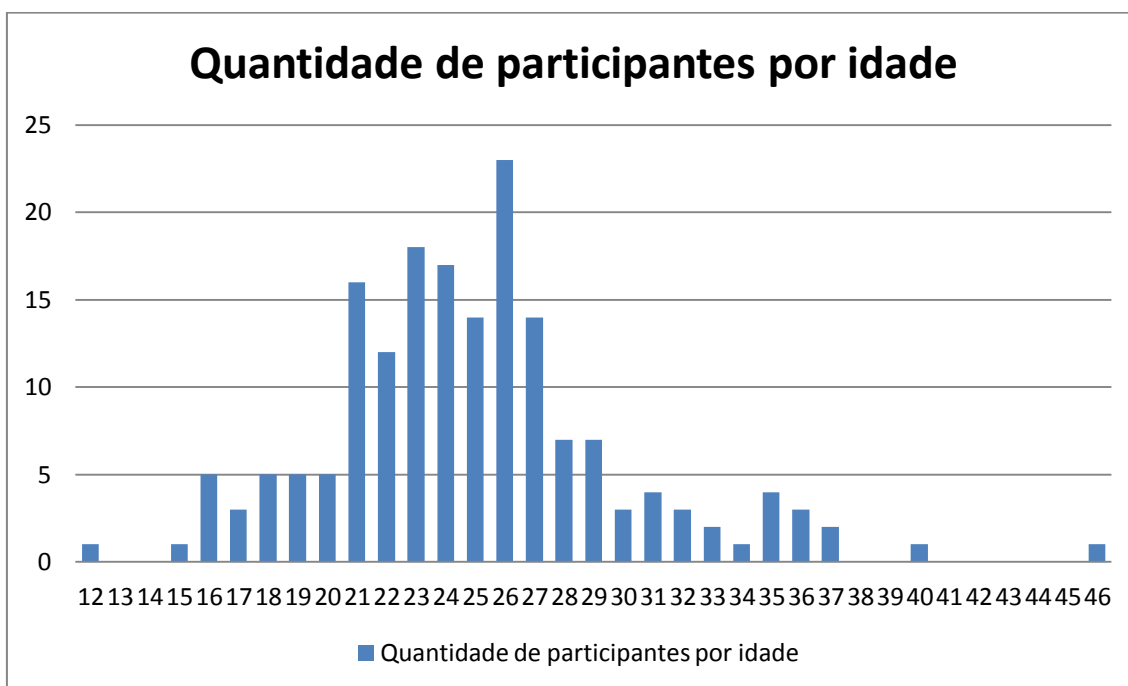


Gráfico 2: Quantidade de participantes por idade

A etapa introdutória do processo de coleta de dados teve início com a apresentação do objetivo da coleta de dados, e uma breve contextualização sobre quais seriam os próximos passos e o que deveria ser feito nas etapas seguintes do processo de coleta de dados, exatamente como mostra a Figura 1.

Objetivo da pesquisa

Esta pesquisa tem como objetivo coletar dados da personalidade de usuários de computador, onde os participantes deverão digitar um texto predefinido (**2 linhas**) e responder um questionário de autoavaliação (**10 questões**).

Entendi

Figura 1: Breve contextualização para o participante, sobre o objeto da ferramenta de coleta de dados

Após contextualizado, foi apresentado para o participante um termo de consentimento livre, como forma de garantir a ciência do participante sobre o sigilo de suas informações e que ele está de acordo com os fins para os quais os dados fornecidos seriam utilizados, deixando claro a possibilidade de desistir a qualquer momento da pesquisa, como mostra a Figura 2.

Para fins de análises estatísticas e registro dos participantes, caso um contato posterior à realização da pesquisa fosse necessário, foi construído um cadastro simplificado de forma que não tomasse muito tempo do participante, e sem a obrigatoriedade de ser completamente preenchido, assim como mostra a Figura 3.

Termo de consentimento livre

Declaro que fui devidamente esclarecido sobre o objetivo da pesquisa que está sendo desenvolvida, que tenho conhecimento sobre o sigilo do meu nome em relação aos dados por mim fornecidos ou que serão coletados. Estou ciente também do meu direito de recusa e desistência em qualquer momento da pesquisa. Dessa forma, desejo voluntariamente participar deste estudo, e autorizo a utilização dos dados por mim produzidos para os fins aos quais se destina a pesquisa, podendo estes serem posteriormente divulgados entre a comunidade científica.

Eu aceito o termo de consentimento e desejo continuar na pesquisa.

Aceitar

Figura 2: Termo de consentimento livre

Cadastro

Nome:

E-mail:

Idade: *Campo obrigatório

*Campo obrigatório

Masculino Feminino

Com o intuito de garantir sua privacidade os dados de nome e e-mail serão utilizados apenas para contatos futuros sobre a pesquisa, não sendo estes publicados ou compartilhados com terceiros.

Figura 3: Cadastro de participantes

O tempo médio de permanência dos participantes na página da coleta de dados foi de 11,27 minutos, enquanto individualmente as etapas de digitação e de aplicação do questionários tiveram duração média de 1,97 minutos e 1,61 minutos respectivamente, o que indica uma duração média de 3,58 minutos para a realização das duas principais etapas de coleta de dados.

3.1.1 Extração do ritmo de digitação

Ao fim do processo inicial de contextualização, obtenção do consentimento e cadastro do participante, a etapa de coleta de dados é efetivamente iniciada com a apresentação de um texto fixo, predefinido, e igual para todos os participantes, instruindo-os a digitar o texto exatamente como apresentado, por duas vezes seguidas em um campo de texto, assim como apresentado na Figura 4.

O texto "**ao digitar este pequeno texto estou auxiliando no desenvolvimento de uma pesquisa, permitindo que os dados do meu modo de digitar possam ser analisados para o desenvolvimento de novas tecnologia**", escolhido para a coleta de dados foi elaborado de forma que não fosse necessária a utilização de "teclas especiais" (ex: *shift* e *caps lock*) para a capitalização de letras nem adição de acentuação, contendo apenas um símbolo referente à pontuação, não pertencendo a um contexto ou nicho específico, e composto por um vocabulário simples, com palavras de uso comum e presentes no dia-a-dia da maior parte das pessoas. Todo o cuidado na elaboração do texto foi oriundo da preocupação em reduzir as discrepâncias entre amostras de diferentes usuários, que poderiam ser causadas pelo uso de palavras pouco conhecidas ou por um texto de difícil compreensão.

Ritmo de digitação

Por favor, digite o texto abaixo exatamente como é apresentado:

ao digitar este pequeno texto estou auxiliando no desenvolvimento de uma pesquisa, permitindo que os dados do meu modo de digitar possam ser analisados para o desenvolvimento de novas tecnologias

ao digitar este pequeno texto estou auxiliando no desenvol|

Enviar

1/2

Figura 4: Coleta de dados do ritmo de digitação

A partir deste processo, é possível identificar um ritmo de digitação para cada participante, sendo este formado por um conjunto de características extraídas de cada uma das teclas pressionadas durante o processo de coleta de dados. Tais características são conhecidas como *Hold time* (H), *Up time* (U) e *Down-Down time* (DD), e respectivamente representam o tempo que uma tecla ficou pressionada, o tempo decorrido entre soltar a tecla atual e o intervalo de tempo decorrido entre pressionar a tecla atual e pressionar a próxima tecla.

A extração de tais características foi realizada através do mapeamento dos eventos de teclado *keydown* e *keyup*, de forma que pudessem ser armazenadas informações das teclas que foram pressionadas/soltas, e o momento em que cada um desses eventos aconteceu. A Figura 5 mostra exemplos de como essas informações são organizadas após capturadas, onde cada tecla é representada por um *keyCode* (código que identifica a tecla na tabela ASCII⁴), e seus respectivos *hold time* e *up time*.

⁴ ASCII ou Código padrão americano para intercâmbio de informação é um dos padrões para representação de caracteres alfanuméricos e símbolos em sistemas computacionais.

```

{
  "user": {
    ...
    "keystroke01": {
      "keys": [{
        "keyCode": 65,
        "holdTime": 167,
        "upTime": 1
      },
      ...
      {
        "keyCode": 73,
        "holdTime": 105,
        "upTime": 23
      }
    ]
  },
  "keystroke02": {
    "keys": [{
      "keyCode": 65,
      "holdTime": 152,
      "upTime": 40
    },
    ...
    {
      "keyCode": 73,
      "holdTime": 104,
      "upTime": 0
    }
  ]
}
}

```

Figura 5: Exemplos de dados brutos obtidos a partir do ritmo de digitação, representados no formato JSON⁵

3.1.2 Extração dos traços de personalidade

The Big Five Model, ou os Cinco Grandes Traços, é um modelo construído para representação da personalidade de um indivíduo, tendo sido criado na área de Psicologia, e é atualmente o modelo cientificamente mais aceito e vastamente estudado. Este modelo de representação da personalidade classifica a personalidade como sendo composta por cinco grandes traços que determinam cada uma das características de ser, pensar e agir de um indivíduo, que ao serem combinados permitem diferenciar pessoas através de um mapeamento de características da personalidade.

Os cinco grandes traços são divididos em: (i) abertura; (ii) realização; (iii) extroversão; (iv) socialização; (v) neuroticismo. Para que seja possível realizar uma análise mais precisa e detalhada da personalidade, esses cinco grandes traços podem ser divididos em subconjuntos de características conhecidas como facetas, que permitem a identificação de informações mais específicas sobre a personalidade de um indivíduo. As facetas de cada um dos traços de personalidade são apresentadas na Tabela 2.

Traço da personalidade (Modelo <i>Big Five</i>)	Facetas
Extroversão	Amigabilidade

⁵ JSON, ou *JavaScript Object Notation* é um dos formatos padrão para representação e transmissão de dados entre aplicações *web*.

	Gregarismo
	Assertividade
	Nível de atividade
	Procura por excitação
	Bom humor
Socialização	Confiança
	Moralidade
	Altruísmo
	Cooperação
	Modéstia
	Compaixão
Realização	Autoeficácia
	Ordem
	Senso de dever
	Empenho
	Autodisciplina
	Prudência
Neuroticismo	Ansiedade
	Raiva
	Depressão
	Autopercepção
	Falta de moderação
	Vulnerabilidade
Abertura	Imaginação
	Interesses artísticos
	Emotividade
	Senso aventureiro
	Intelecto
	Liberalismo

Tabela 2: Os cinco grandes traços e suas respectivas facetas

A utilização do modelo *Big Five* na abordagem de extração de traços de personalidade através do ritmo de digitação é realizada com o intuito de identificar a relação entre uma determinada característica do ritmo de digitação e um dos traços de personalidade propostos pelo modelo. Durante a realização dos experimentos, posteriormente apresentados na seção 4, foram definidas formas distintas de avaliar tal correlação, visando que ao fim deste trabalho seja possível responder a duas questões: **Q1:** É possível mensurar quanto um determinado traço de personalidade está presente na personalidade de um indivíduo? e **Q2:** É possível determinar quais traços de personalidade se destacam na personalidade do indivíduo?

Uma das formas mais tradicionais de extração dos traços de personalidade de um indivíduo é a aplicação de um questionário contendo uma série de perguntas que ajudam a mapear características específicas de cada personalidade. Existem diferentes questionários baseados no *The Big Five Model*, como é o caso do *Neo-International Personality Item Pool* (NEO-IPIP), composto por 300 questões distribuídas igualmente entre os cinco grandes traços. A aplicação desse questionário é realizada de forma que o participante tenha que escolher a resposta à qual mais se adequa, dentre as cinco alternativas disponíveis, que variam de acordo com a escala Likert (LIKERT, 1932), indo de "concordo totalmente" até "discordo totalmente". Devido à quantidade de perguntas que o compõem, é fácil perceber que a aplicação do NEO-IPIP demanda uma grande parcela de tempo, o que acaba resultando em respostas não precisas

devido ao cansaço do participante, ou mesmo questionários incompletos por questões de desistência, dificultando ainda mais o processo de extração dos traços de personalidade.

Com o intuito de reduzir a parcela de tempo necessária para a aplicação de um questionário de autoavaliação, (GOSLING, RENTFROW e SWANN, 2003) desenvolveram o *Ten Item Personality Inventory* (TIPI), um questionário de autoavaliação derivado do NEO-IPIP e composto por apenas 10 questões. Devido ao seu tamanho reduzido, o TIPI não é capaz de extrair características tão precisas de todas as facetas (do *Big Five*) quanto o NEO-IPIP, porém todos os cinco grandes traços são passíveis de serem mapeados. Sua aplicação é realizada de forma similar ao NEO-IPIP, onde o participante deve escolher a resposta à qual mais se adequa dentre as sete alternativas disponíveis, seguindo a escala Likert de 7 pontos, que também varia de "concordo totalmente" até "discordo totalmente".

Neste trabalho, o questionário TIPI foi utilizado como abordagem para extração da personalidade dos participantes, e conseqüentemente adaptado para se adequar à ferramenta de coleta de dados desenvolvida, de forma que as 10 perguntas presentes no questionário original foram apresentadas de forma sequencial, juntamente com suas respectivas alternativas de resposta, e apenas um botão "próxima", para levar o participante da pesquisa até a próxima pergunta, sem permitir que ele pudesse alterar as respostas das perguntas anteriores. A forma como as perguntas do TIPI foram apresentadas para os participantes é exibida na Figura 6.

The figure shows two side-by-side screenshots of the TIPI questionnaire interface. Both screens are titled "Questionário de autoavaliação" and "Ten-Item Personality Inventory (TIPI)".

The left screen is for a male user. It shows "1/10" and the question "Eu me vejo como um homem...". The trait is "Extrovertido, entusiasta." The response options are:

- Discordo Totalmente
- Discordo Parcialmente
- Discordo um pouco
- Nem discordo nem concordo
- Concordo um pouco
- Concordo Parcialmente
- Concordo Totalmente

 A blue "Próxima" button is at the bottom right.

The right screen is for a female user. It shows "1/10" and the question "Eu me vejo como uma mulher...". The trait is "Extrovertida, entusiasta." The response options are:

- Discordo Totalmente
- Discordo Parcialmente
- Discordo um pouco
- Nem discordo nem concordo
- Concordo um pouco
- Concordo Parcialmente
- Concordo Totalmente

 A blue "Próxima" button is at the bottom right.

Figura 6: Coleta de dados da personalidade (versões para o gênero masculino e feminino)

```
{
  "user": {
    ...
    "tipi": {
      "answers": [5, 4, 7, 6, 6, 7, 6, 5, 6, 2]
    },
    ...
  }
}
```

Figura 7: Representação das respostas de um usuário ao fim do questionário TIPI

Por ser composto por 10 questões, o TIPI destina duas questões para cada traço de personalidade. Sendo a_x o valor da resposta selecionada pelo participante para a pergunta de número "x", o valor final de cada um dos traços de personalidade pode ser calculado ao fim do questionário da seguinte forma:

- Extroversão = $\left(\frac{a_1 + R(a_6)}{2}\right)$
- Socialização = $\left(\frac{R(a_2) + a_7}{2}\right)$
- Realização = $\left(\frac{a_3 + R(a_8)}{2}\right)$
- Neuroticismo = $\left(\frac{R(a_4) + a_9}{2}\right)$
- Abertura = $\left(\frac{a_5 + R(a_{10})}{2}\right)$

Onde $R(x) = (8 - x)$ representa o inverso da resposta dada pelo participante. A Figura 7 (acima) apresenta a forma como são armazenadas as respostas de um determinado participante após o fim da aplicação do questionário TIPI.

Ao fim da coleta de dados, o resultado obtido através da aplicação do questionário TIPI é apresentado ao participante, no formato de um gráfico de pizza, destacando o traço de personalidade que foi mapeado com o maior valor dentre os cinco traços analisados. Além de exibir o valor percentual obtido em cada um dos traços de personalidade, uma breve descrição das características de cada um dos traços de personalidade também é apresentada, assim como mostra a Figura 8 e a Figura 9.



Figura 8: Exibição do resultado do TIPI, em formato gráfico, ao fim do processo de coleta de dados

Neuroticismo (ou instabilidade emocional) é a tendência para experimentar emoções negativas, como raiva, ansiedade ou depressão. Aqueles com um grau elevado de neuroticismo são emocionalmente reactivos e vulneráveis ao stress.

Extroversão é caracterizada por emoções positivas e pela tendência para procurar estimulação e a companhia dos outros. Este traço é marcado pelo profundo envolvimento com o mundo exterior. Os extrovertidos gostam de estar com pessoas, e são usualmente vistos como sendo cheios de energia.

Socialização é a tendência para ser compassivo e cooperante em vez de suspeito e antagonista face aos outros. Este traço reflete diferenças individuais na preocupação com a harmonia social. Eles são geralmente respeitosos, amigáveis, generosos, prestáveis e dispostos a fazer compromissos. Pessoas "amigáveis" têm também uma visão otimista da natureza humana. Elas acreditam que as pessoas são basicamente honestas, decentes e dignas de confiança.

Realização é a tendência para mostrar autodisciplina, orientação para os deveres e para atingir os objetivos. Este traço mostra uma preferência pelo comportamento planeado em vez do espontâneo.

Abertura é o interesse pela arte, emoção, aventura, ideias fora do comum, imaginação, curiosidade, e variedade de experiências. Este traço distingue as pessoas imaginativas das convencionais. As pessoas com elevada abertura são intelectualmente curiosas, apreciadoras da arte, e sensíveis à beleza. Elas tendem a ser, comparadas com as pessoas "fechadas", mais criativas, a prestar mais atenção aos seus sentimentos e a terem opiniões não convencionais.

Figura 9: Breve descrição sobre cada um dos traços de personalidade analisados

3.2 Pré-processamento

A etapa de pré-processamento consiste na remoção de dados irrelevantes, tais como informações repetidas e/ou pouco representativas, e a realização de transformações nos dados, de forma que o resultado gerado pelo pré-processamento seja um conjunto de dados mais representativo que o original. Para a realização dos experimentos deste trabalho, a abordagem de pré-processamento utilizada é a descrita em (MONTALVÃO FILHO e FREIRE, 2006), com

o objetivo de equalizar os dados do ritmo de digitação para o mapeamento de traço da personalidade, tendo em vista que essa mesma abordagem mostrou-se vantajosa quando aplicada a dados de ritmo de digitação com fins de identificação biométrica. tal função de equalização é descrita da seguinte forma:

$$g(x) = \frac{1}{1 + \exp\left(-\frac{K(\log_e(x) - \mu_y)}{\sigma_y}\right)}$$

Onde $K = 1.7$, sendo esta uma constante da equação. Além da equação apresentada anteriormente, alguns dos experimentos realizados requereram a utilização de métodos de pré-processamento, voltados para a adaptação da forma como os dados são representados. Um desses métodos de pré-processamento foi a normalização dos resultados obtidos através da aplicação do questionário TIPI entre os valores 0 e 100, e posteriormente classificados em três categorias distintas, que representam valores considerados de nível "baixo", "médio" ou "alto", como forma de comparar o resultado desse experimento, com os resultados obtidos nos estudos realizados por (COSTA e PORTO, 2011), onde os resultados foram tratados de mesma forma.

De forma similar aos experimentos que utilizaram um processo de normalização e classificação, outro conjunto de experimentos necessitou de um processo de *binarização* nos dados obtidos através do questionário TIPI, para a sua execução.

Esse processo de *binarização* foi realizado de forma que valores maiores ou iguais a 5 foram convertidos para o valor binário 1, enquanto os valores inferiores a 5 foram convertidos para o valor binário 0. Tal conversão foi realizada levando em conta que os resultados originais desses experimentos (antes de *binarização*), estavam representados na escala Likert de sete pontos, cujos valores variam entre 1 e 7. De acordo com a escala Likert, valores maiores ou iguais a 5 são considerados altos, valores inferiores a 4 são considerados baixos e o valor 4 é considerado como neutro/não conclusivo.

3.3 Análise

Redes neurais artificiais (RNAs) são modelos computacionais inspirados no sistema nervoso de seres vivos, que possuem a capacidade de aquisição e manutenção do conhecimento, através da utilização de um conjunto de unidades de processamento, conhecidos como neurônios artificiais, que são interligados por um grande número de interconexões chamadas de sinapses artificiais. Para (HAYKIN, 1999; SILVA, SPATTI e FLAUZINO, 2010) as RNAs são estruturas poderosas comumente utilizadas no processo de reconhecimento de padrões devido a sua capacidade de aprendizado, habilidade de generalização, organização dos dados e tolerância a falhas. Em uma RNA, o conhecimento é adquirido a partir do ambiente onde a rede está inserida, através de um processo de aprendizagem, e armazenado na forma de pesos sinápticos que representam a "força" da conexão entre dois neurônios artificiais. Quanto maior a "força" de uma conexão, maior é o peso desses neurônios no processo de tomada de decisão. Em geral, as redes neurais artificiais são basicamente formadas pela seguinte estrutura:

1. Camada de entrada: Composta por uma ou mais unidades, representando os dados utilizados no processamento da rede.
2. Camadas intermediárias ou escondidas: Composta por uma ou mais camadas, sendo cada uma delas composta por um ou mais neurônios artificiais, este

componente da rede neural é responsável por processar os dados de entrada e produzir a saída/resposta da rede.

3. Camada de saída: Retorna a resposta processada pela rede neural.

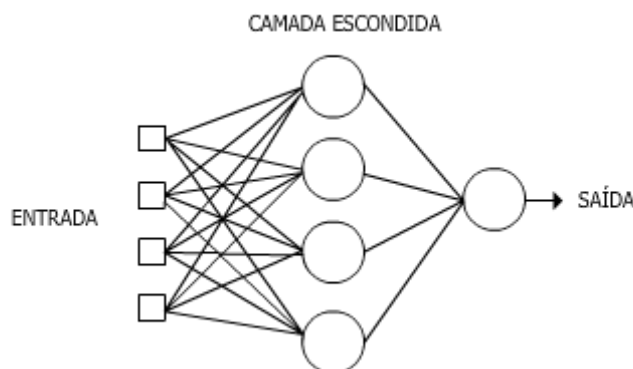


Figura 10: Estrutura básica de uma rede neural

A Figura 10 apresenta uma rede neural composta por 4 entradas, uma única camada escondida contendo 4 neurônios e uma saída.

A partir dos dados obtidos no processo de coleta de dados, uma rede neural artificial foi construída para utilizar os dados do ritmo de digitação como entrada, e retornar informações sobre a personalidade de cada um dos participantes em um formato equivalente ao resultado obtido ao aplicar o questionário TIPI. Diferentes formatos de redes neurais foram utilizados, sendo todos baseados na mesma arquitetura base, composta por uma camada de entrada, uma camada intermediária e uma camada de saída, porém, a quantidade de entradas, a quantidade de neurônios na camada intermediária e a quantidade de saídas na camada de saída variam de acordo com o experimento em questão. Na realização dos experimentos, optou-se por utilizar os dados do ritmo de digitação dos participantes de três formas distintas, com o objetivo de comparar os resultados obtidos em cada uma delas:

1. Utilizar apenas os dados do *Hold time* dos participantes como entrada da rede;
2. Utilizar apenas o *Down-Down time* dos participantes como entrada da rede;
3. Utilizar os dados do *Hold time* e do *Down-Down time* simultaneamente como entrada da rede;

Durante o processo de coleta de dados foram extraídos 376 intervalos de tempo relacionados aos eventos de teclas do teclado, sendo 188 referentes ao *hold time* e 188 referentes ao *up time*. A partir dessas duas características, foram calculados os dados referentes ao *down-down time*, sendo este também composto por 188 intervalos de tempo. Dessa forma, as redes neurais construídas para as abordagens que utilizam apenas o *hold time* ou o *down-down time* possuem 188 entradas na camada de entrada e 188 neurônios na camada intermediária. Por sua vez, as abordagens que utilizam os dados do *hold time* em conjunto com o *down-down time* tem sua camada de entrada composta por 376 entradas e a camada intermediária contendo 376 neurônios.

Para a validação dos resultados obtidos pela rede neural, três abordagens de comparação foram definidas: (i) Comparação do resultado obtido pela rede neural com o resultado obtido pelo questionário TIPI; (ii) Classificar os resultados obtidos pela rede e pelo questionário em

três classes distintas, que identificam um valor "baixo", "médio" ou "alto" para os traços de personalidade. Dessa forma, foi considerado acerto quando o resultado obtido pela rede neural foi classificado como pertencendo à mesma classe que o resultado obtido através do questionário TIPI; (iii) Os resultados obtidos tanto pela rede neural quanto pela aplicação dos questionário TIPI foram binarizados, de forma que resultados com valores maiores ou iguais a 5 foram convertidos para o valor binário 1, enquanto resultados menores que 5 foram convertidos para o valor binário 0.

Para que fosse possível realizar diferentes abordagens de comparação entre os resultados, redes neurais com diferentes estrutura na camada de saída tiveram que ser desenvolvidas. A Figura 11 representa a estrutura da rede neural utilizada na abordagem de validação (i), cujas saída é construída seguindo a escala Likert. Já a Figura 12 e a Figura 13, representam as estruturas das redes neurais das abordagens (ii) e (iii), cujas saídas são respectivamente representadas seguindo a classificação em três faixas de valor e o formato binário.

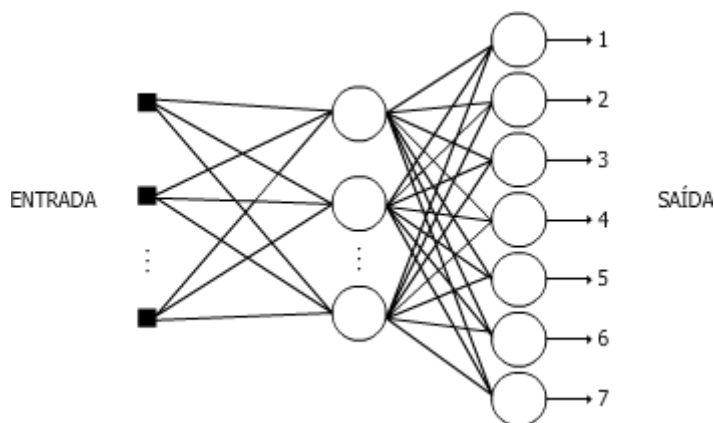


Figura 11: Estrutura da rede neural utilizada para mapear saídas na escala Likert

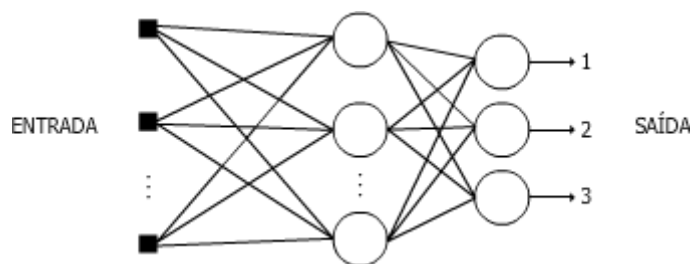


Figura 12: Estrutura da rede neural utilizada para mapear saídas classificadas em três faixas de valor

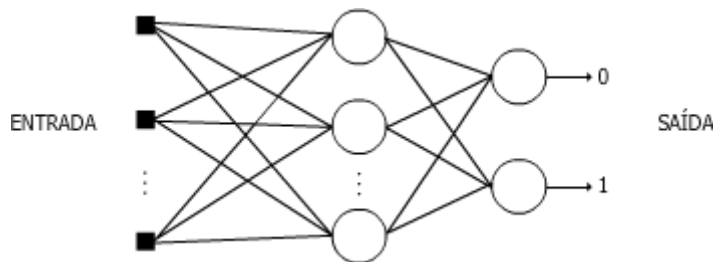


Figura 13: Estrutura da rede neural utilizada para mapear saídas binárias

Devido à arquitetura escolhida para as redes neurais, os experimentos foram realizados de forma que cada traço de personalidade foi analisado individualmente por cada uma das redes neurais, onde a relação entre o ritmo de digitação e traços da personalidade foram mapeados como sendo um problema de classificação, cabendo à rede neural identificar a qual saída da rede os dados do ritmo de digitação está mais relacionado.

Todas as redes neurais utilizadas neste trabalho foram desenvolvidas de forma indireta sobre o Theano, um *framework* desenvolvido em Python, classificado entre os três *frameworks* de *deep learning* mais utilizados pela comunidade científica, tendo como principais características à representação simbólica de expressões através de uma linguagem de descrição de alto nível, flexibilidade para a construção/adaptação de arquiteturas de *deep learning*, suporte a uma vasta gama de bibliotecas, e compatibilidade com uso de CPU e GPU (BAHRAMPOUR, 2016). Por ser um *framework* voltado para a avaliação de expressões matemáticas, usuários de Theano necessitam de uma grande curva de aprendizado até construírem códigos de forma efetiva (BAHRAMPOUR, 2016), por isso, utilizou-se o *framework* Lasagne⁶, construído sobre o Theano, como forma de abstrair a complexidade imposta pela ferramenta original, através da utilização dos métodos e componentes predefinidos para construção simplificada de uma rede neural.

No desenvolvimento deste trabalho foram idealizados 18 tipos diferentes de experimentos, onde cada um representa uma combinação única de três propriedades específicas dos dados extraídos, em busca de uma abordagem que fosse capaz de retornar as melhores taxas de acerto. As propriedades que variam entre cada um dos experimentos são: (i) Dados de entrada da rede; (ii) Equalização da entrada; (iii) Formato da saída da rede.

Dentre as características variadas em cada um dos experimentos, o formato da saída é a característica que identifica o objetivo final da rede neural, relacionado às questões **Q1** e **Q2** propostas na seção 3.1.2. As redes neurais com saída no formato da escala Likert têm como objetivo identificar exatamente quanto um traço de personalidade específico representa na personalidade do indivíduo (**Q1**), de forma que possa ser diretamente comparado com o resultado obtido através da aplicação do questionário TIPI. Já as redes neurais com saída classificada em três faixas de valor possuem o mesmo objetivo que as redes neurais com saída na escala Likert, porém, seguindo uma abordagem de classificação dos resultados equivalente à exposta em (COSTA e PORTO, 2011), com o intuito de comparar os resultados obtidos pelos dois trabalhos. A Tabela 3 descreve os valores utilizados em ambos os trabalhos para a classificação dos resultados em cada uma das faixas especificadas.

Faixa de valor	Classificação
0~44	Baixo
45~55	Médio
56~100	Alto

Tabela 3: Intervalos de valor utilizados na classificação dos traços de personalidade

De forma diferente das redes neurais com saídas na escala Likert e classificadas em três faixas de valor, a rede neural com saídas binárias tem como único objetivo identificar se um traço de personalidade se destacada na personalidade de um indivíduo (**Q2**). Classifica-se como em destaque um traço de personalidade que ao analisado na escala Likert, possuir valor maior ou igual a 5. Ou seja, as respostas do questionário TIPI são pré-processadas, de forma que

⁶ Lasagne está disponível em: <https://github.com/Lasagne/Lasagne>

valores maiores ou iguais a 5 são convertido para o valor binário 1, enquanto o demais valores são convertidos para 0. É considerado acerto da rede neural quando a resposta da rede possuir um valor binário igual ao valor da saída esperada também em formato binário.

A utilização dos dados "brutos" do ritmo de digitação como entrada da rede neural foi escolhida devido a uma falta de uma definição formal por parte da literatura, sobre a relação de uma característica específica do ritmo de digitação, como por exemplo a velocidade de digitação, com um determinado traço de personalidade. Com base nisso, optou-se por passar os dados "brutos" como entrada da rede neural, de forma que a rede pudesse ponderar esses dados, e por si só definir quais características são mais relevantes para cada um dos traços de personalidade.

4 Experimentos e Resultados

Para a realização dos experimentos foram definidas as estruturas de cada uma das abordagens, que determinaram quais dados foram utilizados como entrada da rede neural, o formato desses dados e a utilização ou não de métodos de pré-processamento. Ao todo foram realizados 18 tipos diferentes de experimentos, cada um deles tendo sido aplicado a todos os cinco traços de personalidade analisados.

Os experimentos foram realizados através da aplicação dos dados a uma rede neural de múltiplas camadas (camada de entrada, camada intermediária e camada de saída), ajustada de acordo com o formato dos dados selecionados para cada um dos experimentos. Dessa forma, embora de modo geral a rede neural utilizada em todos os experimentos possuísse uma arquitetura semelhante, a quantidade de entradas, neurônios na camada intermediária e o formato da saída variou de acordo com os dados utilizados e o objetivo de cada experimento.

Visando responder as duas questões expostas na seção 3.1.2, os testes realizados foram divididos em duas vertentes: (i) Testes com saídas binárias, cujos objetivo é determinar se um traço de personalidade está ou não presente na personalidade de um indivíduo; (ii) Saídas não binárias, visando mensurar quanto um traço de personalidade está presente na personalidade de um indivíduo. Tabela 4 ilustra todos os experimentos desenvolvidos e cada uma das suas características.

Experimento	Dados de entrada	Entrada equalizada	Formato da saída
Experimento 1	<i>Hold time</i>	NÃO	Likert
Experimento 2	<i>Hold time</i>	SIM	Likert
Experimento 3	<i>Down-Down time</i>	NÃO	Likert
Experimento 4	<i>Down-Down time</i>	SIM	Likert
Experimento 5	<i>Hold e Down-Down time</i>	NÃO	Likert
Experimento 6	<i>Hold e Down-Down time</i>	SIM	Likert
Experimento 7	<i>Hold time</i>	NÃO	Três faixas de valor
Experimento 8	<i>Hold time</i>	SIM	Três faixas de valor
Experimento 9	<i>Down-Down time</i>	NÃO	Três faixas de valor

Experimento 10	<i>Down-Down time</i>	SIM	Três faixas de valor
Experimento 11	<i>Hold e Down-Down time</i>	NÃO	Três faixas de valor
Experimento 12	<i>Hold e Down-Down time</i>	SIM	Três faixas de valor
Experimento 13	<i>Hold time</i>	NÃO	Binária
Experimento 14	<i>Hold time</i>	SIM	Binária
Experimento 15	<i>Down-Down time</i>	NÃO	Binária
Experimento 16	<i>Down-Down time</i>	SIM	Binária
Experimento 17	<i>Hold e Down-Down time</i>	NÃO	Binária
Experimento 18	<i>Hold e Down-Down time</i>	SIM	Binária

Tabela 4: Experimentos realizados e suas respectivas particularidades

Como forma de simplificar a visualização dos resultados obtidos no experimentos, os gráficos foram criados agrupando os resultados dos experimentos de acordo com a os dados utilizados como entrada e o formato da saída, de forma que cada gráfico apresente duas abordagens lado a lado, sendo a primeira com dados não equalizados e a segunda com dados equalizados.

Os experimentos foram realizados utilizando os dados de entrada selecionados na rede neural, e comparando-os com a saída esperada para cada uma das entradas. Dentre o conjunto de 177 participantes da etapa de coleta de dados, 85 desses participantes foram selecionados para serem utilizados no processo de treinamento da rede neural, enquanto 46 foram utilizados para a fase de validação do treinamento da rede e os últimos 46 para a fase de testes. Os resultados apresentados a seguir são oriundos das respostas obtidas na fase de testes da rede neural.

4.1 Distribuição de probabilidade entre as classes

Para analisar a distribuição de probabilidade entre as classes estudadas de cada um dos traços de personalidade, as amostras obtidas na etapa de coleta de dados foram utilizadas para histogramas que representam a probabilidade de uma amostra aleatória ser classificada em uma das classes analisadas, para cada um dos traços de personalidade estudados, em todas as três abordagens desenvolvidas (Likert, três faixas de valor e binária).

A construção dos histogramas foi realizada com o intuito de confirmar a existência de uma distribuição de probabilidade desigual em todas as três abordagens desenvolvidas, assim como em suas respectivas classes. Os histogramas foram construídos utilizando os dados das amostras previamente segmentadas em três grupos distinto, para serem utilizadas no processo de treinamento da rede neural: (i) Conjunto de treinamento; (ii) Conjunto de validação; (iii) Conjunto de teste.

Além da utilização isolada de cada um desses três conjuntos de amostras, um quarto conjunto denominado *dataset* completo, composto pela união dos três conjuntos descritos anteriormente, também foi utilizado para a construção e posterior análise dos histogramas de distribuição de probabilidades. A análise dos histogramas consiste em verificar qual a probabilidade de uma amostra aleatória ser classificada como pertencente a cada uma das

classes analisadas, onde a soma das probabilidades de todas as classes deve ser igual a 100%. A análise da distribuição de probabilidade entre as classes foi realizada com o objetivo de identificar distribuições desiguais, e consequentemente auxiliar na análise final dos resultados, dando embasamento para a compreensão e o julgamento das taxas de acerto obtidas em cada uma das abordagens desenvolvidas.

Para cada uma das abordagens analisadas, foram realizados histogramas com cada uma das variações de conjuntos de dados (*datasets*) disponíveis. Os *datasets* utilizados estão divididos da seguinte forma: Um conjunto de treinamento composto por 85 amostras, um conjunto de validação composto por 46 amostras, um conjunto de teste composto por outras 46 amostras e o *dataset* completo, compostos por todas as amostras anteriores, com um total de 177 amostras.

4.1.1 Likert

Os gráficos a seguir (Gráfico 3, Gráfico 4, Gráfico 5, Gráfico 6 e Gráfico 7) apresentam os histogramas de cada uma das classes analisadas, de acordo com a sua distribuição de probabilidade seguindo a escala Likert, onde cada um dos histogramas mostra a probabilidade de uma amostra aleatória ser classificada com um determinado valor na escala Likert, que varia entre 1 e 7.

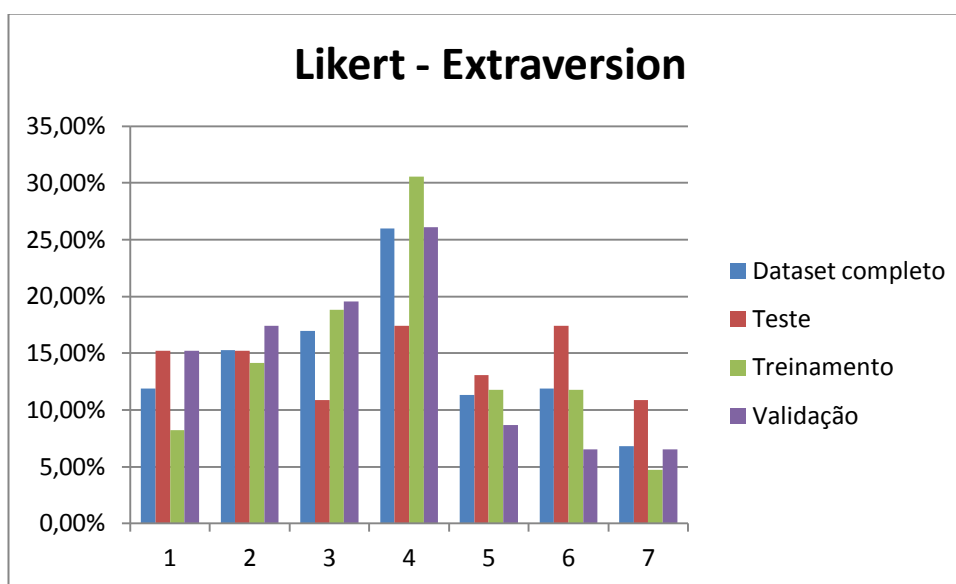


Gráfico 3: Distribuição de probabilidade da extroversão na abordagem da escala Likert

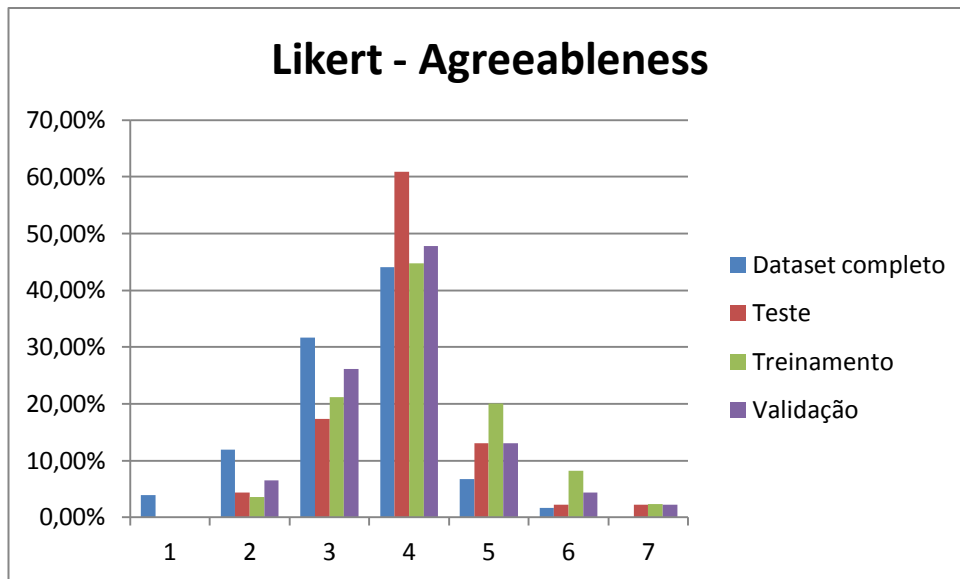


Gráfico 4: Distribuição de probabilidade da socialização na abordagem da escala Likert

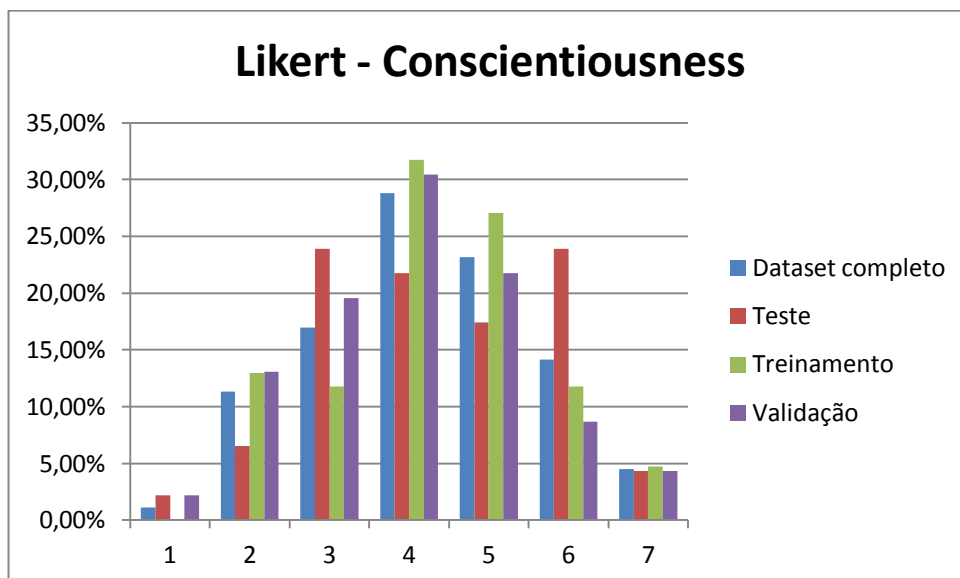


Gráfico 5: Distribuição de probabilidade da realização na abordagem da escala Likert

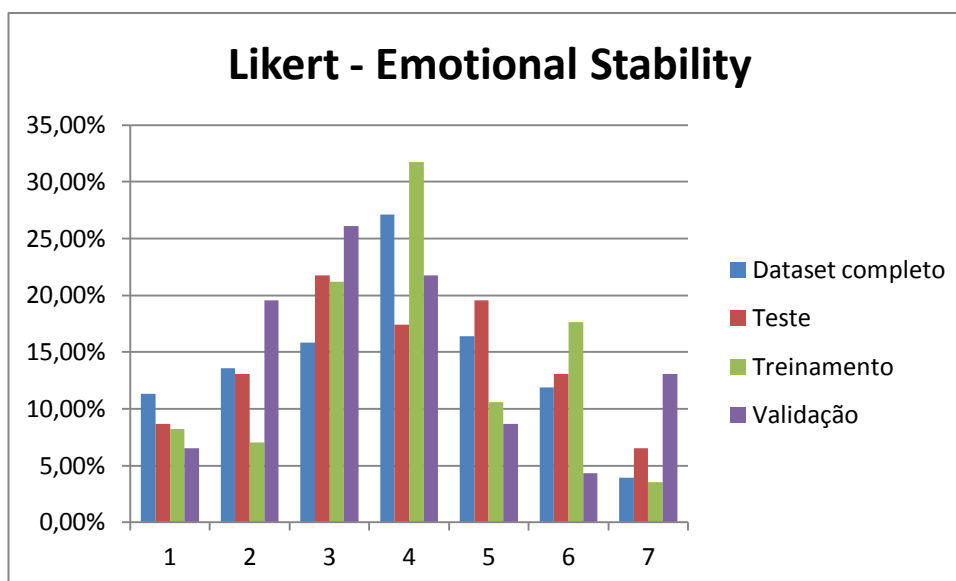


Gráfico 6: Distribuição de probabilidade do neuroticismo na abordagem da escala Likert

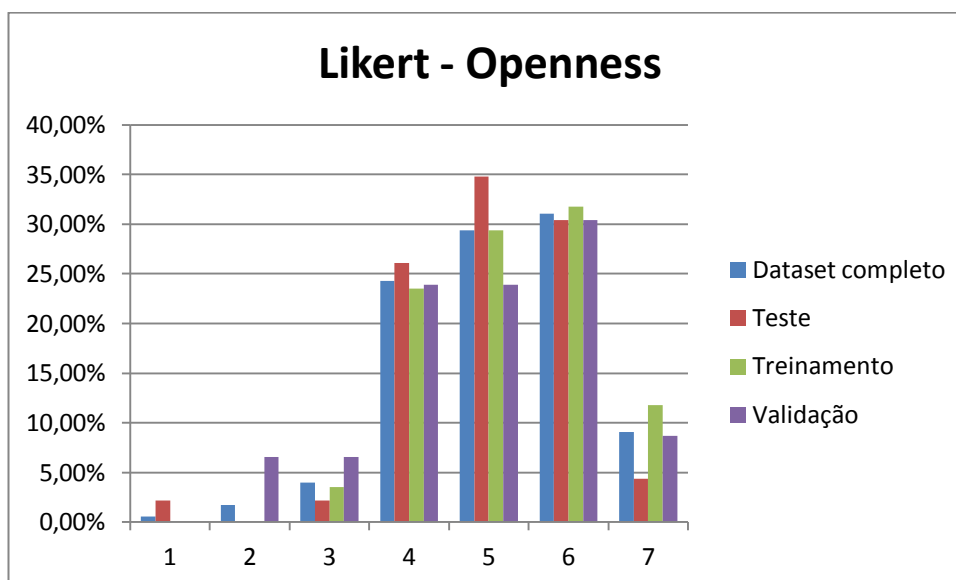


Gráfico 7: Distribuição de probabilidade da abertura na abordagem da escala Likert

Ao analisar os histogramas das distribuições de probabilidade de cada um dos traços de personalidade, é possível perceber a existência de uma distribuição não uniformizada, que devido a sua grande desigualdade entre as classes, pode influenciar diretamente no resultado obtido pelo método de aprendizagem utilizado.

4.1.2 Três faixas

De forma semelhante ao que foi realizado com a abordagem na escala Likert, histogramas foram criados para analisar a distribuição de probabilidade entre cada um dos três possíveis valores da abordagem de três faixas de valor: Baixo, médio e alto, representados nos gráficos a seguir respectivamente pelos números 1, 2 e 3.

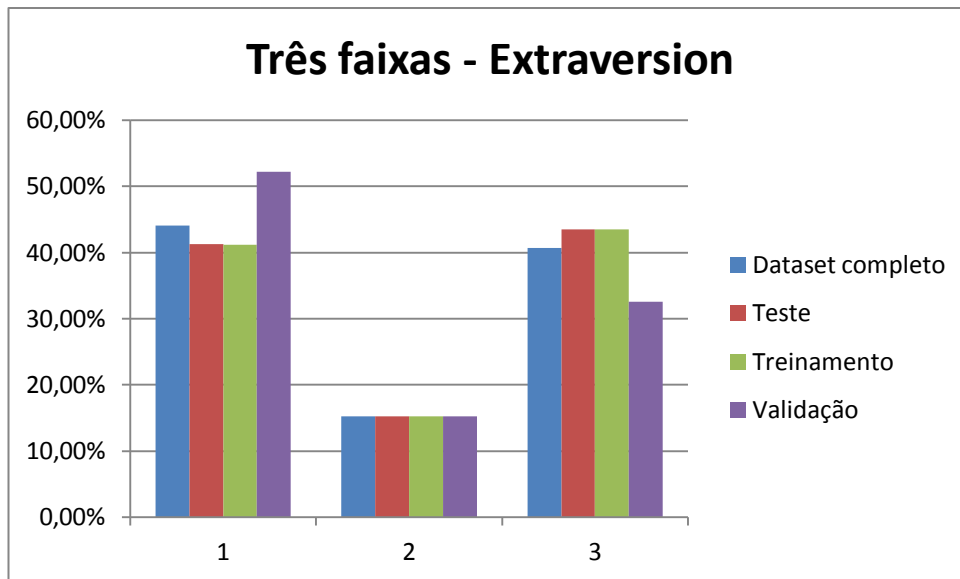


Gráfico 8: Distribuição de probabilidade da extroversão na abordagem de três faixas

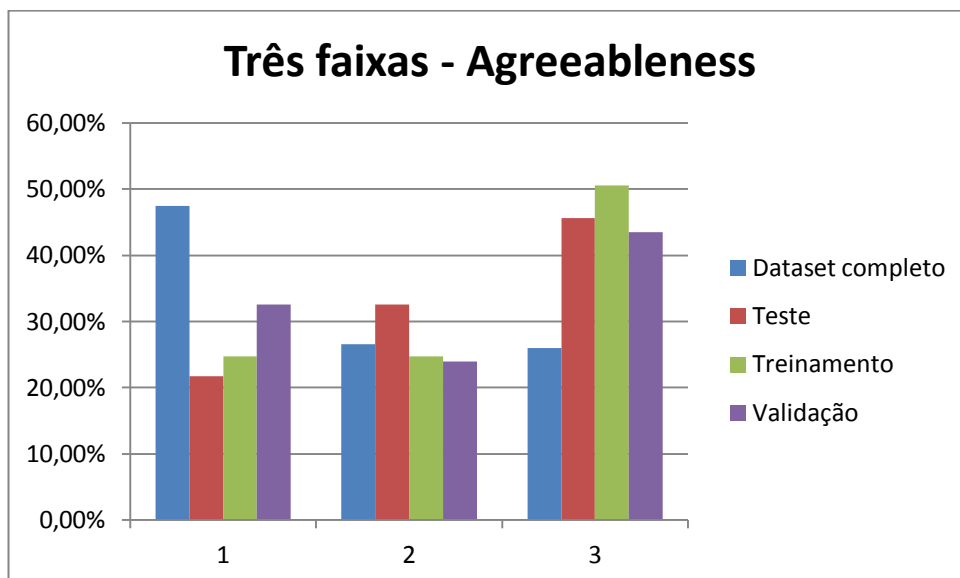


Gráfico 9: Distribuição de probabilidade da socialização na abordagem de três faixas

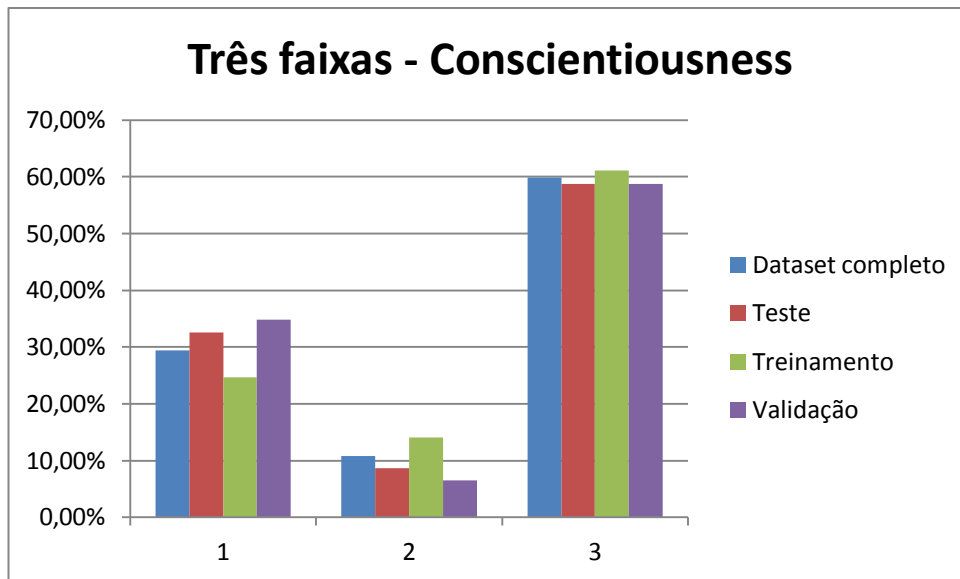


Gráfico 10: Distribuição de probabilidade da realização na abordagem de três faixas

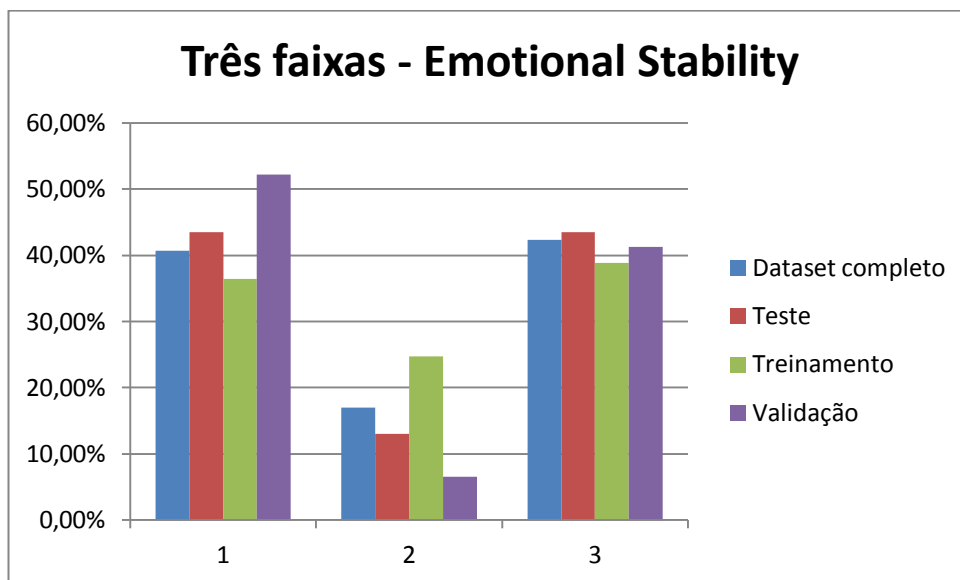


Gráfico 11: Distribuição de probabilidade do neuroticismo na abordagem de três faixas

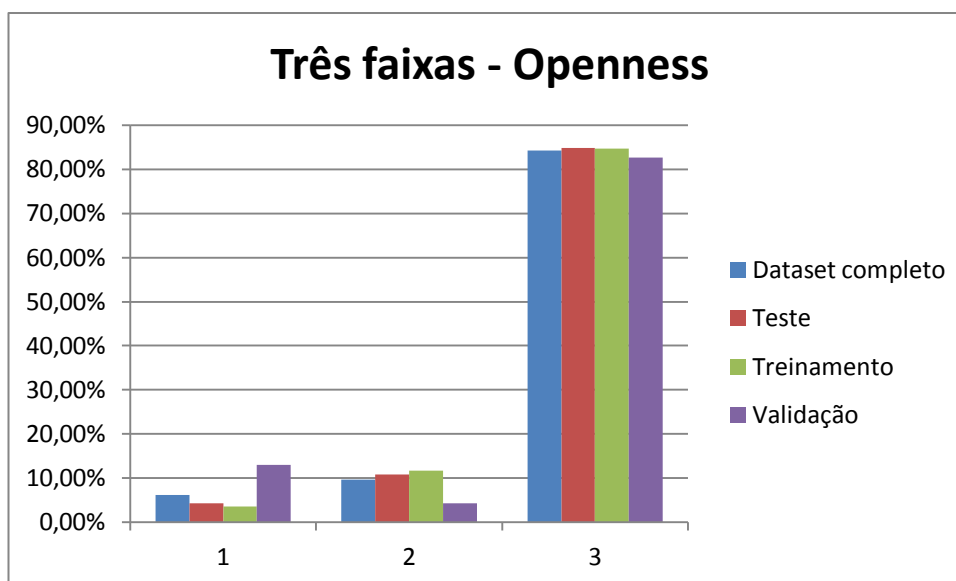


Gráfico 12: Distribuição de probabilidade da abertura na abordagem de três faixas

Ao observar os histogramas da abordagem de três faixas de valor, representados pelo Gráfico 8, Gráfico 9, Gráfico 10, Gráfico 11 e Gráfico 12, nota-se uma grande variação na distribuição de probabilidade entre as classes, destacando o traço de personalidade "*openness*", onde a probabilidade de uma amostra pertencer à faixa 3 (alto) chega a 84,78%, enquanto a faixa 1 (baixo) chega a apenas 13,04%.

4.1.3 Binário

Os histogramas apresentados a seguir (Gráfico 13, Gráfico 14, Gráfico 15, Gráfico 16 e Gráfico 17) exibem as distribuições de probabilidade para os experimentos desenvolvidos segundo a abordagem binária, cujos resultados podem ser classificados em duas classes distintas, representadas pelos valores "0" e "1". Entre as distribuições de probabilidade desiguais identificadas, a abordagem binária chegou a apresentar traços de personalidade com 91,53% de chance de pertencer a uma classe específica, assim como pode ser observado no histograma apresentado pelo Gráfico 14, referente ao traço de personalidade *agreeableness*.

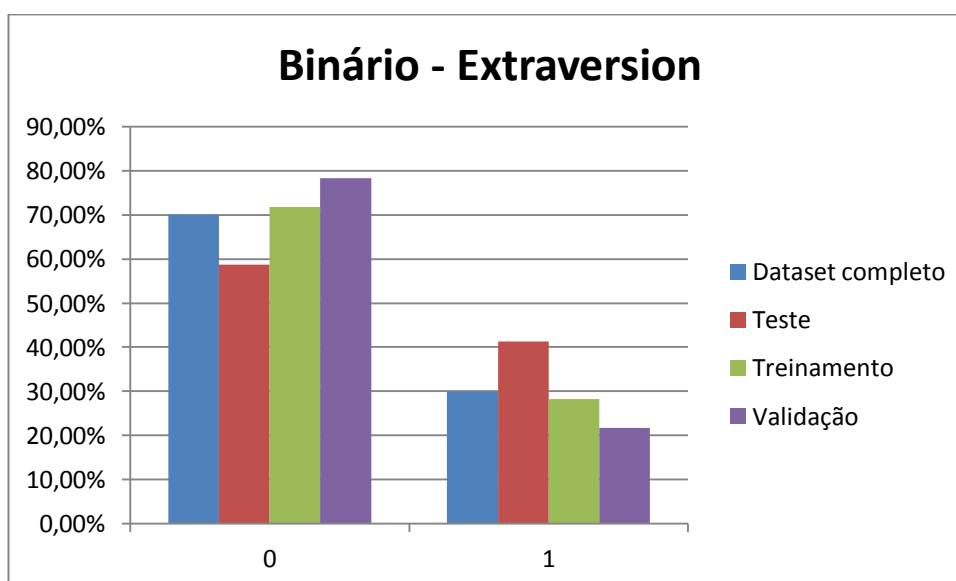


Gráfico 13: Distribuição de probabilidade da extroversão na abordagem binária

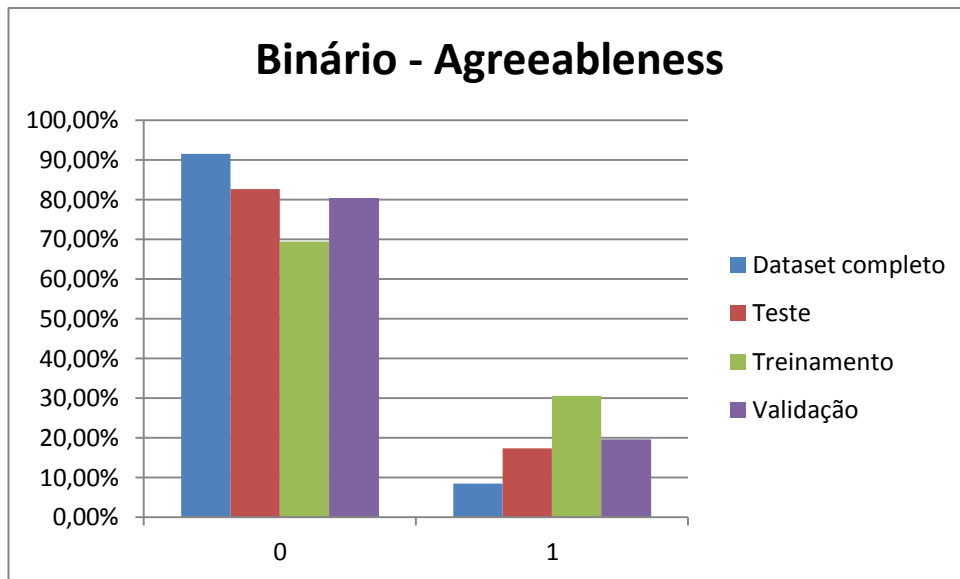


Gráfico 14: Distribuição de probabilidade da socialização na abordagem binária

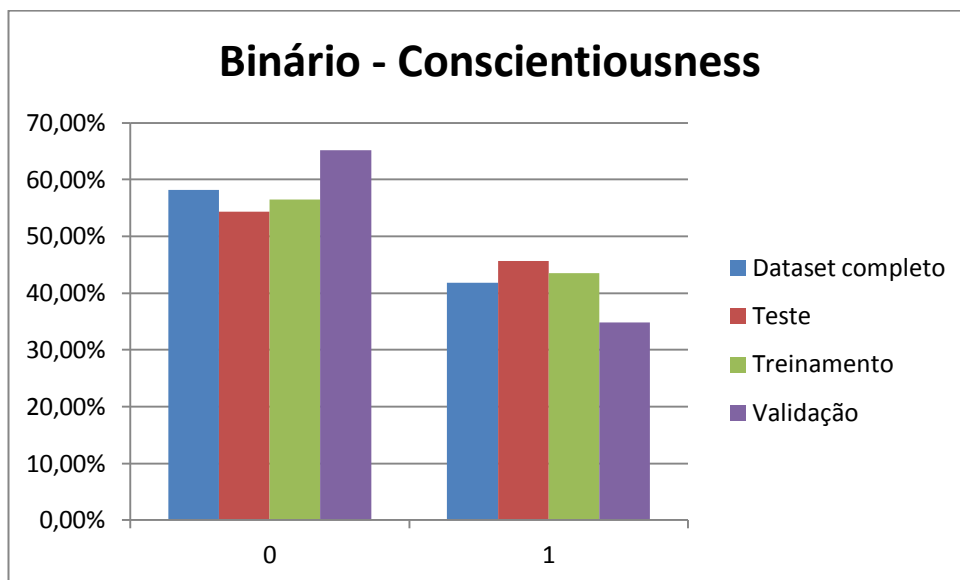


Gráfico 15: Distribuição de probabilidade da realização na abordagem binária

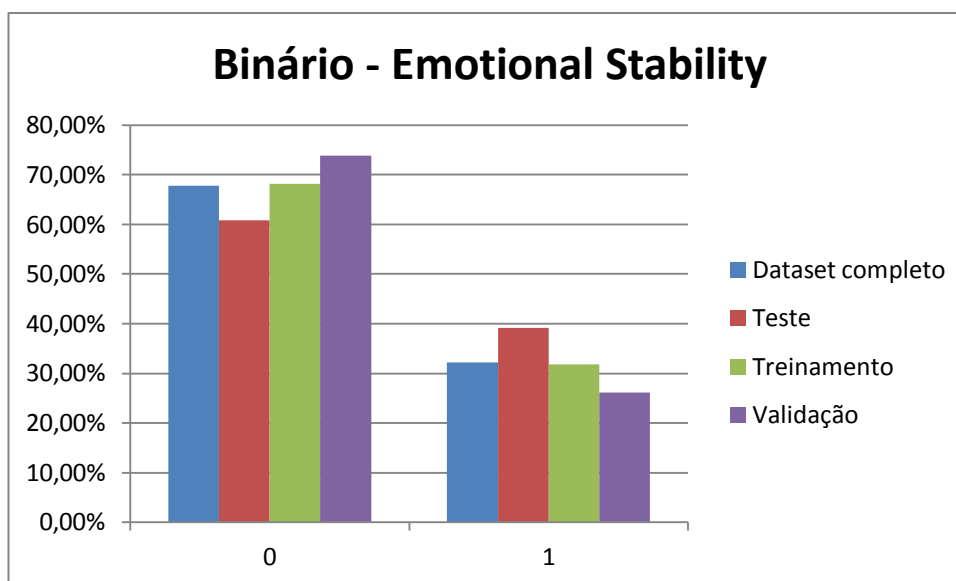


Gráfico 16: de probabilidade do neuroticismo na abordagem binária

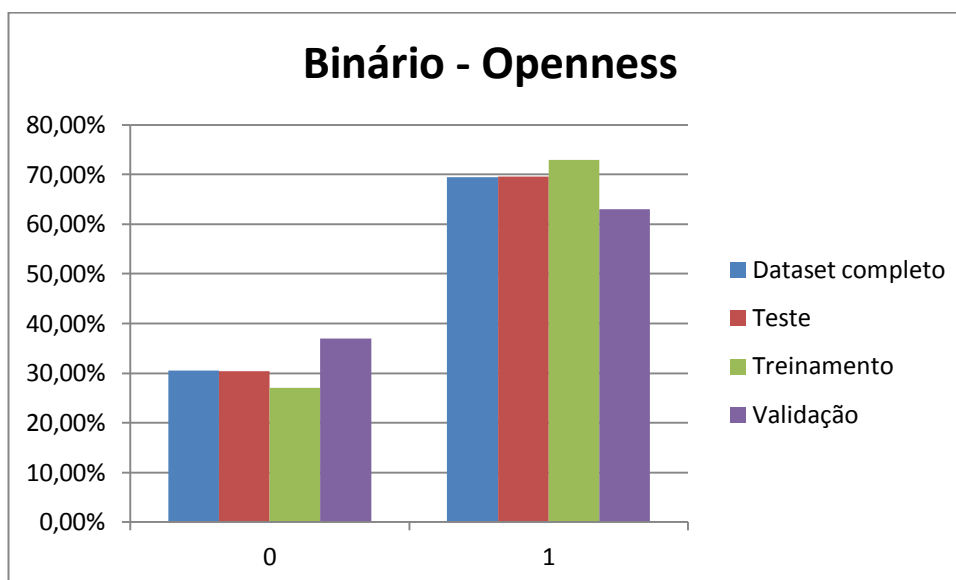


Gráfico 17: Distribuição de probabilidade da abertura na abordagem binária

A partir da observação e análise realizadas nas representações gráficas da distribuição de probabilidade entre as classes de cada uma das abordagens desenvolvidas, podemos fundamentar conclusões sobre os resultados obtidos nos experimentos apresentados a seguir nas seções 4.2, 4.3 e 4.4, de forma que mesmo que altos percentuais de acerto sejam identificados nos resultados dos experimentos, a qualidade/relevância desse resultado final só poderá ser comprovada, caso seja corroborada pela conclusão obtida na análise da distribuição de probabilidade.

4.2 Experimento com saídas na escala Likert

A escala Likert (LIKER, 1932) é um tipo de escala onde os entrevistados devem especificar o seu nível de concordância com uma afirmação, tendo sido desenvolvida especificamente para respostas psicométricas, e utilizada na aplicação de questionários, podendo ser apresentada no formato de 3 pontos, 5 pontos ou 7 pontos. Para a análise dos dados deste

trabalho, utilizou-se a escala Likert composta por 7 pontos, ou seja, 7 opções de resposta para cada uma das perguntas, por ser este o padrão adotado no desenvolvimento do questionário TIPI.

A equalização dos dados foi realizada seguindo o método descrito por (MONTALVÃO e FREIRE, 2006), onde foi utilizada em análises biométricas voltadas para fins de autenticação. Por ter apresentado bons resultados, decidiu-se utilizar tal abordagem visando comparar os resultados obtidos pelos experimentos com e sem a aplicação do método de equalização, embora seja de conhecimento que segundo o que foi afirmado por (MONTALVÃO e FREIRE, 2006) em seu trabalho, a utilização de abordagens de equalização em conjunto com redes neurais pode ser considerada redundante, tendo em vista que devido ao seu processo de aprendizagem, a rede neural efetua uma equalização natural dos dados entrada. A aplicação do método proposto por (MONTALVÃO e FREIRE, 2006) foi realizada utilizando os seguintes valores para média (μ) e desvio padrão (σ): $\mu = 128,4094$ e $\sigma = 842,9373$.

Ao observar o Gráfico 18 é possível identificar a igualdade entre as taxas de acerto de todos os traços de personalidade, quando comparados os resultados das abordagens equalizadas e não equalizadas, mostrando que mesmo com a existência da aleatoriedade natural do processo de aprendizagem de uma rede neural, e a aplicação de uma abordagem de equalização, o Gráfico 18 não apresenta nenhuma dessas influências no resultado final obtido.

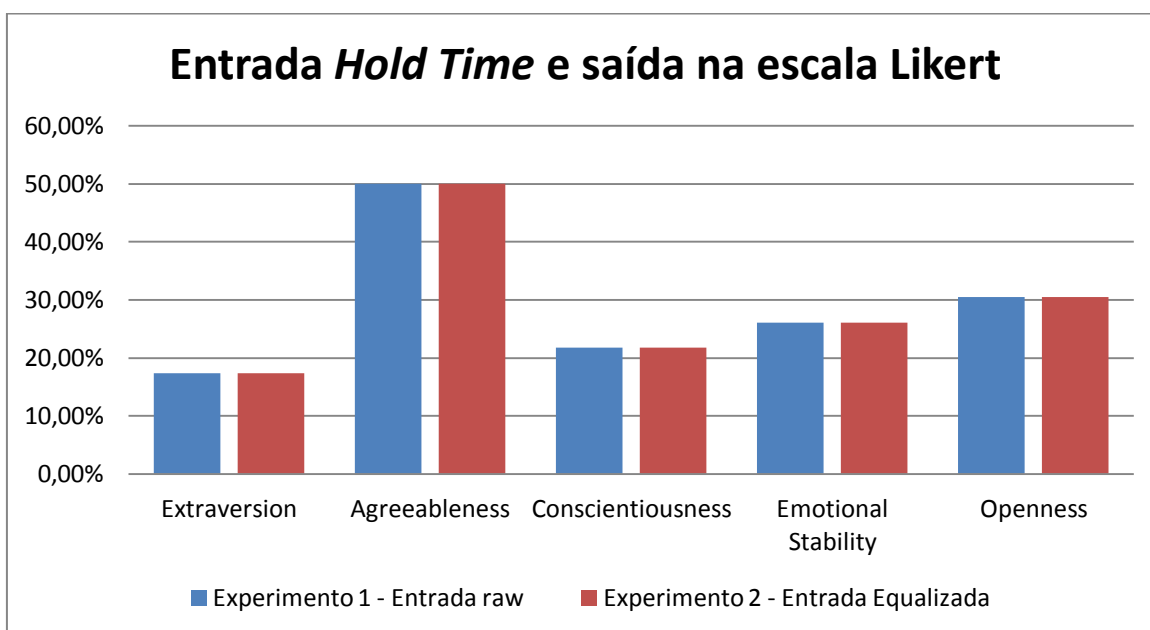


Gráfico 18: Comparação de resultados dos experimentos que utilizam o hold time como entrada e retornam a saída na escala likert.

O Gráfico 18 apresenta a comparação dos resultados obtidos pela rede neural na análise dos dados do *hold time* dos participantes, onde seu objetivo é identificar o quanto um traço de personalidade específico está ligado à personalidade do indivíduo, ao tentar prever um valor entre 1 e 7, que represente cada um dos traços de personalidade, seguindo a escala Likert.

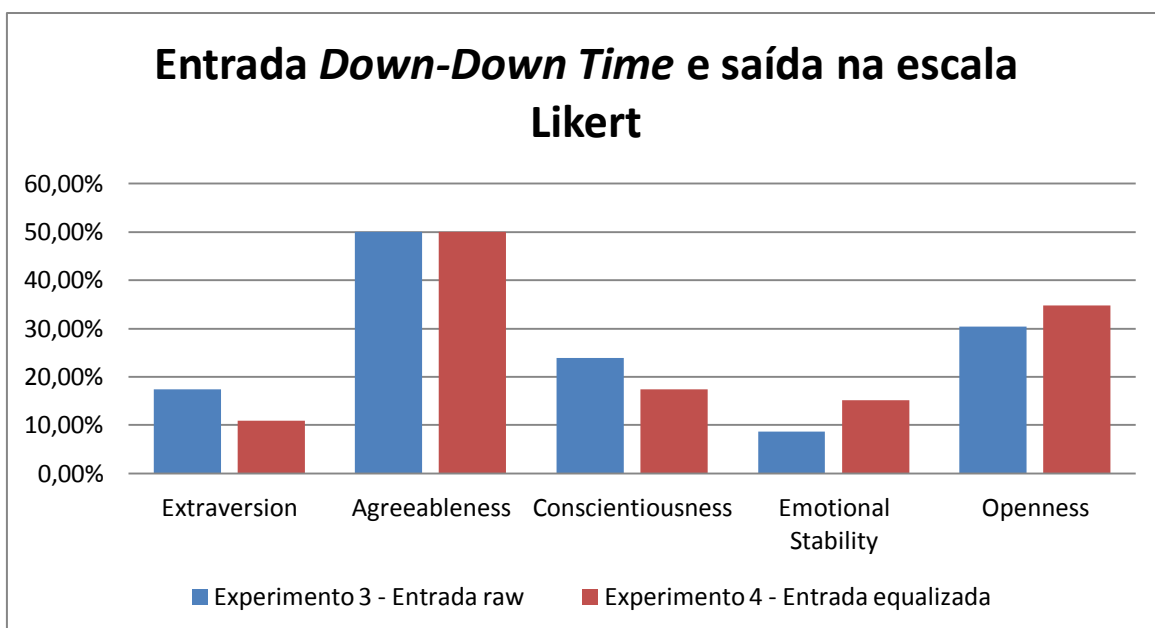


Gráfico 19: Comparação de resultados dos experimentos que utilizam o *down-down time* como entrada e retornam a saída na escala likert.

O *down-down time* é classificado pela literatura como a característica do ritmo de digitação de um indivíduo que carrega o maior número de informação sobre quem digitou. Por este motivo, o *down-down time* foi extraído em conjunto com as demais características do ritmo de digitação (*hold time* e *up time*), com o intuito de fornecer para a rede neural o máximo de informação possível com o mínimo necessário de características. O Gráfico 19 apresenta uma comparação entre a abordagem equalizada e a abordagem não equalizada, com utilização dos dados do *down-down time*, onde os resultados obtidos variaram entre 8,70% e 50%.

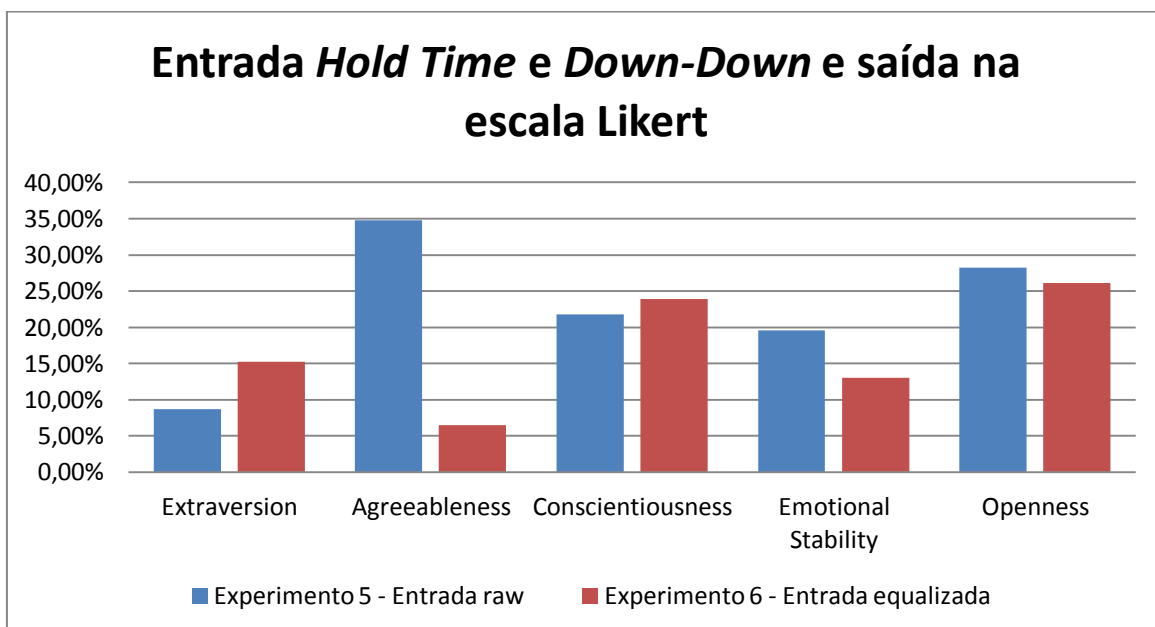


Gráfico 20: Comparação de resultados dos experimentos que utilizam os dados do *hold time* em conjunto com o *down-down time* como entrada e retornam a saída na escala likert.

O Gráfico 20 apresenta os resultados obtidos a partir da análise do *hold time* em conjunto com o *down-down time*, no qual os dados não apresentaram altas taxas de acerto, ficando a maior taxa em torno de 34,78%.

A partir da análise de todos os resultados obtidos pelas abordagens de rede neural com saída na escala Likert, apresentados nos gráficos, Gráfico 18, Gráfico 19 e Gráfico 20, observou-se que as taxas de acerto variaram entre 6,52% e 50%, tendo como média de acerto 24,60%, resultado este insatisfatório para a realização da predição de quanto um traço de personalidade está presente na personalidade de um indivíduo.

O Gráfico 21 apresenta uma comparação dos resultados obtidos em todos os experimentos realizados seguindo a abordagem na escala Likert. Embora o Gráfico 21 não apresente nenhum dado inédito, é a partir da sua observação que pode-se analisar lado a lado os resultados obtidos em cada um dos experimentos apresentados até o momento.

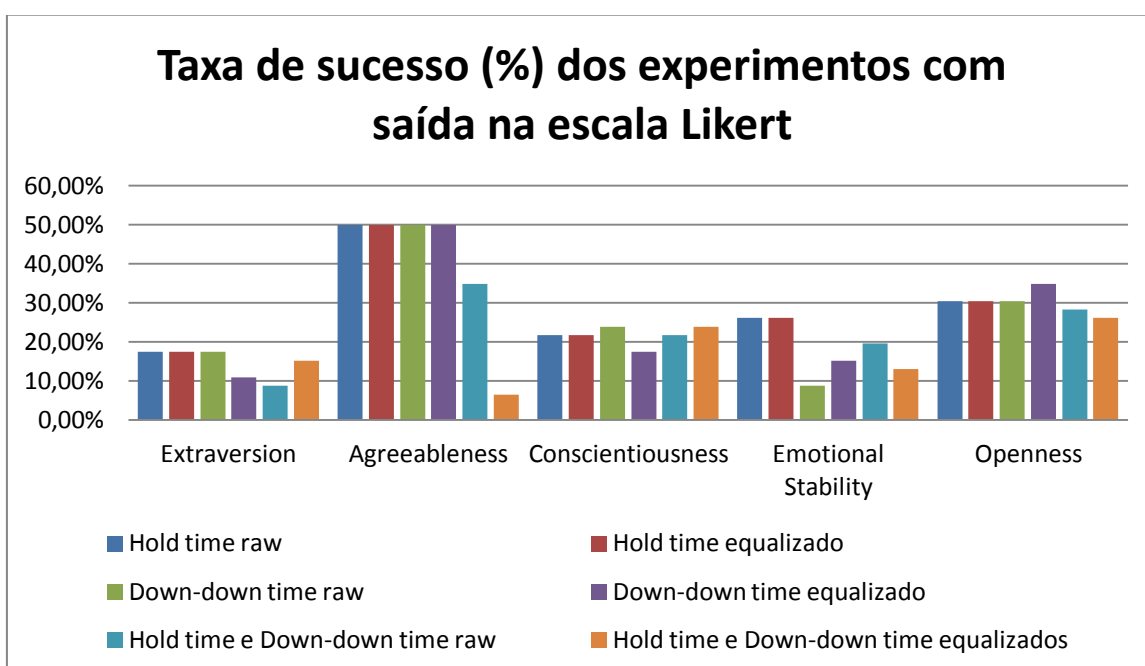


Gráfico 21: Comparação dos experimentos com saída na escala Likert

4.3 Experimentos com saídas classificadas em três faixas de valor

Prevendo a possível obtenção de resultados insatisfatórios através dos experimentos que utilizam a escala Likert como saída da rede neural, considerou-se a possibilidade de classificar os resultados obtidos na referida escala em três faixas de valores distintas. Essa análise foi realizada visando mensurar o nível (baixo, médio ou alto) de influência de um determinado traço de personalidade sobre a personalidade do indivíduo. Tal abordagem foi inspirada no trabalho desenvolvido por (COSTA e PORTO, 2011), onde, em uma etapa de pós-processamento, os dados eram normalizados e representados como um valor entre 0 e 100, para que em seguida tais valores fossem classificados em três segmentos distintos (baixo, médio e alto).

A classificação desses valores em ambos os trabalhos foi realizada através das faixas de valor definidas como: Baixo para valores entre 0 e 44; Médio para valores entre 45 e 55; E alto para valores entre 56 e 100. A partir dessa abordagem, além de ser possível mensurar os dados

obtidos de forma mais abrangente, permite-se também que seja realizada uma comparação entre os resultados obtidos por este trabalho e os resultados obtidos pelo *PersonaliKey*, ferramenta desenvolvida por (COSTA e PORTO, 2011).

A classificação das saídas obtidas após o processamento da rede neural em três faixas distintas de valor, permite que uma gama maior de resultados sejam classificados de forma correta. Esse incremento na taxa de acerto é possível, pois, no experimento anterior, para que fosse considerado acerto da rede neural, era necessário que a rede predissesse o valor exato que representa uma traço de personalidade na personalidade de um indivíduo. Já no experimento atual, para ser considerada correta, basta que o resultado da predição seja classificado na mesma faixa de valor que o valor esperado como resposta da rede (dados obtidos através da aplicação do TIPI).

De forma semelhante ao que foi apresentando nos gráficos referentes a abordagem Likert, a classificação dos resultados em três faixas também expõem resultados não conclusivos com relação ao mapeamento de traços da personalidade, tendo em vista que embora o percentual de acerto varie entre 13,04% e 84,78%, esses resultados estão diretamente ligados às chances de acerto ao acaso, em uma abordagem onde uma escolha aleatória teria uma chance de $\frac{1}{3}$, ou seja, 33,33% de chance de acerto e um desvio padrão de $\sqrt{\frac{1}{3} \times \frac{2}{3}} = \frac{\sqrt{2}}{3} = 47,14\%$, sendo esta exatamente a faixa na qual os resultados obtidos se encaixam, assim como mostram os gráficos Gráfico 22, Gráfico 23 e Gráfico 24.

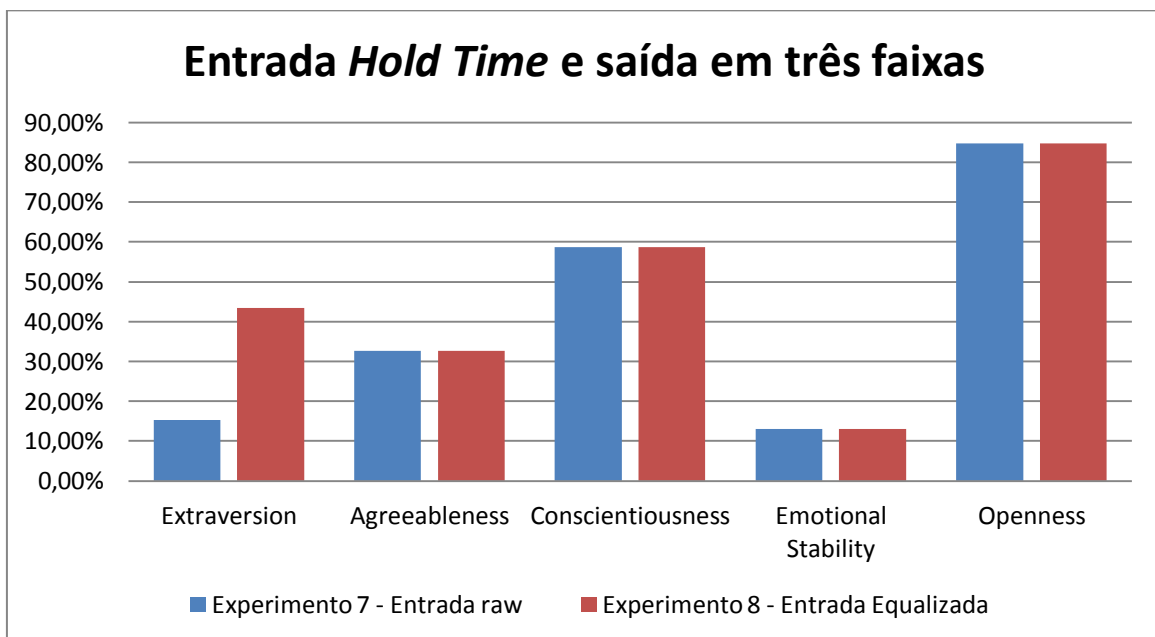


Gráfico 22: Comparação de resultados dos experimentos que utilizam o hold time como entrada e retornam a saída classificada em três faixas de valor.

O Gráfico 23 apresenta os resultados obtidos através da utilização dos dados do *down-down time* em uma rede neural com saídas classificadas em três faixas de valor, onde *extraversion* e *emotional stability* apresentam as menores taxas de acerto, ficando entre 15,22% e 13,04% respectivamente, enquanto *conscientiousness* e *openness* apresentando os melhores resultados, sendo 58,70% e 84,78% respectivamente. Ao observar o Gráfico 23 em conjunto com os gráficos de distribuição de probabilidade apresentados na seção 4.1.2, nota-se que as

taxas de acerto dos traços de personalidade estão diretamente ligadas a forma como a distribuição de probabilidade entre as classes está organizada. Tal fato pode ser observado com clareza nos resultados obtidos pelos traços de personalidade *agreeableness*, *conscientiousness* e *openness*, onde as taxas de acerto apresentadas no Gráfico 23 são exatamente as que foram apresentadas nos gráficos Gráfico 9, Gráfico 10 e Gráfico 12, os quais expõem como a probabilidade de classificação de uma amostra aleatória está distribuída entre os três níveis (baixo, médio e alto) analisados no experimento de três faixas de valor.

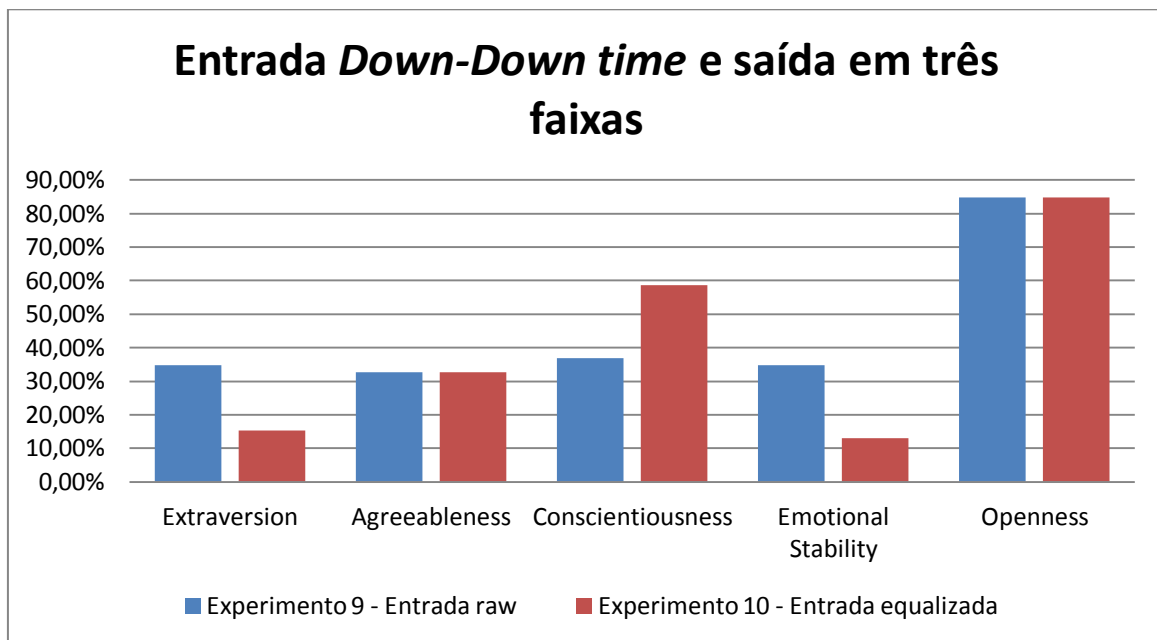


Gráfico 23: Comparação de resultados dos experimentos que utilizam o *down-down time* como entrada e retornam a saída classificada em três faixas de valor.

O Gráfico 24 apresenta os resultados obtidos na aplicação conjunta dos dados do *hold time* e *down-down time* na abordagem de classificação do resultado em três faixas de valores, tendo resultados que mesclam a influencia dos dados do *hold time* (Gráfico 22) e do *down-down time* (Gráfico 23) quando analisados individualmente, apresentando taxas de acerto semelhantes às duas anteriores, em uma mesma abordagem.

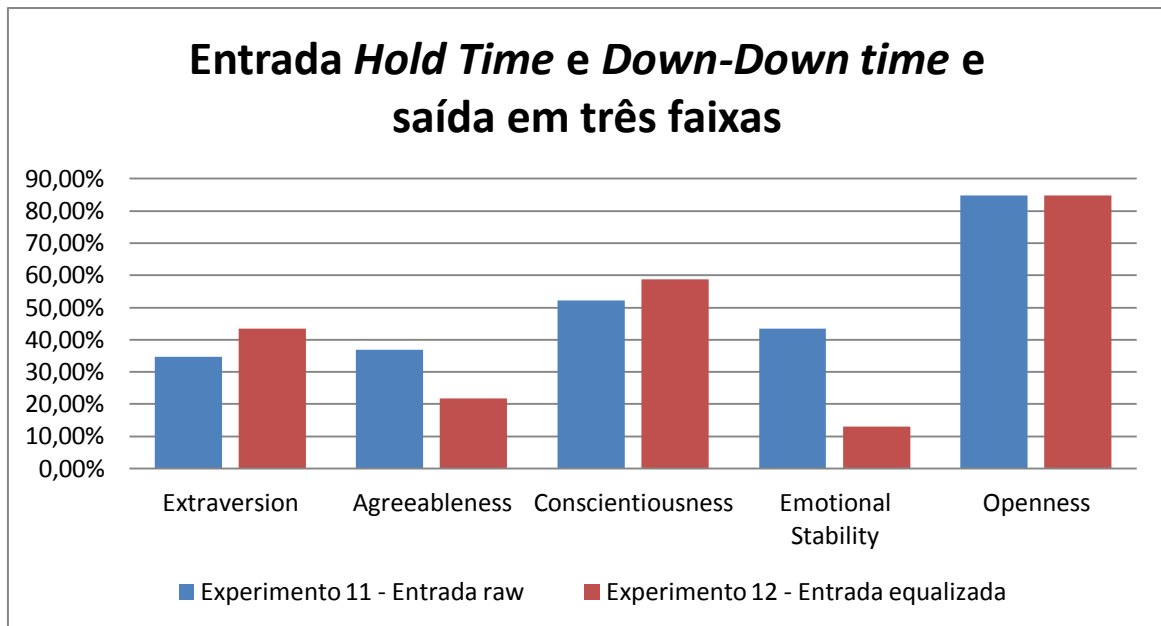


Gráfico 24: Comparação de resultados dos experimentos que utilizam os dados do *hold time* em conjunto com o *down-down time* como entrada e retornam a saída classificada em três faixas de valor.

Embora os resultados obtidos pela abordagem de classificação em três faixas de valor tenha apresentado um resultado aparentemente superior em taxas de acerto, este incremento é oriundo da redução do número de classes mapeadas no experimento, que anteriormente na abordagem de mapeamento através da escala Likert possuía 7 níveis/classes distintos para a classificação de uma amostra, enquanto na abordagem com três faixas de valor, essa quantidade foi reduzida para três (baixo, médio ou alto).

Ao comparar os resultados obtidos pela abordagem de três faixas de valor com os melhores resultados obtidos pelos experimentos do *PersonaliKey* (COSTA e PORTO, 2011), exposto no Gráfico 25, pode-se perceber que três dos cinco traços de personalidade analisados obtiveram resultados inferiores aos apresentados pelo *PersonaliKey*. É importante salientar que o *PersonaliKey* foi desenvolvido utilizando um questionário de autoavaliação composto por 300 questões, o que permite que os resultados por ele obtidos nesse processo sejam mais precisos que os obtidos através da aplicação do questionário TIPI (composto por apenas 10 questões), como realizado neste trabalho.

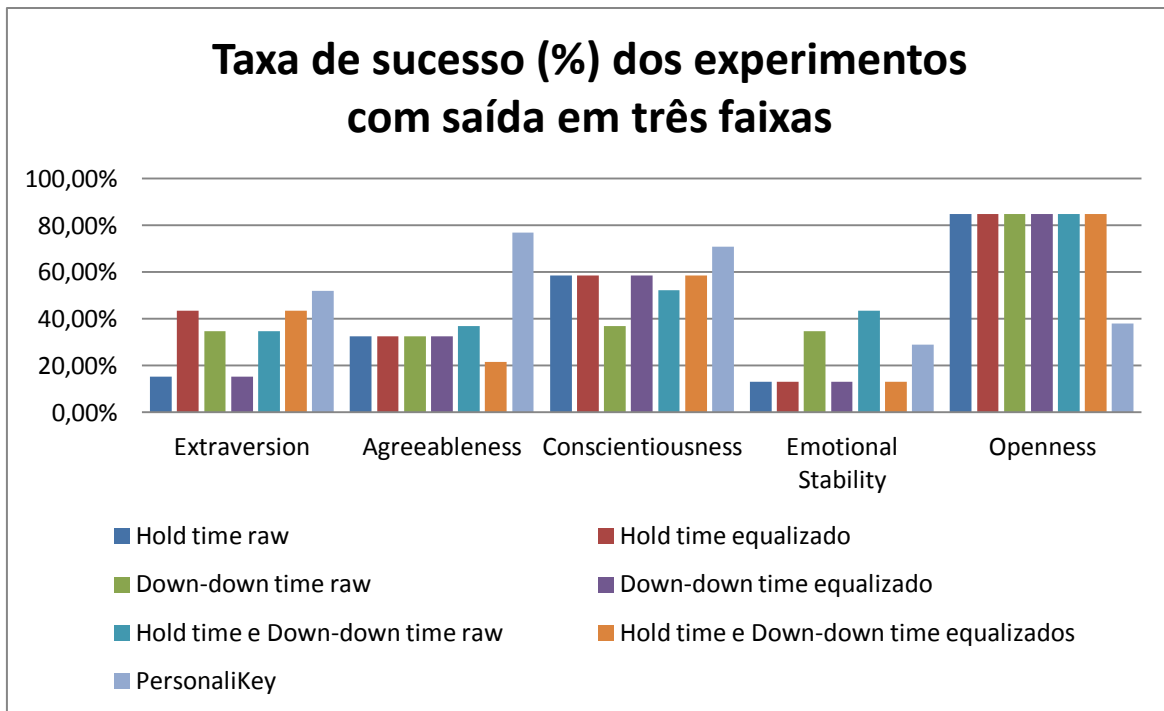


Gráfico 25: Comparação dos experimentos com saída classificada em três faixas de valor em relação ao PersonalKey (COSTA e PORTO, 2011)

O Gráfico 25 ilustra uma comparação entre as abordagens com três faixas de valor, juntamente com os resultados obtidos pelo *PersonalKey*. A comparação realizada com o *PersonalKey* é apenas uma comparação bruta de taxas de acerto, não sendo os dois experimentos diretamente equivalentes, tendo em vista que este estudo não teve acesso aos dados utilizados pelo *PersonalKey*, para que uma comparação pudesse ser realizada com abordagens diferentes aplicadas a um mesmo conjunto de dados.

4.4 Experimentos com saídas binárias

As redes neurais com saídas binárias foram desenvolvidas com o intuito de identificar quais traços de personalidade se destacam na personalidade de um indivíduo, através da *binarização* dos resultados obtidos pela rede, seguindo uma abordagem onde os valores mais representativos são convertidos para o valor binário 1, enquanto os valores menos representativos foram convertidos para o valor binário 0, assim como descrito na seção 3.2.

A abordagem binária permite uma identificação mais clara de quais traços de personalidade se destacam na personalidade de um indivíduo específico, permitindo assim à rede prever quais dos cinco traços de personalidade analisados estão em evidência em uma determinada personalidade, ou seja, foram classificados com um valor maior ou igual a 5 segundo na escala Likert de 7 pontos.

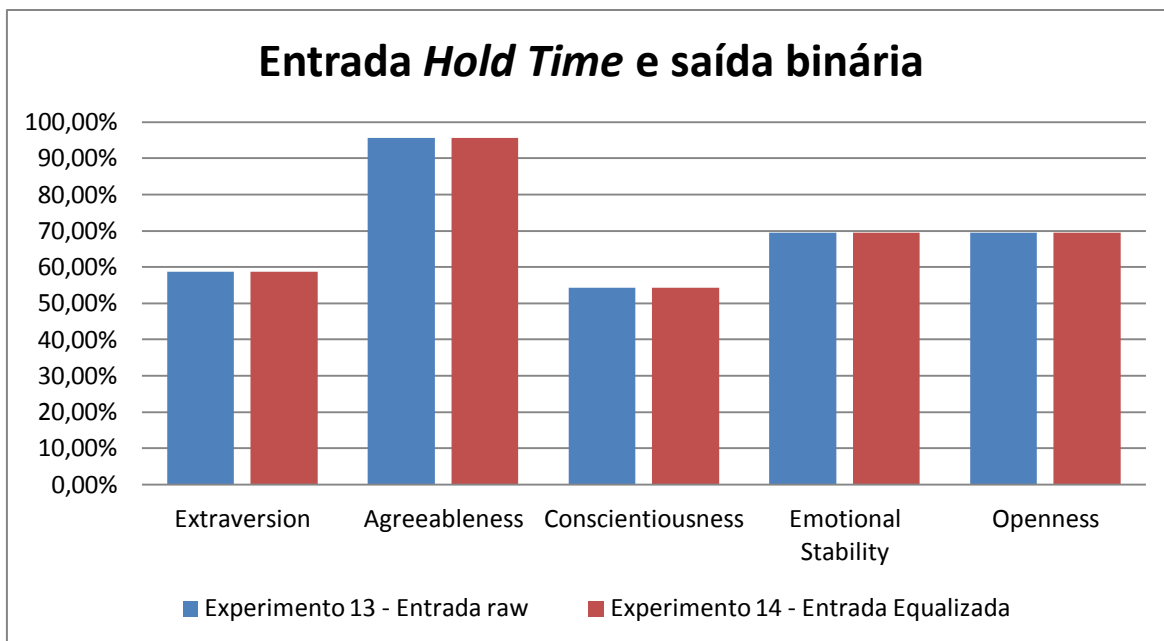


Gráfico 26: Comparação de resultados dos experimentos que utilizam o *hold time* como entrada e retornam a saída binária.

Mesmo com taxas de acerto atingindo valores como 95,65%, não se pode afirmar com certeza que a abordagem exposta no Gráfico 26 é capaz de indicar se os traços de personalidade estudados são passíveis de serem mapeados através da abordagem binária, ou seja, determinar se um traço de personalidade está ou não em destaque na personalidade de um indivíduo, sem antes validarmos os resultados obtidos em conjunto com a análise da distribuição de probabilidade entre as classes utilizadas nessa abordagem;

O Gráfico 26 e o Gráfico 27 apresentam as taxas de acerto da abordagem binária utilizando respectivamente os dados do *hold time* e os dados do *down-down time* como entrada das redes neurais, onde os melhores resultados obtidos para cada um dos traços de personalidade variou entre 50% na identificação do traço *conscientiousness* e 95,65% na identificação do traço de personalidade *agreeableness*, não havendo nenhuma variação significativa nos resultados entre as duas abordagens.

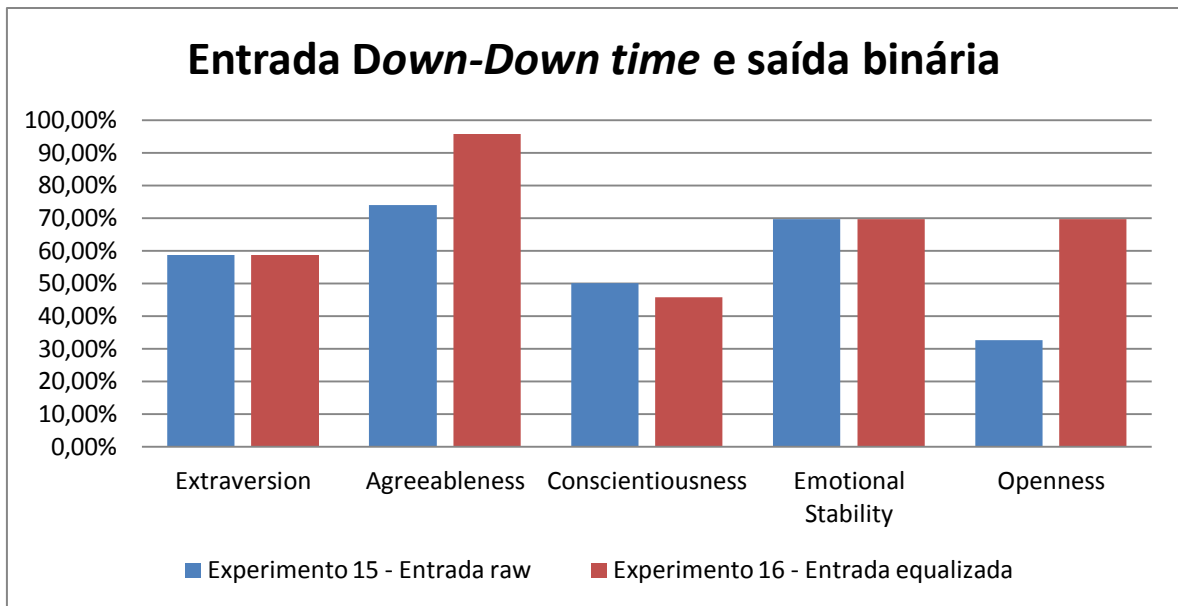


Gráfico 27: Comparação de resultados dos experimentos que utilizam o *down-down time* como entrada e retornam a saída binária.

Com taxas de acerto acima dos 95%, como é o caso do traço de personalidade *agreeableness* como mostram os gráficos Gráfico 26, Gráfico 27 e Gráfico 28, as abordagens binárias em geral apresentaram taxas de acerto consideradas elevadas em uma análise preliminar, porém, quando comparadas com a distribuição de probabilidade das classes analisadas, os resultados obtidos mostram-se insatisfatórios e conseqüentemente inconclusivos para o mapeamento dos traços de personalidade.

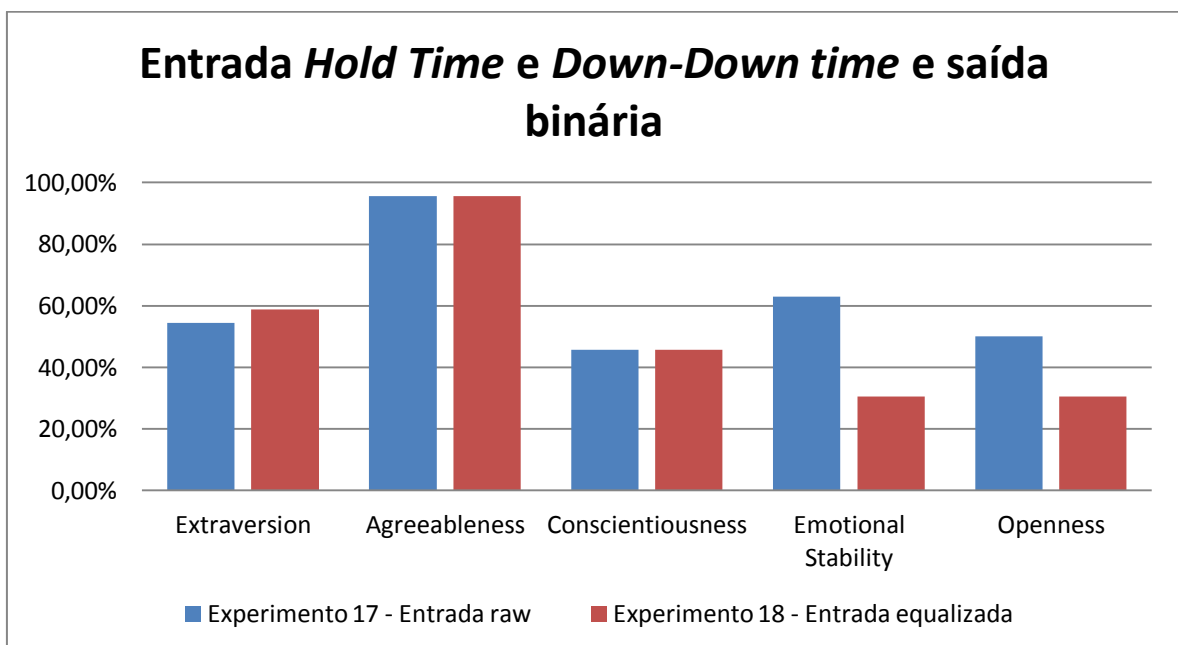


Gráfico 28: Comparação de resultados dos experimentos que utilizam os dados do *hold time* em conjunto com o *down-down time* como entrada e retornam a saída binária.

De forma semelhante ao que foi visto no Gráfico 21 e no Gráfico 25 apresentados em conjunto com as abordagens anteriores, o Gráfico 29 apresenta um comparativo entre os resultados obtidos nos experimentos realizados seguindo a abordagem binária.

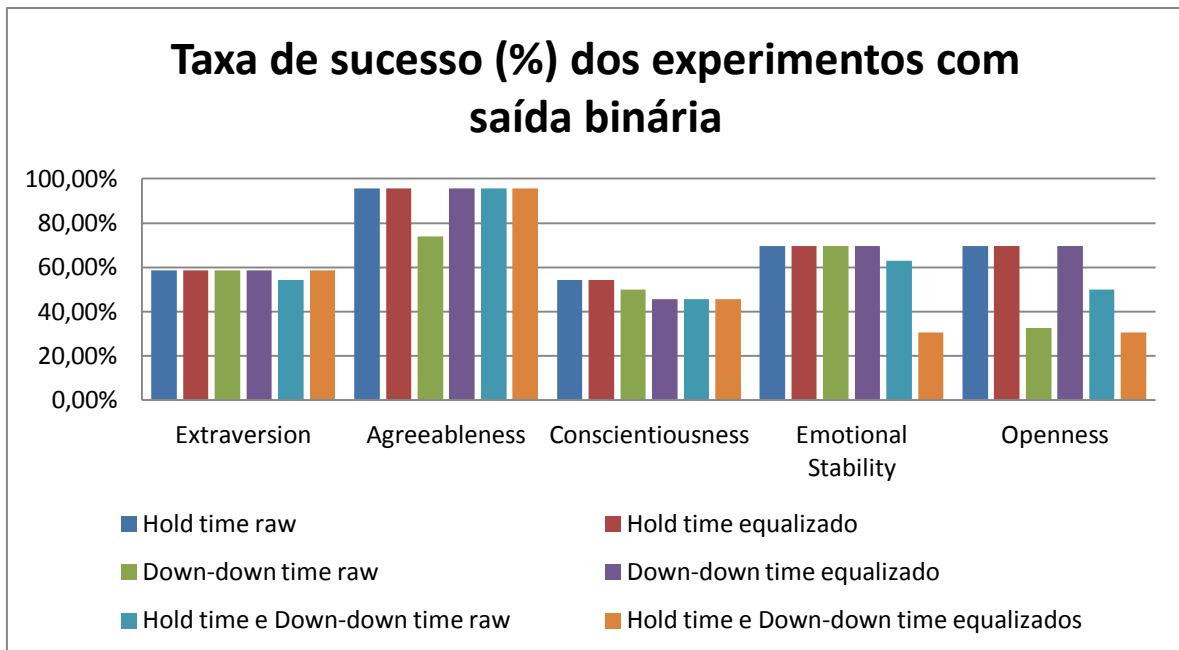


Gráfico 29: Comparação dos experimentos com saída binária

O Gráfico 29 expõem as taxas de acerto de cada um dos experimentos realizados através da aplicação da abordagem binária, cujos resultados variaram entre 30,43% e 95,65%. Embora tenham apresentado taxas de sucesso elevadas em alguns experimentos, de modo geral os resultados obtidos através dos experimentos realizados com as três abordagens estudadas não foram satisfatórios, no que tange o objetivo de mapear os traços de personalidade de um indivíduo através de características do ritmo de digitação. Dessa forma, visando verificar o quão significativo são os resultados obtidos, decidiu-se realizar testes adicionais voltados para uma análise estatística, que serão apresentados na seções 4.5 e 4.6.

4.5 Análise com conhecimento a priori

Como forma de comparar e validar a relevância dos resultados obtidos nos experimentos, decidiu-se realizar um experimento adicional, voltado para a realização de uma análise dos dados utilizados como saída esperada das redes neurais através de uma abordagem com conhecimento a priori, ou seja, onde a distribuição de probabilidade das classes analisadas é conhecida. Essa análise foi realizada para cada um dos três tipos de saída produzida pelas redes neurais (escala Likert, classificação em três faixas de valor e binária), com o objetivo de identificar qual das possíveis saídas mais se repete (moda) no conjunto de treinamento de cada uma das abordagens, e posteriormente utilizar essa resposta como única resposta para prever os dados do conjunto de teste. Dessa forma, visa-se comprovar que os experimentos desenvolvidos utilizando redes neurais não vão além de uma abordagem puramente estatística, não sendo estes resultados relevantes o suficiente para a predição dos traços de personalidade de um indivíduo. O Gráfico 30, Gráfico 31 e Gráfico 32 ilustram os resultados obtidos na comparação desses experimentos.

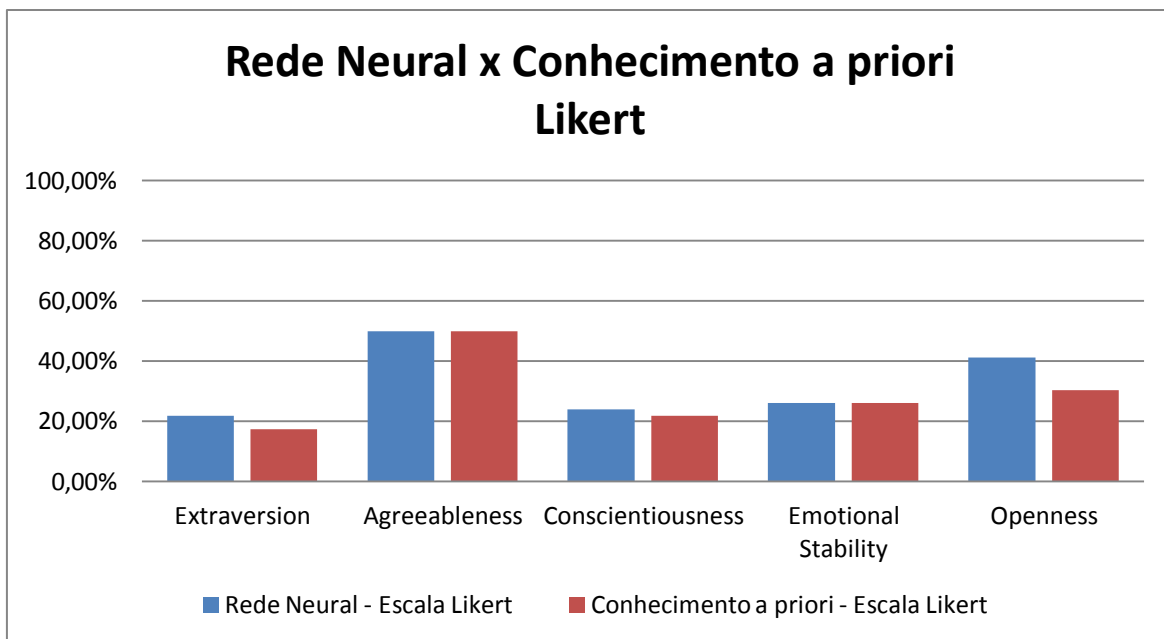


Gráfico 30: Comparação dos resultados na escala Likert, com uma abordagem com conhecimento a priori

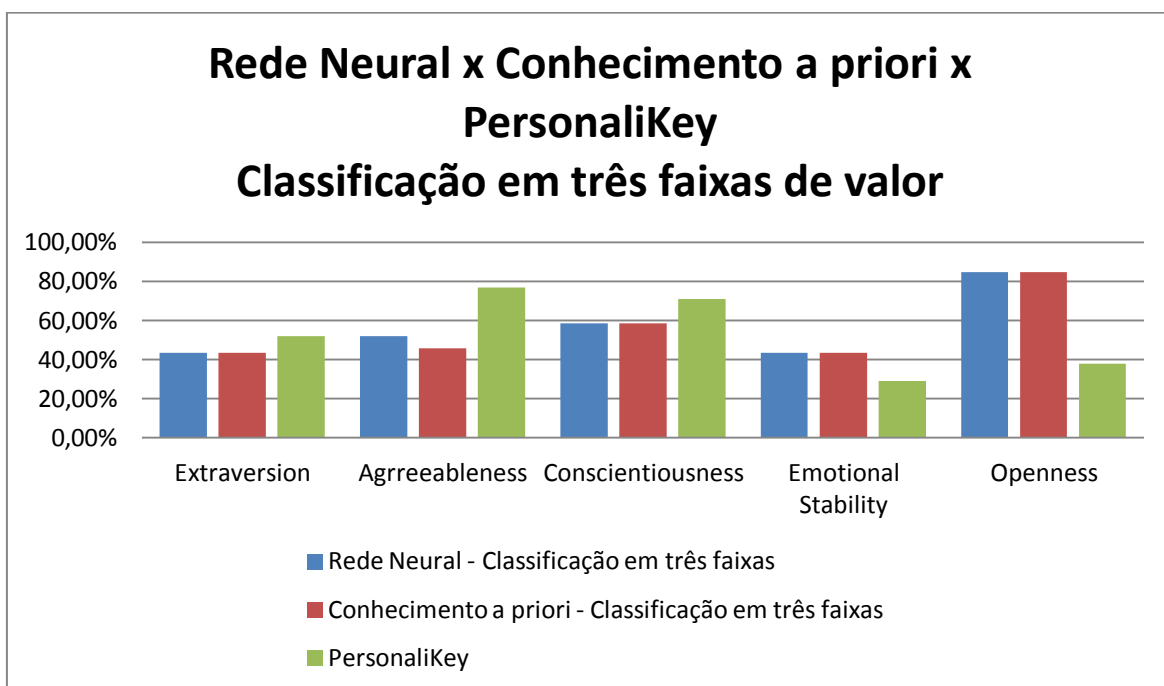


Gráfico 31: Comparação entre os melhores resultados classificados em três faixas, a abordagem com conhecimento a priori e os melhores resultados do PersonalKey

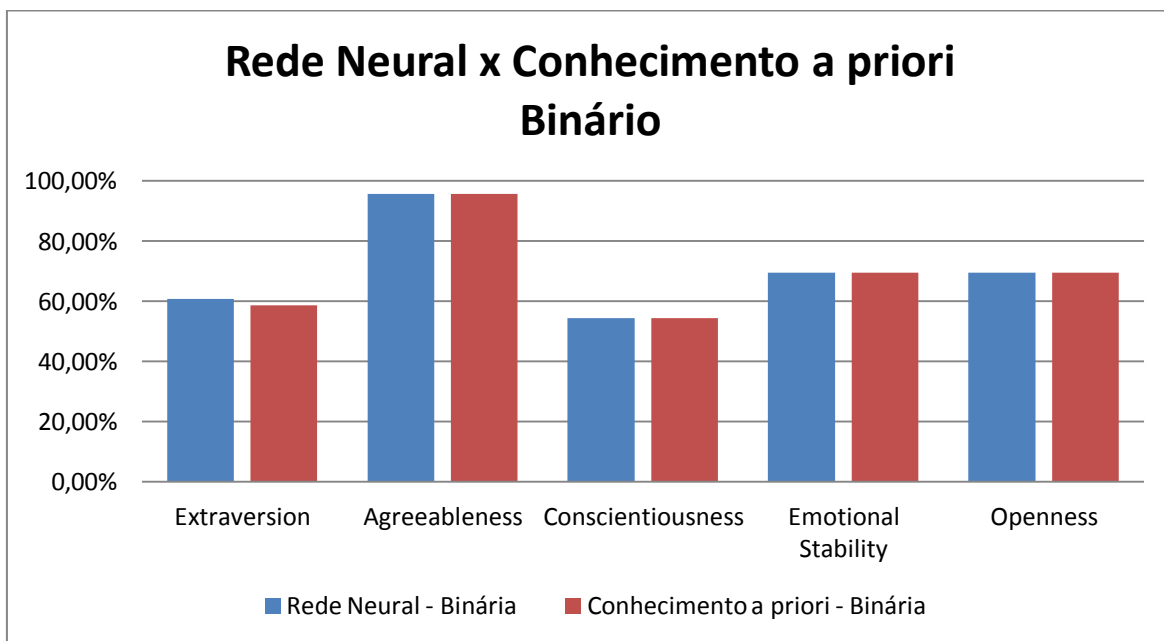


Gráfico 32: Comparação dos resultados binários, com uma abordagem com conhecimento a priori

Ao observar o Gráfico 30, Gráfico 31 e Gráfico 32, nota-se que, embora alguns dos resultados obtidos pelas redes neurais sejam superiores aos da análise com conhecimento a priori, não é possível afirmar com segurança que as redes neurais são capazes de inferir a personalidade de um indivíduo, através das abordagens realizadas.

4.6 Teste de hipótese

Visando verificar a confiabilidade dos resultados obtidos nos experimentos previamente apresentados, uma série de testes de hipótese foram realizados, um para cada traço de personalidade, em cada uma das abordagens desenvolvidas, visando confirmar que os resultados obtidos através dos experimentos são insatisfatórios para prever os traços da personalidade de um indivíduo através de características do ritmo de digitação.

Teste de hipótese é um método estatístico para análise de amostras através da teoria de probabilidades. Para a realização de um teste de hipótese é necessária a existência de duas hipóteses, conhecidas como (i) Hipótese nula (H_0); e (ii) Hipótese alternativa (H_1). A hipótese nula é a hipótese que assumimos como sendo verdadeira para a realização do teste de hipótese, enquanto a hipótese alternativa, é a hipótese que será considerada válida caso a hipótese nula seja rejeitada.

Durante a execução de um teste de hipótese dois tipos distintos de erros podem ocorrer. O erro tipo I é a probabilidade da hipótese nula (H_0) ser rejeitada caso ela seja efetivamente verdadeira. Já o erro tipo II é a probabilidade da hipótese alternativa (H_1) ser rejeitada quando ela é efetivamente verdadeira. Para os testes de hipótese realizados, temos como hipótese nula (H_0) a hipótese de que utilizar uma abordagem baseada em redes neurais é tão ou menos eficaz que escolher o resultado mais provável, enquanto a hipótese alternativa (H_1) indica o contrário, ou seja, utilizar uma abordagem baseada em redes neurais é mais eficaz que escolher o resultado mais provável.

Para a realização dos teste de hipótese, foi utilizada a fórmula $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$, onde \bar{x}

representa a taxa de acerto da rede neural, μ representa a taxa de acerto da abordagem com conhecimento a priori, ambos expresso em número de pessoas, n representa o tamanho da população do teste e σ representa o desvio padrão população, também expresso em número de pessoas e calculado a partir da variância de uma distribuição binomial, que pode ser representado pela equação $\sqrt{(n \cdot \mu(1 - \mu))}$. O teste de hipótese realizado nesse estudo foi construído com um nível de confiabilidade de 95%, e um nível de significância de 5%, sendo este um teste de hipótese unilateral à direita, com o intuito de provar que a hipótese alternativa (H_1) é verdadeira, obtendo-se resultados superiores ao da média da amostra analisada. Para um teste de hipótese com essas características, a região crítica referente ao nível de significância é representada por $p = 1,64$. Dessa forma, para que a hipótese alternativa seja considerada verdadeira, ou seja, rejeitar a hipótese nula, o resultado obtido através do teste de hipótese, representado por "z", deve ser superior a 1,64 ($z > p$).

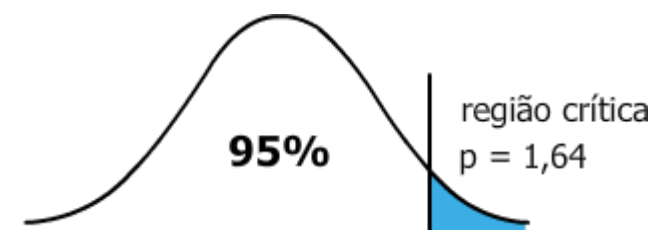


Figura 14: Exemplificação do nível de confiabilidade e da região crítica do teste de hipótese

As tabelas a seguir (Tabela 5, Tabela 6 e Tabela 7) apresentam os resultados obtidos através dos testes de hipótese realizados para cada uma das abordagens previamente desenvolvidas (escala Likert, três faixas de valor e binária), sendo "z" o valor resultante que deve ser comparado com o valor da região crítica do teste, visando verificar qual das hipóteses é verdadeira.

Para a realização dos testes de hipótese, utilizou-se o conjunto de teste (utilizado no treinamento das redes neurais), com tamanho $n = 46$. Os resultados exibidos na coluna "resultado da rede neural em porcentagem" são referentes aos melhores resultados (maior taxa de acerto) da abordagem na detecção do traço de personalidade em questão. Os valores de média ($\frac{\mu}{n}$) e desvio padrão (σ) foram calculados a partir da população do conjunto de teste, utilizando-se das equações descritas anteriormente.

A seguir são apresentadas as tabelas Tabela 5, Tabela 6 e Tabela 7 que respectivamente representam os testes de hipótese realizados com os dados referente a abordagem na escala Likert, abordagem com três faixas de valor e abordagem binária.

Traço de personalidade	Resultado da rede neural em porcentagem ($\frac{\bar{x}}{n}$)	Média esperada da população em porcentagem ($\frac{\mu}{n}$)	Desvio padrão (σ)	Tamanho da população (n)	Resultado do teste de hipótese (z)
Extraversion	0,2174	0,1739	2,570738348	46	5,2766

Agreeableness	0,5000	0,5000	3,391164992	46	0,00
Conscientiousness	0,2391	0,2174	2,797514425	46	2,4244
Emotional Stability	0,2609	0,2609	2,978181529	46	0,00
Openness	0,4130	0,3043	3,12075799	46	10,8665

Tabela 5: Teste de hipótese dos experimentos na abordagem com escala Likert

Traço de personalidade	Resultado da rede neural em porcentagem ($\frac{\bar{x}}{n}$)	Média esperada da população em porcentagem ($\frac{\mu}{n}$)	Desvio padrão (σ)	Tamanho da população (n)	Resultado do teste de hipótese (z)
Extraversion	0,4348	0,4348	3,362193901	46	0,00
Agreeableness	0,5217	0,4565	3,378319623	46	6,02
Conscientiousness	0,5870	0,5870	3,339487073	46	0,00
Emotional Stability	0,4348	0,4348	3,362193901	46	0,00
Openness	0,8478	0,8478	2,436140925	46	0,00

Tabela 6: Teste de hipótese dos experimentos na abordagem de três faixas de valor

Traço de personalidade	Resultado da rede neural em porcentagem ($\frac{\bar{x}}{n}$)	Média esperada da população em porcentagem ($\frac{\mu}{n}$)	Desvio padrão (σ)	Tamanho da população (n)	Resultado do teste de hipótese (z)
Extraversion	0,6087	0,5870	3,339487073	46	2,0309
Agreeableness	0,9565	0,9565	1,38312815	46	0,00
Conscientiousness	0,5435	0,5435	3,378319623	46	0,00
Emotional Stability	0,6957	0,6957	3,12075799	46	0,00
Openness	0,6957	0,6957	3,12075799	46	0,00

Tabela 7: Teste de hipótese dos experimentos na abordagem binária

Ao analisar o resultado de cada um dos testes de hipótese desenvolvidos, representados pela coluna "z" nas tabelas Tabela 5, Tabela 6 e Tabela 7, é possível perceber que apenas cinco deles obtiveram um resultado superior ao valor $p = 1,64$ que representa a região crítica dos testes. Dessa forma, com base nos resultados inferiores a $p = 1,64$, obtidos na maioria dos testes de hipótese, é possível afirmar com um nível de confiabilidade de 95% que os resultados obtidos com os testes de hipótese corroboram os resultados apresentados pelos próprios experimentos e pela abordagem com conhecimento a priori, em que as características extraídas através do ritmo de digitação não apresentam informações suficientes para a realização de um

mapeamento dos traços da personalidade de um indivíduo. Além disso, a similaridade entre os resultados obtidos através de uma abordagem baseada na escolha do traço de personalidade mais provável, e de uma abordagem baseada em redes neurais pode ser destacada a partir da quantidade de resultados (z) obtidos com valor zero, que representam uma igualdade nas taxas de acerto obtidas a partir da aplicação das abordagens supracitadas.

5 Conclusões

Neste trabalho foram realizados 18 tipos de experimentos diferentes, em busca de informações sobre a correlação do ritmo de digitação de um indivíduo e seus respectivos traços de personalidade. Para realização de tais experimentos, um *dataset* foi construído a partir do agrupamento de dados do ritmo de digitação extraído de 177 participantes voluntários, juntamente com características da personalidade de cada um, definidas como traços de personalidade, através de uma ferramenta de coleta de dados desenvolvida especificamente para a realização deste estudo, e disponibilizada online. Os experimentos realizados foram divididos em duas vertentes, de forma que pudessem responder à duas questões distintas **Q1**: É possível mensurar quanto um determinado traço de personalidade está presente na personalidade de um indivíduo? e **Q2**: É possível determinar quais traços de personalidade se destacam na personalidade do indivíduo?

Experimentos voltados para responder a questão **Q1** por sua vez foram divididos em outras duas vertentes, a primeira, voltada para a mensuração do valor de cada um dos traços de personalidade na personalidade de um indivíduo, enquanto a segunda preocupou-se em classificar o valor mensurado de cada traço de personalidade em três faixas de valores distintas, visando determinar se tal traço de personalidade representava uma influência de nível baixo, médio ou alto na personalidade do indivíduo. O primeiro experimento voltado para a mensuração do valor que um traço de personalidade tem sobre a personalidade de um indivíduo obteve como resultado as seguintes taxas de acerto para os traços de personalidade analisados: Extroversão 21,74%; Socialização 50,00%; Realização 23,91%; Neuroticismo 26,09% e Abertura 41,30%.

O segundo modelo de experimentos também voltados para responder à **Q1** foram inicialmente realizados de forma semelhante aos experimentos anteriores, porém, mapeando o resultado obtido para cada um dos traços de personalidade em uma das três faixas de valor predeterminadas, que representam traços de personalidade uma influência baixa na personalidade do indivíduo, traços de personalidade com influência média e traços de personalidade com alto nível de influência. Uma das grandes vantagens dessa abordagem para este estudo, é que além de permitir que o mapeamento dos traços de personalidade dos participantes fosse realizado de forma mais abrangente, bastando que a rede neural predissesse um valor que representasse a mesma classificação (baixo, médio ou alto) do resultado esperado pelo processo de validação da rede neural, essa abordagem também permitiu que os resultados obtidos pelos experimentos fossem comparados aos resultados obtidos pela ferramenta *PersonaliKey*, desenvolvida no estudo realizado por (COSTA e PORTO, 2011), a partir do qual a abordagem de mapeamento dos traços de personalidade em três faixas de valor foi inspirada.

Ao analisar os resultados obtidos por este experimento, principalmente quando comparado aos resultados do *PersonaliKey*, as taxas de acerto obtidas em sua maioria não apresentam grande relevância, sendo elas: Extroversão 43,48%; Socialização 52,17%; Realização 58,70%; Neuroticismo 43,48% e Abertura 84,78%. Além de pouco expressivos, os resultados obtidos nessa abordagem, assim como nas demais abordagens, não obtiveram um resultado satisfatório para o mapeamento de traços de personalidade através do ritmo de digitação, estando de modo geral quase iguais aos resultados da abordagem com conhecimento a priori.

Para responder à questão **Q2**, foi desenvolvida uma abordagem binária onde o objetivo dos experimentos era identificar se um traço de personalidade está ou não em destaque na

personalidade um indivíduo. Nessa abordagem, traços de personalidade cujos resultados mapeados na escala Likert, a partir do questionário TIPI de autoavaliação da personalidade, fossem maiores ou iguais ao valor 5, seriam convertidos para o valor binário 1, enquanto os valores inferiores a 5 seriam convertidos para o valor binário 0. Esse tipo de experimento binarizado resultou em taxas de acerto elevadas quando comparadas aos demais experimentos, sendo: Extroversão 60,87%; Socialização 95,65%; Realização 54,35%; Neuroticismo 69,57% e Abertura 69,57%. Porém, o incremento observado em relação aos demais experimentos é devido à redução na quantidade de classes analisadas no experimentos (apenas duas classes, zero ou um), e a distribuição de probabilidade entre as classes, resultando em taxas de acerto equivalentes as da abordagem com conhecimento a priori.

Por fim, ao analisar os resultados obtidos em cada um dos experimentos realizados, nota-se não ter sido possível mapear de forma satisfatória os traços de personalidade de um indivíduo, a partir do ritmo de digitação, utilizando-se das abordagens descritas neste trabalho. Dessa forma, fica evidente que a utilização exclusiva de características do ritmo de digitação não apresentou informações suficientes para a realização do mapeamento dos traços da personalidade de indivíduo.

6 Trabalhos futuros

Mesmo conhecendo as limitações de um teclado convencional de computador, como fonte de informação de características capazes de diferenciar indivíduos entre si, este tipo de abordagem é incentivada por (SOLANKI e SHUKLA, 2014; NAHIN et al, 2014; KOLAKOWSKA et al, 2013), que confirmam os benefícios da utilização do ritmo de digitação, por ter como equipamento necessário um teclado convencional de computador, sendo este pouco dispendioso e já vastamente utilizado na maioria dos sistemas computacionais, além de ser uma alternativa de abordagem de extração automática, não intrusiva e facilmente adaptável para diferentes sistemas computacionais.

Devido aos resultados insatisfatórios, acredita-se não ser possível mapear com clareza os traços da personalidade de um indivíduo através do ritmo de digitação. Por outro lado, acredita-se na possibilidade de sucesso no desenvolvimento de estudos voltados para novas abordagens e experimentos focados no mapeamento de informação com uma maior relação com as funções motoras do ser humano, como por exemplo as emoções e o estado emocional, assim como apresentando em (ZIMERMANN et al, 2013). Tal trabalho pode ser realizado a partir da adaptação da ferramenta de coleta de dados já desenvolvida, utilizando-se do mesmo processo de coleta de dados do ritmo de digitação, adaptando apenas o questionário de autoavaliação utilizado, de forma que um questionário de autoavaliação do estado emocional seja exibido.

7 Bibliografia

ARYA, A. et al. *A text analysis based seamless framework for predicting human personality traits from social networking sites*. *International Journal of Information Technology and Computer Science (IJITCS)*, v. 4, n. 10, p. 29, 2012.

ALLPORT, G. W. *Pattern and Growth in Personality*. New York: Holt, Rinehart and Winston, Inc., 1961.

BAHRAMPOUR, S. et al. *Comparative Study of Deep Learning Software Frameworks*. *Research and Technology Center, Robert Bosch LLC*, 2016.

CARVER, C. S. SCHEIER, M. F. *Personality is a dynamic organization, inside the person, of psychophysical systems that create the person's characteristic patterns of behavior, thoughts and feelings*. *Perspectives on personality*. Boston, 2000.

CHUDÁ, D.; KRÁTKY, P. *Usage of computer mouse characteristics for identification in web browsing*. *Proceedings of the 15th International Conference on Computer Systems and Technologies*. ACM, p. 218-225, 2014.

COSTA, W. S.; PORTO, S. M. **PersonaliKey: Uma ferramenta de extração de traços de personalidade através do ritmo de digitação**. Universidade Federal de Sergipe. Aracaju, 2011.

EPP, C.; LIPPOLD, M.; MANDRYK, R. L. *Identifying emotional states using keystroke dynamics*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, p. 715-724, 2011.

GILL, A. J.; OBERLANDER, J. *Perception of e-mail personality at zero-acquaintance extraversion takes care of itself; neuroticism is a worry*. *Proceedings of the 25th Annual Conference, Hillsdale, NJ*, p. 456-461, 2003.

GOSLING, S. D.; MEHL, M. R.; PENNEBAKER, J. W. *Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life*. *Journal of Personality and Social Psychology*, 2006.

GOSLING, S. D.; RENTFROW, P. J.; SWANN, W. B. *A very brief measure of the Big-Five personality domains*. *Journal of Research in Personality*, v. 37, n. 6, p. 504-528, 2003.

HAYKIN, S. **Redes Neurais: Princípios e prática**. 2 ed. Prantice Hall, Inc., 1999.

HERNANDEZ, J. et al. *Under pressure: sensing stress of computer users*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, p. 51-60, 2014.

IBM, *Personality Insights, Watson Developer Cloud*. Disponível em <<http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/personality-insights.html>> Acessado em 01/08/2015.

KAKLAUSKAS, A; KRUTINIS, M; SENIUT, M. *Biometric mouse intelligent system for student's emotional and examination process analysis*. In: *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on*. IEEE. p. 189-193, 2009.

KHAN, I. A. et al. **Measuring personality from keyboard and mouse use**. *15h European conference on Cognitive ergonomics: the ergonomics of cool interaction*, 2008. 184-191p.

KHAN, I. A. et al. **Predicting programmers' personality via interaction behaviour with keyboard and mouse**. PeerJ PrePrints, v. 3, p. e1441, 2015.

KHANNA, P; SASIKUMAR, M. **Recognising emotions from keyboard stroke pattern**. *International journal of computer applications*, v. 11, n. 9, p. 1-5, 2010.

KOLAKOWSKA, A. et al. **Emotion recognition and its application in software engineering**. In: *Human System Interaction (HSI), The 6th International Conference on. IEEE*, p. 532-539, 2013.

LALI, P. et al. **Your mouse can tell about your emotions**. In: *Computer and Knowledge Engineering (ICCKE), 4th International eConference on. IEEE*. p. 47-51, 2014.

LIKERT, R. A. **Technique for the Measurement of Attitudes**. *Archives of Psychology*, p. 1-55, 1932.

MONTALVÃO FILHO, J. R.; FREIRE, E. O. **On the equalization of keystroke timing histograms**, *Pattern Recognition Letters*, v. 27, p. 1440-1446, 2006.

MONTALVÃO FILHO, J. R. et al. **Contributions to empirical analysis of keystroke dynamics in passwords**. *Pattern Recognition Letters*, v. 52, p. 80-86, 2015.

NAHIN, A. N. H. et al. **Identifying emotion by keystroke dynamics and text pattern analysis**. *Behaviour & Information Technology*, v. 33, n. 9, p. 987-996, 2014.

NUNES, M. A. S. N.; BEZERRA, J. S.; de OLIVEIRA, A. A. **PersonalityML: a markup language to standardize the user personality in Recommender Systems**. *GEINTEC-Gestão, Inovação e Tecnologias*, v. 2, n. 3, p. 255-273, 2012.

NUNES, M. A. S. N. et al. **Computação Afetiva e sua influência na personalização de Ambientes Educacionais: gerando equipes compatíveis para uso em AVA na EaD**. *Educação e Ciberespaço: Estudos, propostas e desafios*., Aracaju, SE, v. 1, p. 308-347, 2010.

NUNES, M. A. S. N.; HU, R. **Personality-based Recommender Systems: An overview**. In: *ACM Conf. on Recommender System. Dublin. Proceedings of (RecSys '12)*. New York: ACM, p. 5-7, 2012.

NUNES, M. A. S. N.; CAZELLA, S. C. **O que a Personalidade revela? Fidelizando clientes web através de Sistemas de Recomendação e Traços de Personalidade**. In: *(Webmedia 2011 Minicursos): Patricia Vilain e Valter Roesler. (Org.)*. Tópicos em Banco de Dados e Multimídia e Web. Porto Alegre: SBC, v. 1, p. 91-122, 2011.

PERVIN, L. A.; CERVONE, D; JOHN, O. **Personality: Theory and Research**. 9 ed., Wiley, 2004.

PICARD, R. W. **Affective Computing**. Mit Media Laboratory Perceptual Computing. Technical Report n. 321, 1995.

PORTO, S. M.; NUNES, M. A. S. N. **Desenvolvimento de Metodologias de Extração de Perfil Psicológico de Usuário para Aplicação em Sistemas de Recomendação objetivando Personalização de Produtos e Serviços em E-commerce.** Universidade Federal de Sergipe. Aracaju, 2011.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais: Para engenharia e ciências aplicadas.** São Paulo, Artliber Editora, 2010.

SOLANKI, R.; SHUKLA, P. *Estimation of the User's Emotional State by Keystroke Dynamics.* *International Journal of Computer Applications*, v. 94, p. 21-23, 2014.

SHEPHERD, S. J. *Continuous authentication by analysis of keyboard typing characteristics.* *Security and Detection*, p. 111-114, 1995.

STATHOPOULOU, I. O.; ALEPIS, E.; TSIHRINTZIS, G. A.; VIRVOU, M. *On assisting a visual-facial affect recognition system with keyboard-stroke pattern information.* *Knowledge-Based Systems*, p. 350-356, 2010.

ZIMMERMANN, P. et al. *Affective computing—a rationale for measuring mood with mouse and keyboard.* *International journal of occupational safety and ergonomics*, v. 9, n. 4, p. 539-551, 2003.