



UNIVERSIDADE DO ESTADO DO AMAZONAS - UEA
ESCOLA SUPERIOR DE TECNOLOGIA - EST
CURSO DE ENGENHARIA DE COMPUTAÇÃO



**DESIGN ACESSÍVEL ORIENTADO POR DADOS: CLASSIFICAÇÃO AUTOMÁTICA DE
DESCRIÇÕES COMPORTAMENTAIS DO PERSONAUT VIA PLN**

YARA NUNES DE OLIVEIRA

Manaus – AM

Junho – 2026

YARA NUNES DE OLIVEIRA

**DESIGN ACESSÍVEL ORIENTADO POR DADOS: CLASSIFICAÇÃO AUTOMÁTICA DE
DESCRIÇÕES COMPORTAMENTAIS DO PERSONAUT VIA PLN**

Trabalho de Conclusão de Curso apresentado
à banca avaliadora do Curso de Engenharia
de Computação, da Escola Superior de
Tecnologia, da Universidade do Estado do
Amazonas, como pré-requisito para obtenção
do título de Engenheira de Computação.

Orientadora: Profa. Dra. Aurea Hiléia da Silva Melo

Manaus – AM

Junho – 2026

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).
Sistema Integrado de Bibliotecas da Universidade do Estado do Amazonas.

O48d	Oliveira, Yara Nunes de Design acessível orientado por dados: Classificação automática de descrições comportamentais do PersonAut via PLN / Yara Nunes de Oliveira. Manaus : [s.n], 2026. 68 f.: color.; 21.0 cm. TCC - Graduação em Engenharia de Computação- Universidade do Estado do Amazonas, Manaus, 2026. Inclui Bibliografia. Inclui Apêndice. Orientador: Áurea Hiléia da Silva Melo. 1. TEA. 2. PLN. 3. Classificação Automática. 4. PersonAut. 5. ProAut. I. Áurea Hiléia da Silva Melo (Orient.) II. Universidade do Estado do Amazonas. III. Título CDU(1997)62:004
------	---



FOLHA DE APROVAÇÃO

DESIGN ACESSÍVEL ORIENTADO POR DADOS: CLASSIFICAÇÃO AUTOMÁTICA DE DESCRIÇÕES COMPORTAMENTAIS DO PERSONAUT VIA PLN

YARA NUNES DE OLIVEIRA

Trabalho de Conclusão de Curso defendido e aprovado pela banca avaliadora constituída pelos professores:

Profª Dra. Áurea Hiléia da S. Melo
Presidente

Prof.Dr. Luis Cuevas Rodrigues
Avaliador

Prof. Dr. Carlos Maurício Serodio Figueiredo
Avaliador

Manaus, 19 de junho de 2026

Dedicatória

Dedico este trabalho à minha mãe, Maria do Socorro, e ao meu pai, Francisco, que por muitas vezes tiveram suas asas cortadas, mas sempre fizeram questão de construir aviões para que eu pudesse voar. Cada conquista alcançada carrega o esforço, os sacrifícios e o amor que dedicaram à minha formação. Sem vocês, este sonho não seria possível. Esta conquista também é de vocês.

Aos meus amigos de graduação, que compartilharam comigo não apenas os desafios acadêmicos, mas também momentos, aprendizados e experiências que levarei para a vida. A presença de vocês tornou essa jornada mais leve e especial.

À professora Áurea Melo, pelos ensinamentos e paciência durante o desenvolvimento deste projeto.

Resumo

O desenvolvimento de aplicações voltadas ao público autista ainda é um desafio para designers e desenvolvedores leigos. O ProAut é um processo que apoia esses profissionais no projeto de interfaces de aplicações para autistas. O GuideAut, por sua vez, é a plataforma criada para apoiar o uso do ProAut, reunindo todos os artefatos necessários para seguir seu processo, entre eles o PersonAut, um modelo comportamental utilizado para gerar perfis de usuários autistas e auxiliar na construção de empatia. Nesse contexto, este trabalho propõe utilizar técnicas de Processamento de Linguagem Natural (PLN) para classificar automaticamente descrições comportamentais fornecidas por familiares/profissionais e organizá-las nas categorias que compõem o PersonAut. O objetivo principal é desenvolver um modelo capaz de identificar, a partir de textos naturais, dimensões como Atividades que Acalmam, Atividades que Estressam, Aspectos Sociais e Familiares, Estereotipias/Manias e Aspectos Tecnológicos, permitindo o apoio na geração do PersonAut dentro da plataforma GuideAut. Para isso, adotou-se o fluxo padrão de ciência de dados, incluindo revisão bibliográfica, definição de requisitos, coleta e pré-processamento dos dados, seleção e treinamento de modelos de PLN e preparação para integração com o GuideAut. Como principal contribuição, espera-se que os resultados auxiliem profissionais que utilizam a plataforma, automatizando a construção de perfis comportamentais estruturados e contribuindo indiretamente para o desenvolvimento de soluções mais acessíveis e adequadas às necessidades de pessoas autistas.

Palavras-Chave: TEA, PLN, Classificação Automática, PersonAut, ProAut.

Abstract

Developing applications for autistic audiences remains a challenge for lay designers and developers. ProAut is a process that supports these professionals in designing application interfaces for autistic individuals. GuideAut, in turn, is a platform created to support the use of ProAut, bringing together all the necessary artifacts to follow its process, including PersonAut, a behavioral model used to generate profiles of autistic users and assist in building empathy. In this context, this work proposes to use Natural Language Processing (NLP) techniques to automatically classify behavioral descriptions provided by family members/professionals and organize them into the categories that make up PersonAut. The main objective is to develop a model capable of identifying, from natural texts, dimensions such as Calming Activities, Stressful Activities, Social and Family Aspects, Stereotypes/Manias, and Technological Aspects, allowing support in the generation of PersonAut within the GuideAut platform. To achieve this, the standard data science workflow was adopted, including literature review, requirements definition, data collection and preprocessing, selection and training of NLP models, and preparation for integration with GuideAut. The main contribution is expected to be that the results will assist professionals using the platform by automating the construction of structured behavioral profiles and indirectly contributing to the development of more accessible and suitable solutions for the needs of autistic individuals.

Keywords: ASD, NLP, Automatic Classification, PersonAut, ProAut.

Sumário

Lista de Tabelas	viii
Lista de Quadros	x
Lista de Figuras	xi
Introdução	1
1.1 Justificativa	2
1.2 Objetivos	3
Objetivo geral	3
Objetivos específicos	4
1.3 Metodologia	4
Fundamentação Teórica	6
2.1 Autismo e acessibilidade digital	6
2.2 Processamento de Linguagem Natural	8
2.3 Classificação de texto	8
2.4 ProAut	9
2.4.1 PersonAut	10
2.4.2 CRISP-DM	12
2.5 Trabalhos relacionados	13
Desenvolvimento	17
3.1 Entendimento do Problema	17

3.2	Entendimento e Preparação dos Dados	19
3.2.1	Construção do Dataset	20
3.2.2	Coleta de Dados	20
3.2.3	Definição das Categorias e Anotação	21
3.2.4	Distribuição do Dataset	22
3.2.5	Divisão entre Treino e Teste	23
3.3	Implementação do Pipeline de PLN	24
3.3.1	Normalização e Limpeza	24
3.3.2	Vetorização e Representação Textual	25
3.4	Modelagem	26
3.4.1	Seleção do Modelo Base	26
3.4.2	Resultados da Validação Cruzada	27
3.4.3	Avaliação dos Modelos Clássicos	28
3.4.4	Análise da Matriz de Confusão	31
3.4.5	Resultados com Embeddings Contextuais	33
3.5	Avaliação	35
3.6	Implantação	37
3.6.1	Prova de Conceito da Integração	40
3.7	Síntese Geral dos Resultados	41

Considerações Finais	43
-----------------------------	-----------

A Formulário aplicado aos responsáveis	45
---	-----------

Lista de Tabelas

3.1	Resultados médios da validação cruzada	28
3.2	Comparação geral entre os modelos avaliados	36
3.3	Resultados da prova de conceito da solução integrada	41

Lista de Quadros

2.1	Comparação entre os trabalhos relacionados e a proposta deste trabalho	16
3.2	Requisitos Funcionais do Sistema	18
3.3	Requisitos Não Funcionais do Sistema	19
3.4	Exemplos de heterogeneidade textual presentes nas respostas	19
3.5	Exemplo de respostas coletadas	21
3.6	CrITÉrios utilizados para anotação das categorias	22

Lista de Figuras

1.1	Fluxo metodológico adotado no projeto.	5
2.2	Modelo PersonAut em sua versão descritiva. Adaptado de Melo (2021).	12
2.3	Modelo PersonAut em sua versão sintética. Adaptado de Melo (2021).	12
3.4	Distribuição das categorias no dataset	23
3.5	Distribuição das categorias nos conjuntos de treino e teste	24
3.6	Exemplo de normalização e limpeza textual	25
3.7	Distribuição bootstrap do F1-score macro do modelo TF-IDF + SVM	29
3.8	F1-score por categoria utilizando TF-IDF + SVM	30
3.9	Precisão e recall por categoria	30
3.10	Matriz de confusão do modelo TF-IDF + SVM	32
3.11	Matriz de confusão normalizada	33
3.12	Comparação entre modelos clássicos e embeddings contextuais considerando F1-score macro (a) e accuracy (b).	35
3.13	Interface para inserção das descrições comportamentais e classificação automática via PLN.	38
3.14	Preenchimento dos dados demográficos da persona.	39
3.15	PersonAut gerado após a classificação automática das descrições comportamentais.	40

Introdução

O desenvolvimento de tecnologias centradas no usuário tem como princípio fundamental a empatia (NORMAN; DRAPER, 1986), isto é, a capacidade de compreender as necessidades, limitações e expectativas dos usuários. No contexto da Interação Humano Computador (IHC), essa compreensão é essencial para projetar interfaces que sejam não apenas funcionais, mas também acessíveis, inclusivas e alinhadas às características do público-alvo. Entre as diversas abordagens utilizadas na IHC para promover empatia durante o processo de design, destaca-se o uso de personas. Personas são representações fictícias, porém fundamentadas em dados reais, que sintetizam comportamentos e desafios de grupos de usuários (PRUITT; ADLIN, 2006).

Elas auxiliam equipes de desenvolvimento a manter o foco nas necessidades reais das pessoas ao longo do ciclo do projeto. Quando se trata do desenvolvimento de aplicações voltadas para pessoas autistas, essa compreensão se torna ainda mais necessária. O Transtorno do Espectro Autista (TEA) envolve características sensoriais, cognitivas, sociais e comportamentais que variam significativamente entre os indivíduos (American Psychiatric Association, 2014). Cada pessoa apresenta particularidades que impactam diretamente sua interação com sistemas digitais, exigindo abordagens de design inclusivas.

Nesse contexto, o processo ProAut(MELO, 2021) propõe uma metodologia estruturada para apoiar o desenvolvimento de interfaces destinadas ao público autista. Entre seus artefatos, destaca-se o PersonAut(MELO et al., 2020b), um modelo comportamental que organiza informações relevantes sobre o usuário em dimensões específicas, promovendo maior empatia e direcionamento durante o design da solução. Entretanto, a construção do PersonAut depende de descrições comportamentais detalhadas, geralmente fornecidas por familiares ou responsáveis.

À medida que o volume dessas informações cresce, torna-se inviável analisar e categorizá-las manualmente. A ausência de uma sistematização adequada pode dificultar a identificação de padrões relevantes e tornar o processo lento e suscetível a inconsistências.

Diante desse cenário, o uso de técnicas de Processamento de Linguagem Natural (PLN) apresenta-se como uma alternativa promissora. Por meio de modelos de classificação supervisionada, é possível analisar automaticamente descrições textuais e categorizá-las em diferentes classes comportamentais (AGGARWAL; ZHAI, 2012; JURAFSKY; MARTIN, 2024) como Atividades que Acalmam, Atividades que Estressam, Aspectos Sociais e Familiares, Estereótipos/Manias e Aspectos Tecnológicos, permitindo a organização estruturada das informações.

Assim, este Trabalho de Conclusão de Curso propõe o desenvolvimento e a aplicação de técnicas de PLN para classificar descrições comportamentais fornecidas por familiares de pessoas autistas, com o objetivo de apoiar a construção estruturada do PersonAut na plataforma GuideAut. A pesquisa segue o fluxo da ciência de dados, contemplando revisão bibliográfica, coleta e preparação dos dados, definição das categorias, treinamento de modelos de classificação e avaliação de desempenho.

A relevância deste estudo reside tanto no avanço técnico proporcionado pela aplicação de PLN quanto no impacto social decorrente da melhoria no processo de construção de perfis comportamentais. Ao apoiar a sistematização dessas informações, espera-se contribuir para o desenvolvimento de interfaces mais acessíveis, informadas e alinhadas às reais necessidades desse público.

1.1 Justificativa

A criação de interfaces acessíveis para pessoas autistas é uma necessidade urgente, uma vez que esse público frequentemente enfrenta barreiras no uso de tecnologias devido a designs inadequados (CHEN et al., 2016). A plataforma GuideAut já representa um avanço nesse sentido, ao proporcionar um ambiente colaborativo para apoiar o processo ProAut e disponibilizar artefatos essenciais, como o PersonAut. Entretanto, a construção manual desse modelo comportamental depende da análise detalhada de descrições fornecidas por familiares, o que se torna inviável à

medida que o volume de dados cresce.

A aplicação de técnicas de Processamento de Linguagem Natural (PLN) surge como uma solução viável e inovadora para esse desafio, permitindo a automação da categorização dessas descrições e a organização estruturada do perfil comportamental. Do ponto de vista tecnológico, a aplicação de PLN em um contexto específico, como a identificação de dimensões comportamentais relacionadas ao TEA, contribui para o avanço do conhecimento na área, explorando técnicas de classificação de textos em um cenário desafiador e relevante (JURAFSKY; MARTIN, 2024; INDURKHYA; DAMERAU, 2010; AGGARWAL; ZHAI, 2012).

Do ponto de vista prático, a automação da classificação de descrições comportamentais utilizadas no PersonAut otimiza o processo de construção desse artefato, beneficiando diretamente desenvolvedores e designers que dependem dele para compreender necessidades sensoriais, sociais e comportamentais de pessoas autistas. Por fim, do ponto de vista social, o projeto promove a inclusão digital, ao facilitar o processo de desenvolvimento de interfaces mais adequadas às características desse público.

Além disso, o projeto alinha-se a demandas atuais por tecnologias mais acessíveis e inclusivas, conforme destacado em iniciativas globais como a Convenção sobre os Direitos das Pessoas com Deficiência da ONU (2006). Esses fatores evidenciam a importância e a viabilidade do projeto, reforçando sua relevância para a comunidade acadêmica e para a sociedade em geral.

1.2 Objetivos

Esta seção apresenta o objetivo geral e os objetivos específicos que orientam o desenvolvimento desta pesquisa.

Objetivo geral

O principal objetivo deste trabalho é desenvolver e aplicar técnicas de Processamento de Linguagem Natural (PLN) para viabilizar a classificação automática de descrições comportamentais fornecidas por familiares de pessoas autistas, apoiando designers e desenvolvedores na etapa do

PersonAut no design de interfaces destinadas a pessoas autistas.

Objetivos específicos

Para alcançar o objetivo geral proposto, foram definidos os seguintes objetivos específicos:

1. Fazer um levantamento de descrições comportamentais relacionadas a pessoas autistas.
2. Construir um dataset anotado relacionado a pessoas autistas, estruturado de acordo com as dimensões que compõem o PersonAut, a partir de dados textuais coletados de familiares e responsáveis.
3. Desenvolver, treinar e avaliar um modelo de Processamento de Linguagem Natural (PLN) baseado em classificação supervisionada para categorizar descrições comportamentais nas dimensões que compõem o PersonAut, utilizando métricas de validação coerentes.
4. Integrar com a plataforma GuideAut, permitindo a geração apoiada por modelo do PersonAut, a partir das informações fornecidas pelos usuários.

1.3 Metodologia

Para o desenvolvimento desta pesquisa foi utilizada a metodologia CRISP-DM (CHAPMAN et al., 2000). O trabalho foi conduzido considerando as seis etapas, que permitiram organizar o projeto desde a definição do problema até a integração com a plataforma GuideAut:

- **Entendimento do Problema:** responsável pela definição do contexto da pesquisa e dos objetivos da solução proposta. A partir disso, são definidos os requisitos do sistema e o escopo da aplicação.
- **Entendimento dos Dados:** consiste em realizar uma exploração inicial das respostas coletadas. Essa análise tem como objetivo compreender a estrutura dos textos e reconhecer problemas relacionados à linguagem.

- **Preparação dos Dados:** é responsável por transformar os dados coletados em um formato adequado para os modelos de classificação supervisionada.
- **Modelagem:** com os dados no formato correto, são avaliados diferentes algoritmos de classificação para verificar qual deles, ou qual conjunto de parâmetros, resolve a atividade de classificação de maneira mais satisfatória.
- **Avaliação:** os modelos desenvolvidos são avaliados com base em métricas como acurácia, precisão, recall e F1-score. A partir dos resultados obtidos nessa etapa, é selecionado um modelo principal da solução proposta. Além da avaliação quantitativa, foi realizada uma prova de conceito (PoC) para verificar a aplicabilidade da solução no fluxo de construção do PersonAut dentro da plataforma GuideAut.
- **Implantação:** o modelo selecionado é preparado para integração com a plataforma GuideAut.

A Figura 1.1 apresenta o fluxo metodológico adotado neste trabalho.

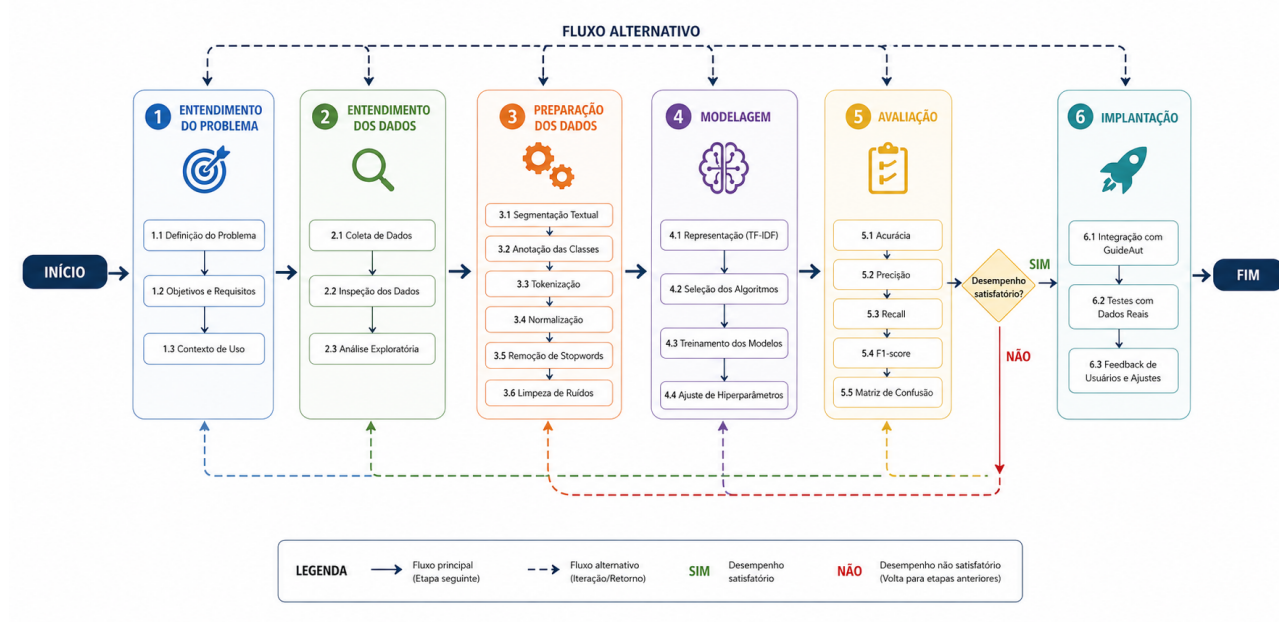


Figura 1.1: Fluxo metodológico adotado no projeto.

Fonte: Elaborado pela autora e produzido com apoio do ChatGPT (2026).

Fundamentação Teórica

Neste capítulo, são apresentados os principais conceitos teóricos que embasam o desenvolvimento deste trabalho. São abordados princípios relacionados a autismo e acessibilidade digital bem como fundamentos do Processamento de Linguagem Natural (PLN).

2.1 Autismo e acessibilidade digital

O Transtorno do Espectro Autista (TEA) se classifica como um transtorno do neurodesenvolvimento (American Psychiatric Association, 2014). Caracterizado por comportamentos estereotipados, déficits no desenvolvimento social, acadêmico e aprendizagem em alguns casos (American Psychiatric Association, 2014). Por ser um espectro, se trata de um grupo heterogêneo, com diversas características, déficits e graus, que vai de leve a grave. Pessoas neurodivergentes tendem a ter mais de um transtorno associado, por exemplo, muitos indivíduos com TEA têm deficiência intelectual. Logo, para o diagnóstico de autismo, é observado tanto fatores de comunicação social quanto comportamentos repetitivos (p. ex., abanar as mãos, estalar os dedos) e rigidez mental (American Psychiatric Association, 2014). A rigidez cognitiva é uma característica notável do autismo, o apego a rotinas, texturas, cores e até formatos está ligado a essa resistência à mudanças. Rituais e planejamento, trazem ao paciente uma fonte de prazer, motivação e conforto (American Psychiatric Association, 2014).

Existem muitos tratamentos possíveis para melhoria da vida das pessoas com TEA. Locatelli explica que por meio de uma equipe médica com profissionais variados e o suporte familiar é possível proporcionar a esses indivíduos uma qualidade maior no dia a dia. Com a crescente evolução do mundo digital, as tecnologias têm se tornado uma grande aliada nesse processo,

servindo de suporte aos pais e terapeutas e trazendo maior independência à pessoa com TEA (CHIEN et al., 2015). Porém, muitas das tecnologias desenvolvidas, são voltadas para os indivíduos autistas mas não pensadas para eles, tornando o acesso desconfortável e, por vezes, restrito. A luta pela inclusão digital tem ganhado grande força, nas últimas décadas, dado a crescente digitalização da sociedade (SOUSA; NETO; PESSOA, 2023). Ladeira e Amaral (LADEIRA; AMARAL, 1999) conceituam inclusão como um processo que percorre e afeta durante toda a vida do indivíduo e tem como objetivo proporcionar melhor qualidade de vida. Visto isso, a necessidade da inclusão também para esse público se torna necessária.

Como mencionado anteriormente, a necessidade de manter padrões específicos é uma necessidade evidente na vida das pessoas com TEA. Na construção de aplicações voltadas a esse público não é diferente, é necessário diretrizes claras e que ofereçam suporte às diferentes necessidades especiais. Desde o design, fluxo do uso até o desenvolvimento deve ser baseado em cores, layout, sons e imagens sugeridas por profissionais. Nesse sentido, o ProAut (MELO, 2021), propõe um apoio aos desenvolvedores que desejam criar aplicações inclusivas. Um dos artefatos do ProAut, o PersonAut, define um modelo de personas que ajuda no processo de empatia para a criação de interfaces exclusivas para pessoas autistas.

Neste trabalho, foi feito um levantamento com pais e responsáveis para ajudar no preenchimento do PersonAut, onde há seções que definem atividades que acalmam, estressam e a relação que os indivíduos autistas têm com a tecnologia. Porém, como esse levantamento foi feito por meio de formulários, há uma despadronização nas respostas o que gera uma dificuldade na consulta e uso dessas informações. Por exemplo, vários pais responderam: “Meu filho gosta de tomar banho” ou “Se acalma em atividades com mar e piscina” que estariam enquadrados na mesma categoria, “água”. A World Wide Web Consortium (W3C) também tem desenvolvido as guidelines, que são recomendações que buscam guiar no desenvolvimento de aplicações acessíveis (World Wide Web Consortium (W3C), 2023). Seguindo essas diretrizes, é possível criar um meio digital inclusivo para pessoas com necessidades especiais.

2.2 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é uma área da Inteligência Artificial (IA) que está crescendo ao longo dos últimos anos. Tem por objetivo extrair padrões e sentido de textos escritos em linguagem humana, usada para comunicações do cotidiano, como o português, espanhol ou francês (INDURKHYA; DAMERAU, 2010). O PLN abrange o estudo de diversos aspectos linguísticos e semânticos, como morfologia, sintaxe, semântica e pragmática, permitindo que sistemas computacionais analisem e processem textos em idiomas naturais, como o português, o inglês, o espanhol e o francês (KHURANA et al., 2022; JURAFSKY; MARTIN, 2024). Essa área também lida com desafios complexos, como ambiguidades, polissemia, sarcasmos e expressões idiomáticas, que tornam a linguagem humana rica, porém difícil de ser interpretada por máquinas.

A maior dificuldade computacional do uso das PLN's está na característica de que a linguagem humana geralmente é escrita de forma incorreta e entendê-las exige a necessidade de conectar conceitos e significados e não apenas uma sequência de caracteres (ALMEIDA; RASSI, 2024). A PLN pode ser inserida em diversas atividades como classificação de textos, e recuperação e extração de informações. Nesse sentido, nota-se a importância e o potencial do uso das NLP's para a padronização de informações em grande volumes de dados.

2.3 Classificação de texto

A classificação de texto é uma das tarefas fundamentais do Processamento de Linguagem Natural (PLN) e consiste em atribuir categorias ou rótulos a documentos de acordo com seu conteúdo. Essa técnica tem como objetivo organizar grandes volumes de informações textuais, facilitando a análise, a busca e a tomada de decisão automatizada em diferentes contextos (AGGARWAL; ZHAI, 2012).

O processo de classificação de texto geralmente envolve duas etapas principais: a representação textual e a aprendizagem do modelo. Na primeira etapa, o texto é convertido em uma

forma numérica que possa ser interpretada por algoritmos computacionais. Essa representação pode ser feita por meio de abordagens tradicionais, como o modelo Bag-of-Words (BoW) e o Term Frequency–Inverse Document Frequency (TF–IDF), ou através de métodos mais recentes, baseados em vetores de palavras (word embeddings), como Word2Vec, GloVe e BERT. Na segunda etapa, os dados representados são utilizados para treinar um modelo de aprendizado de máquina, capaz de identificar padrões linguísticos e semânticos que diferenciam as classes.

Diversos algoritmos podem ser aplicados nesse processo, como Naive Bayes, Máquinas de Vetores de Suporte (Linear SVC), Redes Neurais, e, mais recentemente, modelos baseados em aprendizado profundo (Deep Learning), que alcançam resultados superiores em tarefas complexas de linguagem. A classificação de texto possui uma ampla gama de aplicações práticas, incluindo análise de sentimentos em redes sociais, detecção de spam em e-mails, classificação de notícias por tema, identificação automática de documentos jurídicos e análise de feedbacks de usuários.

Essas aplicações demonstram o potencial da técnica para transformar dados textuais não estruturados em informações úteis e acionáveis, permitindo que organizações extraiam valor de grandes bases de texto. Dessa forma, a classificação de texto representa uma etapa essencial dentro do PLN, unindo técnicas linguísticas e computacionais para compreender e categorizar automaticamente a linguagem humana, contribuindo significativamente para o avanço de sistemas inteligentes de processamento de informações.

2.4 ProAut

O ProAut é um processo desenvolvido para apoiar equipes de software na prototipação de interfaces voltadas para pessoas autistas, especialmente aquelas de baixo funcionamento. Proposto por (MELO, 2021), o processo foi estruturado a partir do Design Thinking, integrando princípios do Design Centrado no Usuário e do Design Participativo, a fim de assegurar que familiares, cuidadores, terapeutas e, se possível, os próprios autistas contribuam com a construção das soluções.

Uma das principais motivações para o ProAut é de que grande parte das aplicações des-

tinadas ao público autista carece de diretrizes claras para o projeto de interface. Além disso, observa-se uma lacuna ainda maior em estudos que considerem autistas de baixo funcionamento, cuja participação direta no processo de design é limitada e exige metodologias e artefatos específicos para representar suas necessidades. O processo ProAut organiza-se em quatro fases principais: Imersão, Análise, Ideação e Prototipação, cada uma composta por atividades e artefatos próprios.

Esses artefatos ajudam equipes pouco experientes no domínio do autismo a compreender o contexto do usuário, mapear características sensoriais e cognitivas relevantes, estruturar requisitos e transformar esse conhecimento em interfaces adequadas. Entre eles destacam-se o Formulário de Caracterização do Autista (FCA), os Canvas de Análise, a Tabela de Requisitos/Restrições (TRR) e os modelos PersonAut e EmpathyAut (MELO et al., 2020a, 2020b), considerados um dos principais diferenciais do processo.

2.4.1 PersonAut

O PersonAut é um dos artefatos centrais da fase de Análise do ProAut e constitui o principal foco deste trabalho. Ele consiste em um modelo de persona especialmente desenvolvido para representar pessoas autistas de baixo funcionamento, suprindo a limitação dos modelos tradicionais de personas, que não capturam adequadamente aspectos sensoriais, comunicacionais e comportamentais dessa população. Segundo (MELO, 2021), o PersonAut foi projetado para registrar informações que influenciam diretamente o design de interfaces, tais como:

- características sensoriais (hiper e hipo-reatividade);
- preferências visuais e de interação;
- gatilhos de estresse e elementos que regulam ou desregulam emocionalmente o usuário;
- comportamentos repetitivos ou padrões motores que afetem o uso da aplicação;
- nível de comunicação e
- necessidades de apoio do cuidador durante a tarefa

O modelo é disponibilizado em dois formatos: um mais descritivo (Figura 2.2) e outro mais sintético (Figura 2.3), permitindo que as equipes escolham o formato mais adequado ao seu processo. Ambos incluem seções específicas para registrar observações obtidas por meio de entrevistas com cuidadores, terapeutas e familiares, formando uma representação consistente do usuário quando a participação direta não é possível.

Ao incorporar elementos comportamentais e sensoriais típicos do TEA, o PersonAut garante que a equipe de design mantenha as necessidades reais do usuário no centro das decisões durante todo o processo. O PersonAut também desempenha papel relevante ao fornecer um referencial de categorias e atributos que inspiraram a estruturação do dataset e das classes utilizadas no sistema de classificação automática de descrições comportamentais.

Além de existir como modelo conceitual, o PersonAut está integrado à plataforma GuideAut, um ambiente desenvolvido para apoiar a aplicação do processo ProAut. Na plataforma, designers e desenvolvedores, inserem informações comportamentais sobre a pessoa autista para construção da persona. Entretanto, as informações coletadas por meio de entrevistas/formulários com pais e responsáveis ocorre, predominantemente, em formato textual livre, o que dificulta sua organização. Nesse contexto, este trabalho busca apoiar a automatização parcial da construção do PersonAut por meio da classificação automática dessas descrições comportamentais, permitindo que informações semelhantes sejam agrupadas.



Figura 2.2: Modelo PersonAut em sua versão descritiva. Adaptado de Melo (2021).

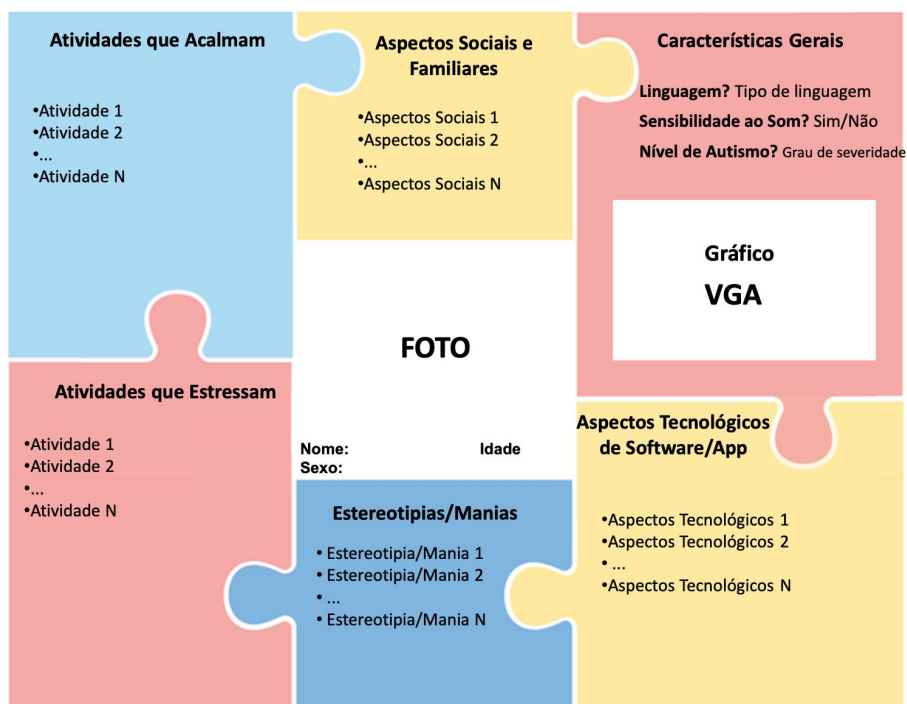


Figura 2.3: Modelo PersonAut em sua versão sintética. Adaptado de Melo (2021).

2.4.2 CRISP-DM

A *Cross Industry Standard Process for Data Mining* (CRISP-DM) é uma das metodologias mais utilizadas para condução de projetos de mineração de dados e ciência de dados (SHIMAOKA;

FERREIRA; GOLDMAN, 2024). Desenvolvido no final da década de 1990, o modelo foi criado com o objetivo de estabelecer um processo padronizado que pudesse ser aplicado em qualquer setor de atuação (CHAPMAN et al., 2000).

Uma das principais características da CRISP-DM é sua abordagem iterativa e flexível. Ela permite que as etapas sejam revisitadas sempre que necessário, possibilitando ajustes contínuos conforme o domínio sobre o setor ou sobre os dados aumentem (CHAPMAN et al., 2000). Essa característica torna o modelo adequado para projetos que envolvem dados reais, que necessitem de refinamento durante o processo.

A CRISP-DM é composta por seis fases principais: Entendimento do Negócio (*Business Understanding*), Entendimento dos Dados (*Data Understanding*), Preparação dos Dados (*Data Preparation*), Modelagem (*Modeling*), Avaliação (*Evaluation*) e Implantação (*Deployment*). Essas etapas abrangem todo o ciclo de vida de um projeto de ciência de dados, desde a compreensão do problema até a disponibilização da solução.

Devido à sua ampla aceitação na literatura e na indústria, o CRISP-DM é frequentemente adotado em pesquisas que envolvem mineração de dados, aprendizado de máquina e Processamento de Linguagem Natural, fornecendo uma estrutura organizada para o desenvolvimento, validação e implementação de soluções baseadas em dados (ALVES, 2023).

2.5 Trabalhos relacionados

Nesta seção, foram revisadas pesquisas relacionadas à inclusão de pessoas com TEA na tecnologia, destacando metodologias, limitações e como este trabalho se diferencia. Em (SOUZA; BENITEZ; CARMO, 2021), apresenta o interesse que indivíduos com TEA possuem pela tecnologia e a ascensão da criação de aplicações voltadas a esse público. Nesse artigo, é levantada a necessidade da criação de diretrizes de acessibilidade para que essas interfaces sejam acessíveis.

Nesse sentido, é proposta uma revisão integrativa (RI) dos estudos nessa área para a consolidação de pesquisas empíricas envolvendo pessoas com TEA. Foram revisados sete artigos para essa revisão, onde foi observado que ainda existe uma lacuna na análise de diretrizes de acessibilidade no mundo digital. Em especial a autistas adultos de nível leve, que têm poucos

estudos empíricos direcionados a eles, o que de acordo com (SOUZA; BENITEZ; CARMO, 2021) dificulta generalizações. A revisão dos artigos baseou-se nas convenções da Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

Dos resultados, foram encontrados 365 publicações acerca do tema, em três bancos (Scopus, Web of Science e Google Scholar), dessas 365 foram selecionadas 7 para compor a RI com contribuições do Reino Unido (3), Itália (1), Estados Unidos (1), Austrália (1) e Brasil (1), estudos envolvendo adultos e crianças com TEA e cuidadores/terapeutas. Nota-se a relevância dessa RI no meio acadêmico para estudos futuros acerca da consolidação das diretrizes de interfaces inclusivas. Diferente dessa abordagem, o presente artigo está voltado para a aplicação de PLN para a criação automática do PersonAut, artefato do ProAut, direto na plataforma GuideAut, criando assim uma base firme e consistente de um guia para a criação de aplicações voltadas ao público com TEA.

O trabalho apresentado em (PASSERINO; MONTARDO, 2007) propõe analisar os conceitos de inclusão social e digital entre si. Segundo Passerino e Montardo (2007), “Considera-se inclusão, portanto, o processo estabelecido dentro de uma sociedade mais ampla que busca satisfazer necessidades relacionadas com qualidade de vida, desenvolvimento humano, autonomia de renda e equidade de oportunidades e direitos para os indivíduos e grupos sociais que em alguma etapa da sua vida encontram-se em situação de desvantagem com relação a outros membros da sociedade.” relacionando essa definição com a desigualdade digital, que muitas vezes é associado apenas a falta de aparelhos ou falta de acesso a rede mas no artigo é evidenciado que é também a impossibilidade de uso, capacidade técnica e identificação com as aplicações desenvolvidas no mercado. Entendendo-se que a verdadeira inclusão digital está em combinar hardware e software. Com esse artigo foi possível relacionar a necessidade de inclusão no meio social e digital com a obrigatoriedade de existir normas e convenções padronizadas a cerca da criação de interfaces inclusivas.

O trabalho apresentado em (BORGES, 2023) faz uma análise comparativa entre algoritmos clássicos e modelos como BERT. O estudo parte do contexto da mineração de texto, enfatizando a importância da classificação automática para organização de grandes volumes de dados

textuais em categorias previamente definidas.

Para a realização dos experimentos, foram utilizadas quatro bases de dados públicas, sendo duas em português e duas em inglês, com cenários de classificação binária e multiclasse. Os resultados demonstraram que o BERT obteve desempenho superior aos métodos tradicionais em métricas como acurácia e F1-score, evidenciando a capacidade dos modelos contextuais de compreender melhor o significado dos textos. Entretanto, o estudo também destaca que os algoritmos clássicos ainda são relevantes em cenários com menor volume de dados, devido ao menor custo de treinamento e processamento.

Embora o trabalho tenha como foco a comparação entre diferentes técnicas de classificação textual, ele não aborda um domínio específico de aplicação. Diferentemente dessa proposta, o presente trabalho aplica técnicas de PLN ao contexto do TEA, utilizando descrições comportamentais fornecidas por familiares e responsáveis para apoiar a construção automatizada do PersonAut na plataforma GuideAut.

O trabalho apresentado em (VANDENBUSSCHE et al., 2022) investiga a aplicação de técnicas de PLN e Aprendizado de Máquina na análise automática de relatos textuais produzidos por pacientes com diferentes tipos de cefaleia. O objetivo do estudo foi verificar se descrições escritas pelos próprios pacientes poderiam ser utilizadas para identificar automaticamente o tipo de condição associada. Para a realização da pesquisa, foram coletados relatos escritos por 121 pacientes diagnosticados com enxaqueca ou cefaleia em salvas. Os textos foram submetidos a diferentes etapas de processamento linguístico. Posteriormente, algoritmos de aprendizado de máquina, como Regressão Logística, Máquinas de Vetores de Suporte (SVM) e Naive Bayes, foram usados para classificar automaticamente as descrições de acordo com o diagnóstico correspondente.

Os resultados demonstraram que as técnicas de PLN foram capazes de identificar diferenças linguísticas significativas entre os grupos analisados, revelando padrões compatíveis sobre essas condições médicas. Além disso, os modelos baseados em Regressão Logística e SVM apresentaram os melhores desempenhos de classificação, alcançando valores de F1-score entre 0,82 e 0,86, superando os classificadores baseados em Naive Bayes. Embora o estudo esteja inserido no con-

texto clínico e voltado à identificação de diferentes tipos de cefaleia, sua abordagem apresenta semelhanças com a proposta deste trabalho pois também utiliza descrições textuais produzidas por usuários como fonte para a classificação. Entretanto, diferentemente da pesquisa de Vandenbussche et al. (2022), que busca apoiar o diagnóstico de condições médicas, o presente trabalho aplica técnicas de PLN para categorizar descrições comportamentais relacionadas ao TEA.

O Quadro 2.1 apresenta uma síntese comparativa entre os trabalhos analisados e a proposta deste trabalho, destacando seus focos, limitações e contribuições.

Estudo	Foco e Tecnologia	Lacuna Identificada	Como este trabalho aborda
Souza, Benitez e Carmo (2021)	Revisão integrativa sobre acessibilidade digital para pessoas com TEA e diretrizes para interfaces inclusivas.	Identifica escassez de estudos empíricos e ausência de diretrizes consolidadas para aplicações voltadas ao público autista.	Utiliza o PersonAut e o GuideAut como instrumentos para apoiar o desenvolvimento de soluções mais alinhadas às necessidades de pessoas com TEA.
Passerino e Montardo (2007)	Discussão sobre inclusão digital e social, enfatizando acessibilidade e participação tecnológica.	Não propõe mecanismos computacionais para apoiar o desenvolvimento de sistemas inclusivos.	Aplica técnicas computacionais para automatizar a organização de informações comportamentais utilizadas na construção de personas inclusivas.
Borges et al. (2023)	Comparação entre algoritmos clássicos de classificação textual e modelos baseados em BERT.	Não explora aplicações em domínios específicos relacionados ao TEA ou acessibilidade.	Aplica e compara modelos clássicos e embeddings contextuais em um domínio específico relacionado ao TEA.
Vandenbussche et al. (2022)	Aplicação de PLN e Aprendizado de Máquina para classificação automática de relatos clínicos.	Foco voltado ao apoio diagnóstico em contexto médico, sem relação com acessibilidade ou construção de personas.	Utiliza PLN para classificação de descrições comportamentais com finalidade de apoio à construção automatizada do PersonAut.

Quadro 2.1: Comparação entre os trabalhos relacionados e a proposta deste trabalho

Desenvolvimento

Este capítulo apresenta o processo de desenvolvimento do sistema de classificação baseado em PLN para viabilizar a classificação automática de descrições comportamentais relacionadas ao PersonAut. Conforme discutido nos capítulos anteriores, o PersonAut é um dos principais artefatos do processo ProAut, sendo utilizado para representar características comportamentais relevantes de pessoas autistas e auxiliar designers e desenvolvedores na criação de interfaces mais acessíveis.

Atualmente, a construção desse artefato depende da análise manual de descrições textuais fornecidas por familiares e responsáveis. Esse processo exige esforço significativo para interpretar, organizar e categorizar as informações coletadas, especialmente quando há grande volume de dados. Diante desse cenário, este trabalho propõe a utilização de técnicas de PLN para classificar automaticamente essas descrições comportamentais nas categorias que compõem o PersonAut, permitindo apoiar sua geração dentro da plataforma GuideAut.

Para isso, foi desenvolvido um pipeline completo envolvendo coleta e preparação dos dados, construção do dataset anotado, modelagem supervisionada, avaliação dos modelos e integração da solução à plataforma. O desenvolvimento foi estruturado seguindo as etapas da metodologia CRISP-DM apresentadas na seção 1.3 do capítulo 1.

3.1 Entendimento do Problema

A etapa de Entendimento do Problema tem como objetivo compreender o contexto da pesquisa, identificar as necessidades da solução proposta e definir os objetivos que devem ser alcançados (CHAPMAN et al., 2000). Nessa etapa, os requisitos do sistema foram definidos de forma

alinhada ao objetivo geral do projeto: viabilizar a classificação automática de descrições comportamentais fornecidas por familiares de pessoas autistas, estruturando-as nas dimensões que compõem o PersonAut. Esses requisitos orientam o desenvolvimento e garantem que o protótipo seja consistente com as necessidades do processo ProAut e da plataforma GuideAut.

Os requisitos do sistema foram definidos a partir da análise do processo atual de construção do PersonAut, dos objetivos estabelecidos para esta pesquisa e das necessidades de integração com a plataforma GuideAut. Para sua elaboração, foram consideradas as atividades realizadas manualmente por designers e desenvolvedores durante a análise das descrições comportamentais, e as etapas necessárias para transformar essas informações em categorias estruturadas. A partir desse levantamento, foram identificadas as funcionalidades essenciais para viabilizar a classificação automática dos textos e os requisitos de qualidade necessários para garantir a confiabilidade da solução proposta.

A seguir, são apresentados os requisitos funcionais definidos para o sistema.

ID	Descrição
RF01	O sistema deve receber textos descritivos inseridos por familiares de pessoas autistas.
RF02	O sistema deve pré-processar automaticamente os textos utilizando técnicas de PLN.
RF03	O modelo deve classificar os textos nas categorias do PersonAut.
RF04	O sistema deve gerar automaticamente classificações estruturadas para apoiar a construção do PersonAut.
RF05	O pipeline deve permitir treinamento e avaliação de diferentes modelos supervisionados.

Quadro 3.2: Requisitos Funcionais do Sistema

Também foram definidos requisitos não funcionais relacionados ao desempenho, modularidade e capacidade de integração da solução.

ID	Descrição
RNF01	O sistema deve apresentar desempenho mensurável utilizando métricas como precisão, recall e F1-score.
RNF02	O pipeline deve lidar com textos escritos de maneira informal e heterogênea.
RNF03	O modelo deve possuir arquitetura modular para futura integração com a plataforma GuideAut.
RNF04	O processo de treinamento deve evitar vazamento de informações entre treino e teste.

Quadro 3.3: Requisitos Não Funcionais do Sistema

3.2 Entendimento e Preparação dos Dados

A etapa de entendimento dos dados consiste na compreensão da estrutura textual das respostas coletadas e identificar padrões linguísticos relevantes para a classificação supervisionada (CHAPMAN et al., 2000).

Como as respostas do formulário foram produzidas de forma espontânea pelos responsáveis, observou-se elevada heterogeneidade textual, incluindo erros ortográficos, abreviações, linguagem coloquial, emojis, repetições e frases extensas. O quadro 3.4 apresenta exemplos observados nas respostas coletadas.

Característica	Exemplo
Erro ortográfico	“ele fica muito agitado quando esculta baruios alto”
Abreviação	“qnd vai ao mercado fica nervoso”
Linguagem coloquial	“ele ama mexer no celular o tempo todo”
Repetição de caracteres	“fica muuuito irritado com buzina”
Frases extensas	“quando sai da rotina da escola e vai para lugares muito cheios ele começa a ficar nervoso, tapa os ouvidos e tenta se esconder”
Múltiplas informações semânticas	“gosta de brincar sozinho mas se irrita quando tiram o tablet dele”

Quadro 3.4: Exemplos de heterogeneidade textual presentes nas respostas

Diante disso, foi necessário realizar etapas de entendimento e preparação dos dados para garantir sua adequação ao processo de classificação supervisionada. As subseções seguintes apresentam o processo de construção do dataset, abrangendo a coleta das respostas, definição

das categorias, anotação manual das descrições e organização dos dados para treinamento e avaliação dos modelos.

3.2.1 Construção do Dataset

Embora existam datasets públicos relacionados ao TEA, a maior parte dessas bases estão voltadas para triagem diagnóstica, sinais clínicos ou exames. Observou-se uma escassez de bases textuais relacionadas ao comportamento da pessoa autista.

Dessa forma, tornou-se necessária a construção de um dataset próprio alinhado às categorias do PersonAut. A criação dessa base permitiu representar aspectos cotidianos, subjetivos e contextuais do comportamento de pessoas autistas a partir da perspectiva de seus responsáveis.

O processo de construção do dataset envolveu:

- coleta de respostas discursivas;
- segmentação textual em unidades menores de sentido;
- definição das categorias semânticas;
- anotação manual supervisionada.

3.2.2 Coleta de Dados

A coleta foi realizada por meio de um formulário discursivo aplicado a pais, mães e responsáveis de pessoas autistas. O formulário foi estruturado com base nas dimensões do PersonAut, contemplando aspectos comportamentais, sensoriais, sociais e tecnológicos.

O instrumento foi dividido em sete seções:

- Informações gerais;
- Atividades que acalmam;
- Atividades que estressam;
- Aspectos sociais e familiares;

- Estereotípias e manias;
- Aspectos tecnológicos;
- Características gerais.

As perguntas foram construídas de forma discursiva, incentivando respostas naturais e detalhadas. O formulário permaneceu disponível entre os dias 26/11/2025 e 02/12/2025. Ao final da coleta, foram obtidas 21 respostas completas. A Tabela 3.5 apresenta exemplos de respostas coletadas.

Texto
“Gosta de brincar na piscina, isso sempre acalma ele.”
“Se irrita facilmente com barulhos muito altos.”
“Prefere brincar sozinho.”
“Passa a mão repetidamente em tecidos macios.”
“Assiste vídeos de trens no tablet por longos períodos.”

Quadro 3.5: Exemplo de respostas coletadas

3.2.3 Definição das Categorias e Anotação

A construção do modelo supervisionado exigiu a definição clara das categorias e a anotação manual das descrições coletadas. As categorias utilizadas foram derivadas diretamente das dimensões presentes no PersonAut, garantindo alinhamento entre o modelo computacional e a lógica do ProAut.

As seguintes categorias foram adotadas:

- Atividades que Acalmam;
- Atividades que Estressam;
- Aspectos Sociais e Familiares;
- Estereotípias/Manias;
- Aspectos Tecnológicos.

Após a segmentação textual, as frases foram anotadas manualmente considerando o contexto das descrições e as dimensões do PersonAut. O processo de anotação foi conduzido por meio da leitura individual de cada segmento textual, atribuindo-se uma categoria a cada entrada. Como apoio ao processo de anotação, foi elaborado um conjunto de critérios para orientar a classificação das frases em cada uma das categorias do PersonAut. Essa etapa permitiu transformar respostas longas e heterogêneas em 456 exemplos supervisionados para treinamento dos modelos de PLN. O Quadro 3.6 apresenta uma síntese dos critérios utilizados durante a rotulagem dos dados.

Categoria	Critério de Anotação	Exemplo
Atividades que Acalmam	Frases relacionadas a situações, atividades ou estímulos que promovem relaxamento, conforto, bem-estar ou autorregulação emocional.	“O Lucas se acalma muito quando escuta músicas suaves”
Atividades que Estressam	Frases associadas a desconforto, irritação, ansiedade, sobrecarga sensorial ou situações que desencadeiam estresse.	“Ele se estressa bastante com barulhos repentinos”
Aspectos Sociais e Familiares	Descrições relacionadas à interação social, comunicação, convivência familiar, amizades, preferências de companhia ou isolamento social.	“Com a família ele é carinhoso”
Estereotipias/Manias	Comportamentos repetitivos, movimentos estereotipados, rotinas rígidas, interesses restritos ou hábitos recorrentes.	“Ele balança o corpo para frente e para trás quando está empolgado ou nervoso”
Aspectos Tecnológicos	Referências ao uso de dispositivos eletrônicos, aplicativos, jogos digitais, vídeos, computadores, tablets ou outras tecnologias.	“Uso do celular/tablet”

Quadro 3.6: Critérios utilizados para anotação das categorias

3.2.4 Distribuição do Dataset

Após as etapas de coleta, segmentação e anotação manual, o dataset final passou a possuir 456 exemplos textuais anotados. Como algumas frases pertenciam simultaneamente a mais de uma categoria, foi aplicada expansão multilabel durante a preparação dos dados.

A Figura 3.4 apresenta a distribuição das categorias no conjunto final.

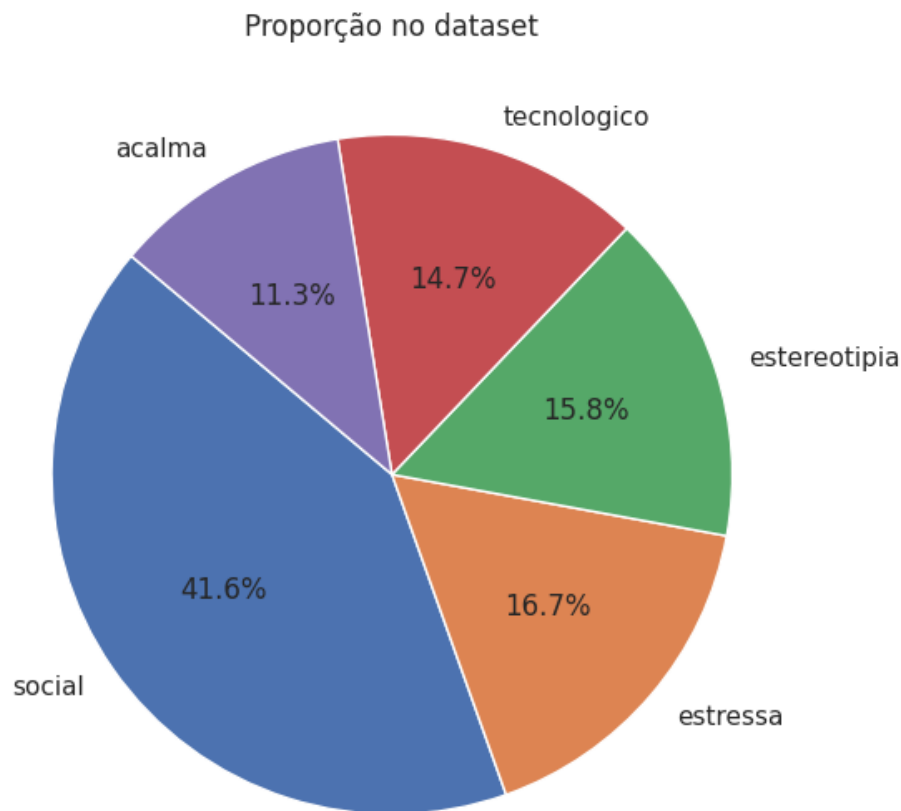


Figura 3.4: Distribuição das categorias no dataset

Houve predominância da categoria *social*, responsável por aproximadamente 41,6% dos exemplos do dataset. Esse desbalanceamento influenciou diretamente os experimentos, especialmente nas métricas de recall das classes minoritárias.

As categorias *acalma* e *estereotipia*, por exemplo, possuem quantidade significativamente menor de exemplos quando comparadas à classe *social*, dificultando o aprendizado supervisionado.

3.2.5 Divisão entre Treino e Teste

O conjunto de dados foi dividido utilizando *StratifiedShuffleSplit*, preservando proporcionalmente a distribuição das categorias nos conjuntos de treino e teste. A Figura 3.5(a) apresenta a distribuição das categorias no conjunto de teste, enquanto a Figura 3.5(b) apresenta a distri-

buição no conjunto de treino.

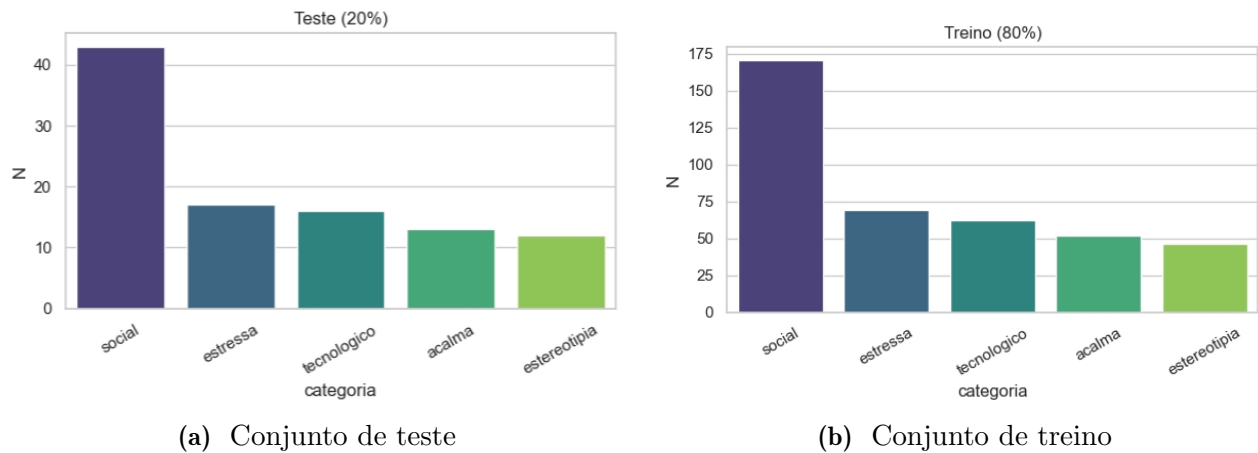


Figura 3.5: Distribuição das categorias nos conjuntos de treino e teste

A distribuição proporcional entre treino e teste foi importante para evitar distorções nas métricas finais, principalmente devido ao tamanho reduzido do dataset.

O conjunto final foi dividido em:

- 400 exemplos para treino;
- 101 exemplos para teste.

3.3 Implementação do Pipeline de PLN

Após a preparação dos dados, foi desenvolvido o pipeline de PLN responsável por realizar a transformação dos textos em representações adequadas para os modelos de classificação. A implementação do pipeline de PLN constituiu a etapa central do desenvolvimento da solução que foi construído de forma modular, permitindo testar diferentes representações textuais e classificadores supervisionados.

3.3.1 Normalização e Limpeza

Nesta etapa foram aplicadas técnicas de limpeza textual com o objetivo de remover ruídos e preservar apenas informações relevantes semanticamente.

As principais transformações incluíram:

- conversão para letras minúsculas;
- remoção de acentos;
- remoção de caracteres especiais;
- remoção de URLs e emojis;
- remoção de stopwords.

A segmentação manual das respostas também foi importante para preservar unidades menores de significado.

A Figura 3.6 apresenta um exemplo do processo de normalização textual aplicado no pipeline.

Texto Original:

“Ele fica MUUUITO irritado com barulhos altos !!!”

Após Normalização e Limpeza:

“fica irritado barulho alto”

Figura 3.6: Exemplo de normalização e limpeza textual

Observa-se que o processo removeu elementos considerados ruído e palavras pouco relevantes semanticamente, preservando os termos mais importantes para a etapa de classificação textual. Destaca-se que essas etapas de pré-processamento foram aplicadas apenas após a divisão treino/teste, evitando *data leakage*.

3.3.2 Vetorização e Representação Textual

Após o pré-processamento, os textos passaram a ser representados numericamente. Inicialmente, foi utilizada a abordagem TF-IDF (*Term Frequency – Inverse Document Frequency*), devido ao seu baixo custo computacional e alta interpretabilidade.

A configuração aplicada utilizou:

- *max_features* = 2000;
- *ngram_range* = (1,2);
- *min_df* = 2.

Além da representação tradicional por TF-IDF, também foram avaliadas representações baseadas em embeddings contextuais utilizando modelos pré-treinados.

Entre os modelos testados destacam-se:

- MiniLM;
- BERTimbau.

Os embeddings gerados por esses modelos permitiram representar semanticamente as frases em espaços vetoriais densos, capturando relações contextuais mais complexas entre palavras e expressões. Por fim, foi realizado o fine-tuning do modelo BERTimbau, ajustando seus parâmetros para a tarefa específica de classificação das descrições comportamentais.

3.4 Modelagem

A etapa de Modelagem da metodologia CRISP-DM consistiu em identificar o modelo supervisionado com melhor capacidade de generalização para a classificação das descrições comportamentais.

3.4.1 Seleção do Modelo Base

Inicialmente foi utilizado o *DummyClassifier* como baseline, configurado com a estratégia *most_frequent*. O objetivo foi estabelecer um desempenho mínimo de referência.

Em seguida, foram avaliados modelos tradicionais de classificação textual, incluindo:

- Multinomial Naive Bayes;
- LinearSVC.

Posteriormente, também foram conduzidos experimentos utilizando embeddings contextuais combinados com classificadores supervisionados. Além da utilização de embeddings contextuais como representação textual para o classificador LinearSVC, também foi avaliada uma abordagem baseada no ajuste fino (fine-tuning) do modelo BERTimbau. Nessa estratégia, o próprio modelo Transformer foi adaptado para a tarefa de classificação das descrições comportamentais do PersonAut, permitindo a atualização de seus parâmetros durante o treinamento supervisionado.

Foram avaliadas combinações entre:

- TF-IDF + LinearSVC;
- MiniLM + LinearSVC;
- BERTimbau + LinearSVC;
- BERTimbau Fine-tuning.

Os modelos baseados em embeddings contextuais apresentaram melhor capacidade de capturar nuances semânticas presentes nas descrições, especialmente em frases curtas e linguisticamente ambíguas. O LinearSVC demonstrou maior estabilidade entre os modelos clássicos avaliados, servindo como principal classificador nos experimentos baseados em TF-IDF e embeddings. Entretanto, o modelo que apresentou o melhor desempenho geral foi o BERTimbau com fine-tuning, adotado como modelo final da solução.

3.4.2 Resultados da Validação Cruzada

Inicialmente, foram avaliados modelos clássicos de classificação textual utilizando representações baseadas em TF-IDF. Os modelos avaliados foram:

- DummyClassifier;
- Multinomial Naive Bayes;
- LinearSVC.

Os experimentos foram conduzidos utilizando validação cruzada estratificada repetida (*RepeatedStratifiedKfold*) com 5 folds e 2 repetições.

A Tabela 3.1 apresenta os resultados médios obtidos.

Tabela 3.1: Resultados médios da validação cruzada

Modelo	Accuracy média	F1-macro médio
TF-IDF + SVM	0.5363 ± 0.0452	0.4426 ± 0.0601
TF-IDF + Naive Bayes	0.5363 ± 0.0287	0.3756 ± 0.0570
DummyClassifier	0.4275 ± 0.0050	0.1198 ± 0.0010

Os resultados demonstram que o modelo LinearSVC apresentou desempenho superior e mais estável quando comparado aos demais modelos clássicos.

Embora o Naive Bayes tenha apresentado accuracy semelhante ao SVM, o F1-score macro inferior indica maior dificuldade em classificar corretamente as categorias minoritárias.

Já o DummyClassifier apresentou desempenho significativamente inferior, confirmando que a tarefa exige aprendizado supervisionado real e que simplesmente prever a classe majoritária não é suficiente para resolver o problema proposto.

3.4.3 Avaliação dos Modelos Clássicos

Após a validação cruzada, o modelo TF-IDF + LinearSVC foi avaliado no conjunto holdout composto por 20% dos dados.

A Figura 3.7 apresenta a distribuição bootstrap do F1-score macro obtido pelo modelo.

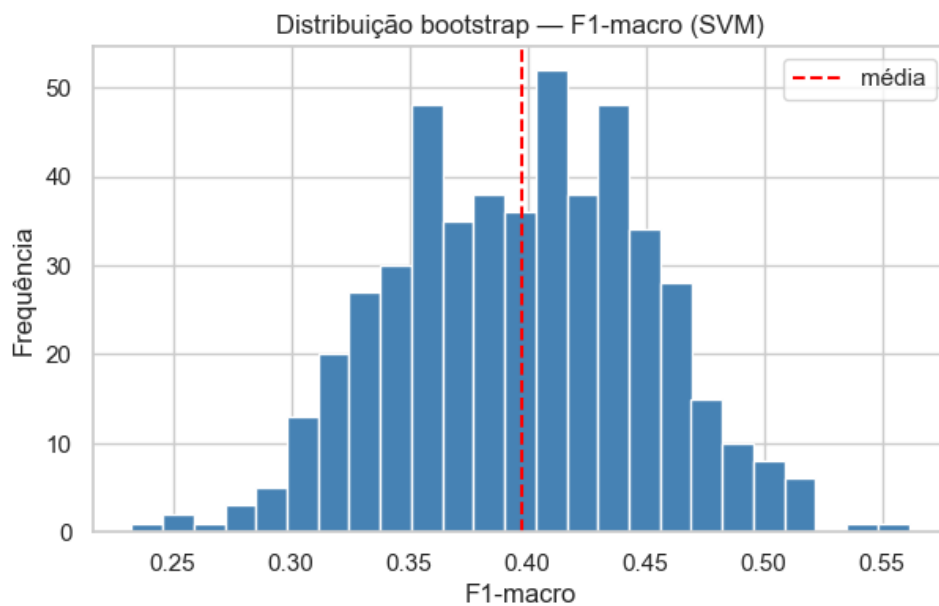


Figura 3.7: Distribuição bootstrap do F1-score macro do modelo TF-IDF + SVM

Observa-se relativa estabilidade do modelo em diferentes reamostragens do conjunto de teste, embora ainda exista variabilidade decorrente do tamanho reduzido do dataset.

Os resultados obtidos pelo modelo TF-IDF + LinearSVC no conjunto holdout foram:

- **Accuracy:** 0.5149, indicando que aproximadamente 51,5% das descrições foram classificadas corretamente.
- **F1-score macro:** 0.4058, refletindo o desempenho médio entre as categorias e considerando simultaneamente precisão e recall.

Apesar do desempenho relativamente satisfatório para um dataset reduzido, observou-se que o modelo apresentou maior dificuldade em classes semanticamente mais ambíguas.

A Figura 3.8 apresenta o F1-score por categoria.

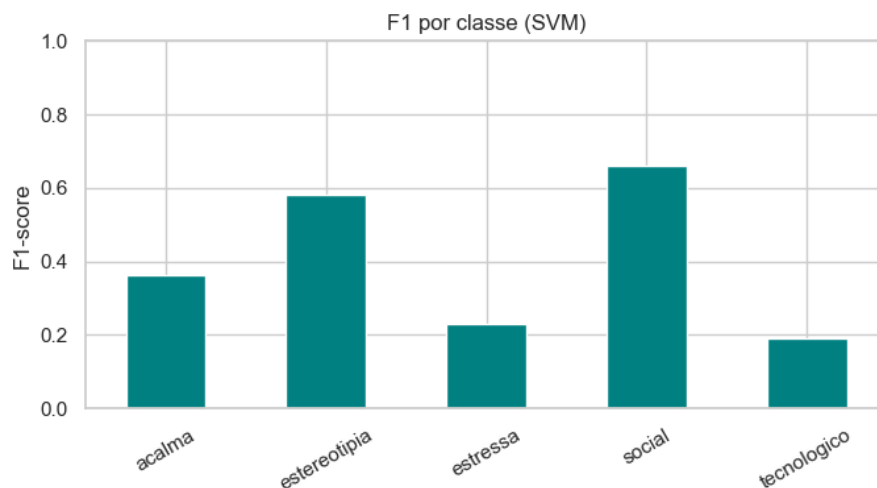


Figura 3.8: F1-score por categoria utilizando TF-IDF + SVM

A categoria *social* apresentou melhor desempenho, principalmente devido à maior quantidade de exemplos disponíveis durante o treinamento.

Por outro lado, as categorias *estressa* e *tecnologico* apresentaram os menores resultados, indicando maior dificuldade do modelo em distinguir frases semanticamente próximas.

A Figura 3.9 apresenta a relação entre precisão e recall para cada categoria.

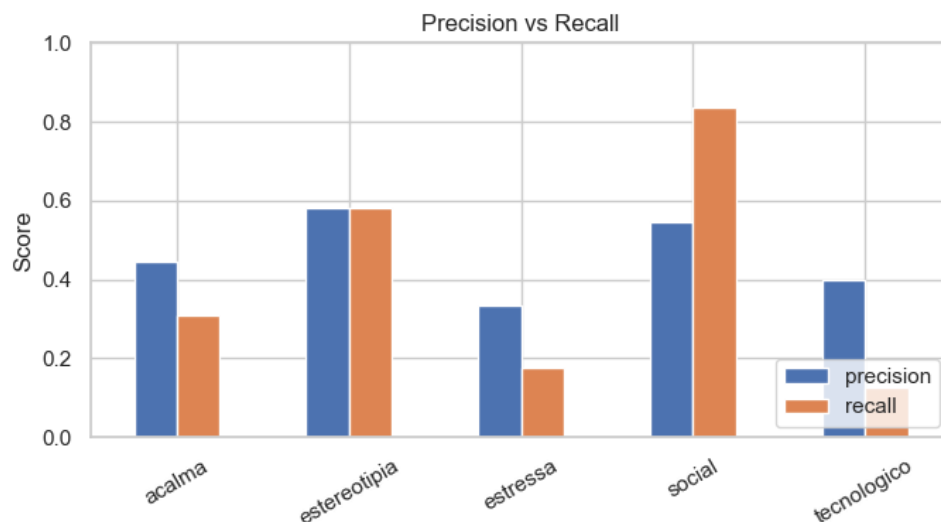


Figura 3.9: Precisão e recall por categoria

Observa-se que algumas categorias apresentam precisão razoável, mas recall reduzido, indicando que o modelo consegue identificar corretamente determinados exemplos, porém deixa de reconhecer diversos casos positivos.

3.4.4 Análise da Matriz de Confusão

A Figura 3.10 apresenta a matriz de confusão do modelo TF-IDF + SVM no conjunto holdout. Os valores da diagonal principal representam as classificações corretas realizadas pelo modelo, e os demais valores indicam confusões entre categorias. Observa-se que a categoria *social* apresentou o maior número de classificações corretas, com 36 instâncias corretamente identificadas. Em contrapartida, as categorias *acalma*, *estressa* e *tecnologico* apresentaram maior quantidade de erros de classificação.

Um padrão observado é a tendência do modelo em predizer a categoria *social*. Por exemplo, 6 instâncias da categoria *acalma*, 10 da categoria *estressa* e 11 da categoria *tecnologico* foram classificadas incorretamente como *social*. Esse comportamento indica forte sobreposição semântica entre as categorias, principalmente porque diversas frases descrevem simultaneamente comportamentos sociais e aspectos emocionais ou tecnológicos. Além do desbalanceamento entre as classes, as confusões observadas também podem ser explicadas pela proximidade semântica entre algumas categorias do PersonAut. Em muitos casos, uma mesma descrição comportamental pode conter elementos relacionados a mais de uma dimensão simultaneamente.

Por exemplo, uma frase como “gosta de assistir vídeos no tablet para se acalmar” apresenta características associadas tanto à categoria *tecnologico* quanto à categoria *acalma*. De forma semelhante, descrições como “prefere ficar sozinho utilizando o celular” podem ser interpretadas como aspectos relacionados à interação social ou ao uso de tecnologia. Essa sobreposição dificulta a definição de fronteiras rígidas entre as categorias e contribui para o aumento das classificações incorretas. Outro aspecto relevante é que as categorias *social* e *estressa* frequentemente compartilham termos semelhantes. Frases contendo palavras como “pessoas”, “amigos”, “conversar” ou “família” podem aparecer tanto em descrições de interação social quanto em relatos de situações que geram desconforto ou ansiedade. Como consequência, o modelo tende a associar esses exemplos à categoria *social*, que também corresponde à classe mais representada no conjunto de treinamento.

Esse comportamento pode ser observado nas elevadas taxas de confusão entre as categorias *estressa*, *tecnologico* e *social*, sugerindo que parte dos erros decorrem não só das limitações

do modelo, mas também da própria subjetividade presente nas descrições comportamentais utilizadas como entrada.

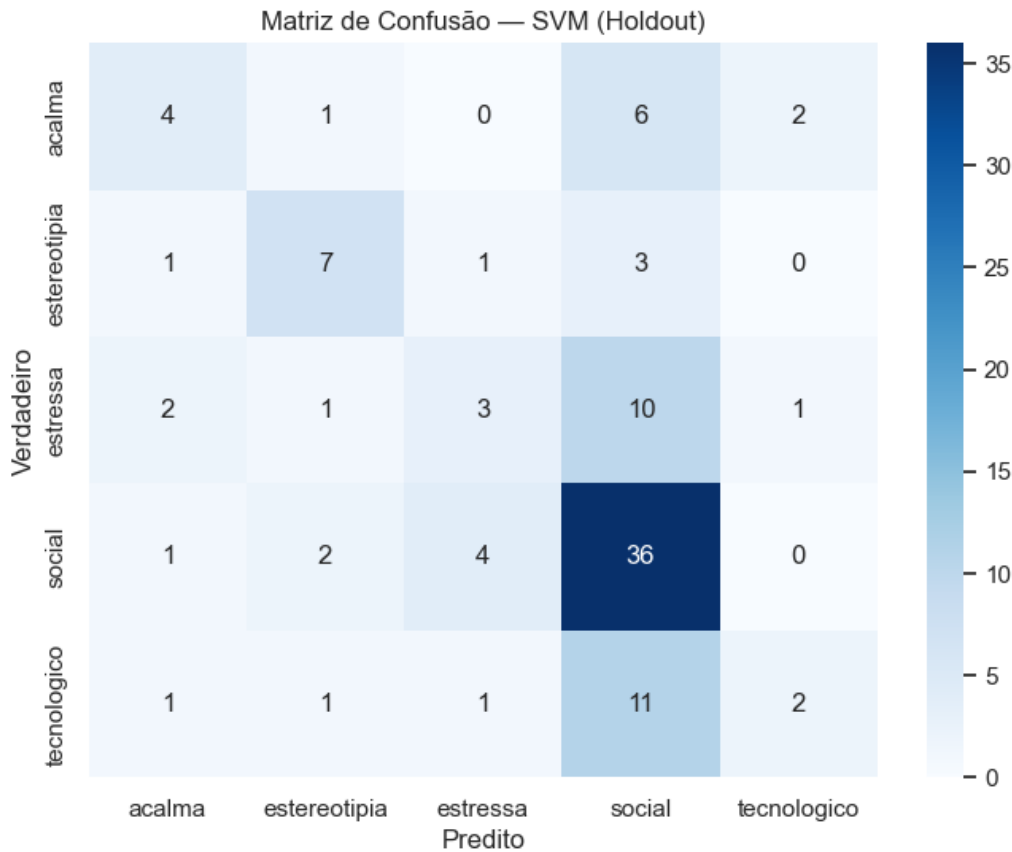


Figura 3.10: Matriz de confusão do modelo TF-IDF + SVM

A Figura 3.11 apresenta a versão normalizada da matriz de confusão, permitindo analisar proporcionalmente o desempenho em cada categoria independentemente da quantidade de exemplos disponíveis. Observou-se que aproximadamente 84% (0.84) das instâncias da categoria *social* foram corretamente classificadas, evidenciando que essa foi a classe com melhor desempenho. A categoria *estereotipia* também apresentou resultado relativamente satisfatório, com cerca de 58%(0.58) de acerto.

Por outro lado, apenas 31% (0.31) das instâncias da categoria *acalma*, 18% (0.18) da categoria *estressa* e 12%(0.12) da categoria *tecnologico* foram corretamente identificadas. Em especial, nota-se que 69%(0.69) dos exemplos da categoria *tecnologico* e 59%(0.59) da categoria *estressa* foram classificados como *social*, reforçando a dificuldade do modelo em separar semanticamente essas classes.

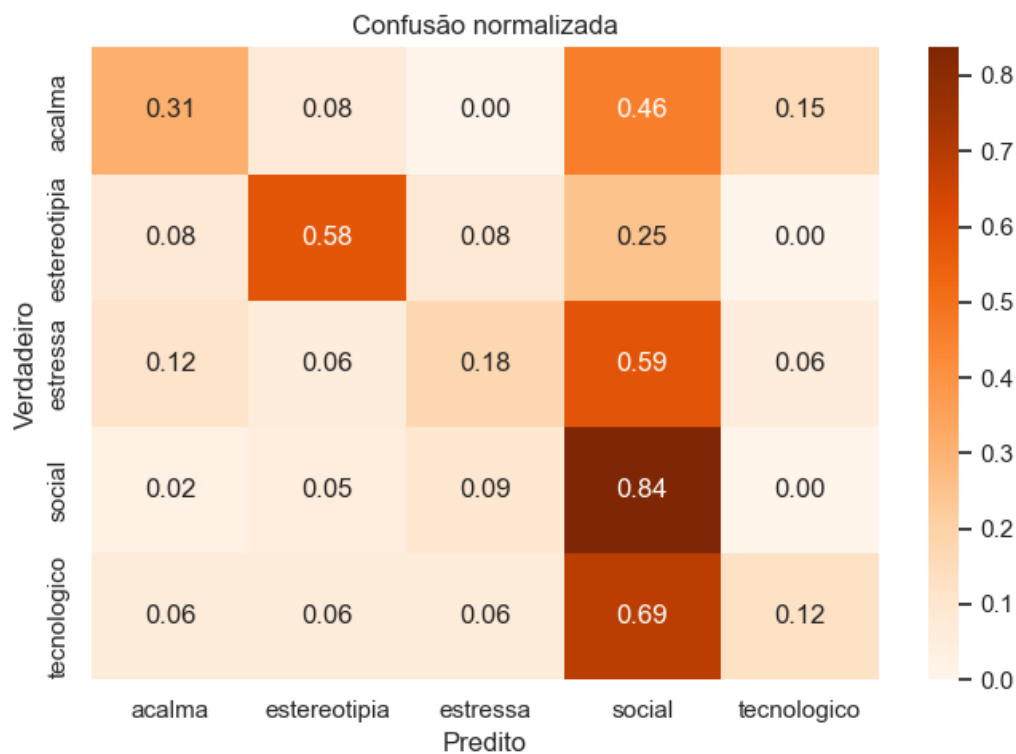


Figura 3.11: Matriz de confusão normalizada

3.4.5 Resultados com Embeddings Contextuais

Além das abordagens tradicionais baseadas em TF-IDF, também foram avaliados modelos utilizando embeddings contextuais.

A hipótese considerada era que representações semânticas profundas poderiam capturar melhor nuances linguísticas presentes nas descrições comportamentais.

MiniLM + SVM

O modelo MiniLM combinado com LinearSVC apresentou os seguintes resultados:

- Accuracy: 0.5842;
- F1-score macro: 0.5047.

Observou-se melhora significativa em relação ao modelo baseado apenas em TF-IDF, principalmente nas categorias *acalma*, *estereotipia* e *social*.

Esse comportamento demonstra que embeddings contextuais conseguem representar melhor relações semânticas presentes nas frases.

BERTimbau + SVM

O modelo BERTimbau + LinearSVC apresentou os seguintes resultados:

- Accuracy: 0.6040;
- F1-score macro: 0.5064.

BERTimbau com Fine-tuning

O modelo BERTimbau com fine-tuning apresentou o melhor desempenho geral entre todos os modelos avaliados.

Os resultados obtidos foram:

- Accuracy: 0.6634;
- F1-score macro: 0.6340.

Os resultados demonstram que o ajuste fino do modelo proporcionou um ganho significativo em relação às abordagens baseadas apenas na utilização de embeddings contextuais combinados com o LinearSVC. Esse desempenho indica que a adaptação dos parâmetros do BERTimbau ao domínio específico das descrições comportamentais permitiu representar de forma mais eficaz as características semânticas presentes no conjunto de dados.

A Figura 3.12(a) apresenta os valores de F1-score macro obtidos pelos modelos avaliados, enquanto a Figura 3.12(b) apresenta os resultados de accuracy no conjunto holdout.

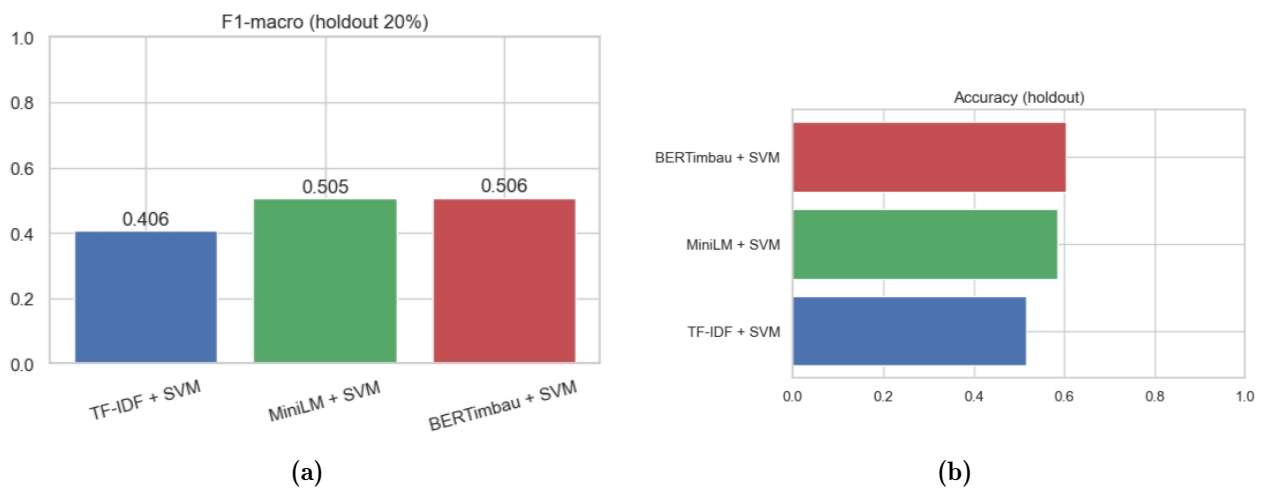


Figura 3.12: Comparação entre modelos clássicos e embeddings contextuais considerando F1-score macro (a) e accuracy (b).

Os resultados demonstram que os modelos baseados em embeddings contextuais apresentaram desempenho superior aos modelos tradicionais baseados apenas em frequência de palavras. Entretanto, o melhor resultado geral foi obtido com o BERTimbau após o processo de fine-tuning, que superou todas as demais abordagens avaliadas. O BERTimbau apresentou desempenho ligeiramente superior ao MiniLM, especialmente na categoria *tecnológico*. Esse resultado é coerente com o fato de o BERTimbau ter sido treinado especificamente para português, tornando-se mais adequado para representar as respostas coletadas.

Além disso diferentemente do TF-IDF, que trabalha com a frequência dos termos, o BERTimbau produz representações vetoriais que permitem capturar significados semelhantes mesmo quando usa palavras diferentes para expressar. Essa característica é bem relevante neste trabalho, pois as descrições comportamentais foram escritas de forma livre pelos responsáveis, apresentando elevada variabilidade lexical e sintática.

3.5 Avaliação

Segundo (CHAPMAN et al., 2000), a etapa de Avaliação tem como objetivo verificar se os modelos desenvolvidos atendem aos objetivos definidos para o projeto. Para isso, a avaliação foi realizada utilizando estratégias de validação estratificada para preservar a distribuição das

classes.

Inicialmente, o conjunto de dados foi dividido em:

- 80% para treino;
- 20% para teste.

A divisão foi realizada utilizando *StratifiedShuffleSplit*.

Durante os experimentos, aplicou-se validação cruzada estratificada repetida (*RepeatedStratifiedKfold*), utilizando 5 folds e 2 repetições, permitindo estimativas mais estáveis das métricas.

As métricas utilizadas foram:

- acurácia;
- precisão macro;
- recall macro;
- F1-score macro.

O F1-score macro foi adotado como principal métrica de comparação devido ao desbalanceamento entre as categorias do dataset. Também foi utilizado *GridSearchCV* para ajuste de hiperparâmetros dos modelos e da vetorização textual. Para os experimentos envolvendo o ajuste fino do BERTimbau, adotou-se uma estratégia de validação específica. Após a separação inicial dos dados em 80

Após os experimentos comparativos, os modelos foram avaliados no conjunto holdout, permitindo analisar sua capacidade de generalização em dados não vistos durante o treinamento.

A Tabela 3.2 apresenta os resultados gerais obtidos pelos principais modelos avaliados.

Tabela 3.2: Comparação geral entre os modelos avaliados

Modelo	Accuracy	F1-Macro
TF-IDF + SVM	0.5149	0.4058
MiniLM + SVM	0.5842	0.5047
BERTimbau + SVM	0.6040	0.5064
BERTimbau fine-tuning	0.6634	0.6340

Os resultados demonstraram que os modelos baseados em embeddings contextuais apresentaram desempenho superior em relação aos modelos clássicos baseados apenas em TF-IDF.

Entre os modelos avaliados, o BERTimbau fine-tuning apresentou o melhor desempenho geral, alcançando accuracy de 0.6634 e F1-score macro de 0.634 no conjunto holdout.

3.6 Implantação

A etapa de Implantação corresponde à disponibilização da solução no ambiente que irá atender às necessidades do usuário final (CHAPMAN et al., 2000). Neste trabalho, após a etapa de avaliação, o modelo selecionado foi disponibilizado para consumo pela plataforma GuideAut, permitindo sua utilização durante o processo de construção do PersonAut. A integração foi realizada por meio de uma API REST responsável por receber descrições comportamentais em linguagem natural e retornar a categoria predita pelo modelo de PLN.

O serviço foi desenvolvido de forma independente da plataforma, permitindo futuras atualizações do modelo sem necessidade de alterações na interface do GuideAut. Esse endpoint recebe uma frase em texto livre e retorna a categoria prevista pelo modelo treinado. O endpoint principal disponibilizado pela API é apresentado a seguir:

Exemplo de requisição:

```
POST /classify
{ "text": "Ele gosta de ouvir música para se acalmar" }
```

Exemplo de resposta:

```
{ "category": "acalma" }
```

A integração foi incorporada ao fluxo de criação do PersonAut dentro da plataforma GuideAut. Inicialmente, o usuário fornece uma descrição textual contendo características comportamentais da pessoa autista. Em seguida, o texto é enviado para a API de classificação, que identifica automaticamente a categoria correspondente para cada trecho analisado.

Os resultados retornados são utilizados para preencher automaticamente as seções do PersonAut, organizando as informações nas dimensões adequadas. Além disso, o fluxo permite que

o usuário complemente informações demográficas e características gerais da persona antes da geração final do artefato.

A Figura 3.13 apresenta a interface utilizada para inserção das descrições textuais e acionamento do processo de classificação automática. A Figura 3.14 apresenta a etapa de preenchimento dos dados demográficos da persona. Por fim, a Figura 3.15 apresenta o PersonAut gerado após a classificação automática e organização das informações fornecidas pelo usuário.

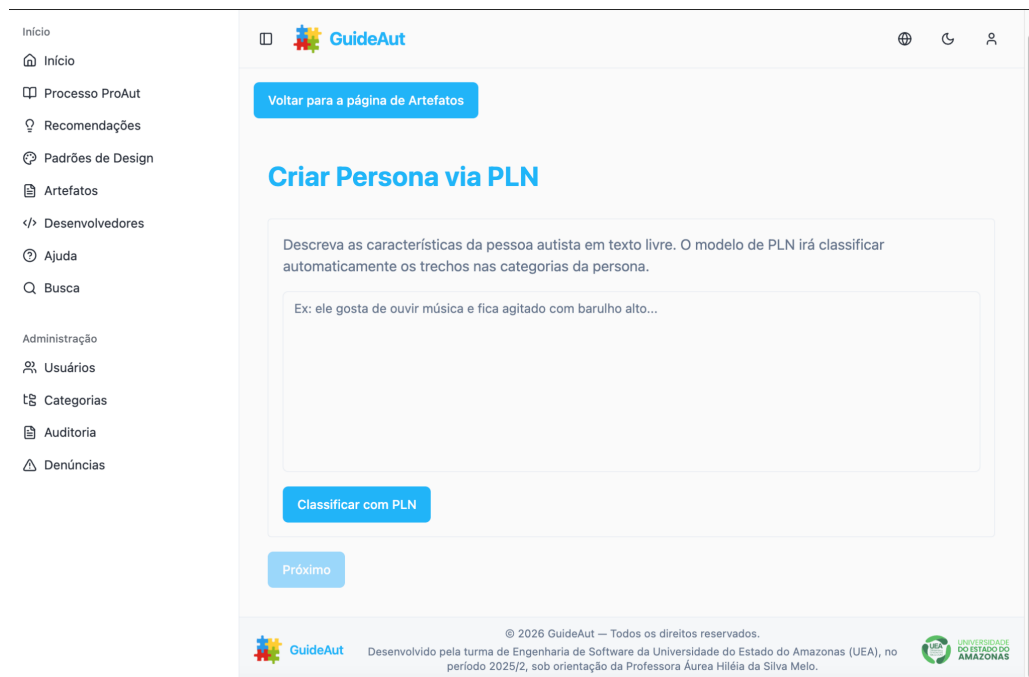
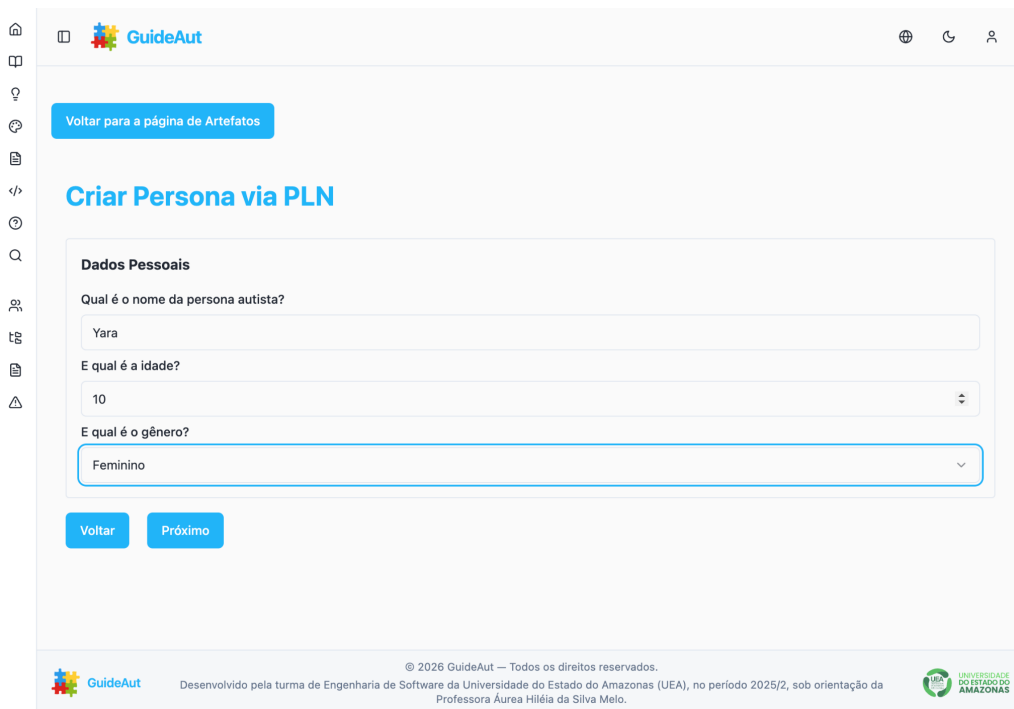


Figura 3.13: Interface para inserção das descrições comportamentais e classificação automática via PLN.



The screenshot displays the 'Criar Persona via PLN' (Create Persona via PLN) form in the GuideAut application. The interface includes a sidebar with navigation icons, a top header with the GuideAut logo and user controls, and a main content area. A blue button at the top left of the main area reads 'Voltar para a página de Artefatos'. The form is titled 'Criar Persona via PLN' and is divided into a 'Dados Pessoais' (Personal Data) section. This section contains three fields: 'Qual é o nome da persona autista?' (What is the name of the autistic persona?) with the value 'Yara', 'E qual é a idade?' (And what is the age?) with the value '10', and 'E qual é o gênero?' (And what is the gender?) with the value 'Feminino'. Below the form are two blue buttons: 'Voltar' (Back) and 'Próximo' (Next). The footer contains the GuideAut logo, copyright information for 2026, and the logo of the Universidade do Estado do Amazonas (UEA).

Dados Pessoais

Qual é o nome da persona autista?

Yara

E qual é a idade?

10

E qual é o gênero?

Feminino

[Voltar](#) [Próximo](#)

© 2026 GuideAut — Todos os direitos reservados.
Desenvolvido pela turma de Engenharia de Software da Universidade do Estado do Amazonas (UEA), no período 2025/2, sob orientação da Professora Áurea Hiléia da Silva Melo.

Figura 3.14: Preenchimento dos dados demográficos da persona.



Figura 3.15: PersonAut gerado após a classificação automática das descrições comportamentais.

3.6.1 Prova de Conceito da Integração

Após a implantação da API e a integração à plataforma GuideAut, foi realizada uma prova de conceito com o objetivo de avaliar o comportamento da solução em cenários próximos ao uso real. Para isso, foram elaboradas 100 descrições comportamentais nunca vistas pelo modelo, distribuídas igualmente entre as cinco categorias do PersonAut (*acalma*, *estressa*, *social*, *estereotipia* e *tecnológico*). As descrições foram inseridas na interface integrada da plataforma, permitindo avaliar o fluxo completo de classificação, desde o envio do texto até a organização automática das informações.

A Tabela 3.3 apresenta os resultados obtidos.

Tabela 3.3: Resultados da prova de conceito da solução integrada

Categoria	Casos Avaliados	Acertos
Acalma	20	14
Estressa	20	16
Social	20	18
Estereotipia	20	13
Tecnológico	20	13
Total	100	74

A solução apresentou concordância em 74 das 100 descrições avaliadas, correspondendo a uma taxa de concordância de 74% (74/100). Os melhores resultados foram observados nas categorias *social* (18 acertos) e *estressa* (16 acertos). Já as categorias *acalma*, *estereotipia* e *tecnológico* apresentaram maior número de erros, principalmente em descrições semanticamente semelhantes às de outras categorias.

Além dos resultados quantitativos, os testes demonstraram o funcionamento adequado do fluxo completo da solução, incluindo a comunicação entre a plataforma GuideAut e a API de classificação, o processamento das descrições textuais e a geração automática das categorias utilizadas na construção do PersonAut. Dessa forma, a prova de conceito permitiu validar a viabilidade técnica da integração proposta e demonstrar o potencial da solução para apoiar a organização apoiada de descrições comportamentais em um cenário próximo ao uso real.

3.7 Síntese Geral dos Resultados

Os experimentos realizados demonstraram que a classificação automática das descrições comportamentais do PersonAut é viável utilizando técnicas de Processamento de Linguagem Natural. Os resultados também evidenciaram que representações baseadas apenas em frequência de palavras possuem limitações para capturar o contexto semântico das descrições, especialmente em frases subjetivas e semanticamente ambíguas.

Os modelos baseados em embeddings contextuais apresentaram melhor capacidade de generalização, obtendo maiores valores de accuracy e F1-score macro. Entre os modelos avaliados, o BERTimbau fine-tuning apresentou o melhor equilíbrio geral entre desempenho e estabilidade,

sendo selecionado como modelo final da solução proposta.

Apesar dos resultados promissores, o trabalho ainda apresenta limitações relacionadas principalmente ao tamanho reduzido do dataset e ao desbalanceamento entre categorias. Ainda assim, os resultados obtidos demonstram o potencial do uso de PLN para apoiar a construção automatizada do PersonAut e contribuir para o desenvolvimento de sistemas mais acessíveis e adaptados ao contexto do TEA.

Considerações Finais

Este trabalho apresentou o desenvolvimento de uma abordagem baseada em Processamento de Linguagem Natural (PLN) para classificação automática de descrições comportamentais relacionadas ao PersonAut. O principal objetivo da pesquisa foi investigar a viabilidade do uso de técnicas de PLN para auxiliar na organização automática de informações textuais fornecidas por responsáveis de pessoas autistas, contribuindo para a construção do PersonAut por designers e desenvolvedores.

Inicialmente, observou-se a ausência de datasets textuais voltados especificamente para descrições comportamentais. A partir daí, construiu-se um dataset próprio alinhado às categorias do PersonAut, com base nas respostas discursivas coletadas por meio de formulário aplicado a responsáveis de pessoas autistas. Após as etapas de segmentação textual e anotação manual obteve-se um conjunto final com 456 exemplos textuais distribuídos entre cinco categorias: *acalma*, *estressa*, *social*, *estereotipia* e *tecnológico*.

Durante o desenvolvimento da solução, foram implementadas etapas de pré-processamento textual, vetorização por TF-IDF e geração de embeddings contextuais utilizando MiniLM, BERTimbau e BERTimbau fine-tuning. Além disso, foram avaliados diferentes classificadores supervisionados, incluindo Multinomial Naive Bayes e LinearSVC. Os resultados obtidos demonstraram que modelos baseados em embeddings contextuais apresentaram desempenho superior aos modelos clássicos baseados apenas em frequência de palavras. Entre os modelos avaliados, o BERTimbau fine-tuning apresentou o melhor desempenho geral, alcançando accuracy de 0.6634 e F1-score macro de 0.6340 no conjunto holdout. Os experimentos também evidenciaram desafios importantes relacionados ao tamanho reduzido do dataset, ao desbalanceamento

entre categorias e à sobreposição semântica entre algumas classes. Ainda assim, os resultados demonstram que descrições comportamentais escritas de maneira espontânea possuem padrões linguísticos suficientemente consistentes para serem modelados computacionalmente por técnicas de PLN.

Além da contribuição técnica, este trabalho também contribui indiretamente para a área de acessibilidade e design inclusivo ao propor uma estratégia capaz de apoiar parcialmente a construção automatizada do PersonAut, reduzindo esforço manual durante o processo de análise textual. Como trabalhos futuros, recomenda-se a ampliação do dataset por meio da coleta de novas respostas junto a familiares, profissionais e instituições especializadas em TEA. Além disso incluir mais de um anotador e fazer análise de concordância entre eles. Também podem ser investigadas abordagens mais avançadas baseadas em arquiteturas Transformer ajustadas especificamente para o domínio do autismo.

Além das contribuições no contexto do PersonAut, os resultados obtidos reforçam o potencial do PLN para apoiar a organização e análise de informações textuais em diferentes contextos. A abordagem desenvolvida neste trabalho pode servir como base para aplicações voltadas à classificação automática de relatos, formulários discursivos e descrições comportamentais em áreas como educação, saúde e acessibilidade, contribuindo para a extração estruturada de conhecimento a partir de dados textuais produzidos em linguagem natural. Por fim, entende-se que o uso de PLN representa uma abordagem promissora também para apoiar a organização automática de descrições comportamentais no contexto do PersonAut, contribuindo para o desenvolvimento de sistemas mais acessíveis, personalizados e alinhados às necessidades de pessoas autistas.

A. Formulário aplicado aos responsáveis

Este anexo apresenta o formulário utilizado para a coleta dos dados junto a pais, mães e responsáveis de pessoas autistas. O instrumento inclui questões discursivas distribuídas segundo as dimensões do PersonAut, além do Termo de Consentimento Livre e Esclarecido (TCLE).

Instrução

O formulário completo, conforme aplicado, encontra-se reproduzido nas páginas seguintes.

Formulário — Perfil Comportamental e Rotinas da Pessoa Autista

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) *

Este formulário faz parte de uma pesquisa de **Trabalho de Conclusão de Curso (TCC)** da aluna **Yara Nunes de Oliveira**, do curso de Engenharia da Computação da **Universidade do Estado do Amazonas (UEA)**.

O estudo tem como objetivo **coletar informações reais sobre comportamentos, características, sensibilidades, estereotípias, atividades e interações sociais de pessoas no espectro autista**, a partir da perspectiva de pais, mães ou responsáveis, com a finalidade de **construir um dataset para análise acadêmica e desenvolvimento de soluções tecnológicas voltadas ao autismo**.

A pesquisa está relacionada diretamente ao desenvolvimento do TCC, cujo foco envolve **a compreensão de padrões comportamentais e interesses de pessoas autistas**, bem como fatores que acalmam, estressam ou influenciam sua rotina.

Dessa forma, você foi convidado(a) a participar desta pesquisa como voluntário(a), devido à sua vivência e experiência direta com a pessoa autista sob sua responsabilidade. A participação consiste em responder a este questionário, cujas informações serão utilizadas **exclusivamente para fins acadêmicos e científicos**, garantindo que nenhum **dado que permita identificar você ou a pessoa autista serão divulgados sob nenhuma circunstância**. Todas as informações coletadas serão tratadas com rigoroso sigilo e confidencialidade, seguindo os princípios éticos da pesquisa científica.

Sua participação é **livre e voluntária**. Você pode recusar-se a participar, retirar seu consentimento ou interromper sua colaboração a qualquer momento, sem qualquer tipo de penalidade ou prejuízo.

Este questionário utiliza a plataforma Google Forms. Dessa forma, a coleta e o armazenamento das respostas estão sujeitos à **Política de Privacidade do Google**,

disponível em:

<https://www.google.com/policies/privacy/>.

47

Ao selecionar a opção “**Aceito participar da pesquisa**” e clicar no botão “**Avançar**”, você estará declarando que leu e compreendeu as informações acima e que concorda em participar de forma livre e esclarecida.

Em caso de dúvidas ou necessidade de esclarecimentos adicionais, você pode entrar em contato com a pesquisadora responsável:

Pesquisadora responsável:

Yara Nunes de Oliveira

Universidade do Estado do Amazonas (UEA) – Escola Superior de Tecnologia (EST)

E-mail: yndo.eng21@uea.edu.br

Orientadora:

Áurea Hiléia da Silva Melo

Universidade do Estado do Amazonas (UEA) – Escola Superior de Tecnologia (EST)

Agradecemos desde já pela sua colaboração e contribuição para o desenvolvimento desta pesquisa.

OBS: Perguntas marcadas com “*” são de resposta obrigatória.

☒ Aceito participar da pesquisa.

☐ Não aceito a participação.

1. Informações Gerais

Essas perguntas ajudam na contextualização e análise posterior

1.1 Nome da pessoa autista: *

Lucas Andrade

1.2 Idade: *

48

8

1.3 Gênero: *

☐

Feminino

☒

Masculino

☐

Prefiro não declarar

1.4 Nível de suporte: *

☐

Nível 1 (leve / precisa de pouco suporte)

☒

Nível 2 (moderado / suporte substancial)

☐

Nível 3 (severo / suporte muito substancial)

☐

Prefiro não informar

49

1.5 Tipo de linguagem predominante: *

- ☐ Não verbal
- ☐ Pouco verbal
- ☐ Frases curtas
- ☒ Comunicação funcional
- ☐ Comunicação avançada
- ☐ Ecológica
- ☐ Outro:

1.6 Ele/ela possui sensibilidade ao som? *

- ☒ Sim
- ☐ Não
- ☐ Às vezes / depende do ambiente
- ☐ Outro:

50

1.7 Qual seu vínculo com ele/ela: *

- ☐ Pai/Mãe
- ☐ Tio/tia
- ☐ Avô/avó
- ☐ Professor(a)
- ☐ Psicopedagogo(a)
- ☐ Terapeuta
- ☐ Psicólogo
- ☐ Fonoaudiólogo
- ☐ Outro:

2. Atividades que Acalmam

O que costuma ajudar a pessoa a se regular, relaxar ou se sentir segura?

2.1 Descreva duas ou mais situações que ajudam ele/ela a se **acalmar/regular**. (ex.: *
ouvir música/sons específicos, uso de celular/tablet, brinquedos sensoriais etc.)

O Lucas se acalma muito quando escuta músicas suaves, principalmente trilhas instrumentais. Ele também gosta de ficar balançando na rede no quarto dele — isso sempre ajuda quando ele está mais agitado. Outra coisa que funciona é brincar com massinha ou com aqueles brinquedos de textura, porque ele gosta muito de sentir coisas diferentes nas mãos.

3. Atividades que Estressam

Gatilhos comportamentais, situações desconfortáveis ou demandas difíceis.

3.1 Descreva duas ou mais situações que costumam **estressar /desregular** ele/ela. (ex.: * barulho alto, multidões, mudança de rotina etc.)

Ele se estressa bastante com barulhos repentinos, como fogos, moto acelerando ou alguém batendo porta. Lugares muito cheios, tipo supermercados, também deixam ele muito ansioso e inquieto. Mudanças de rotina, como trocar o horário do banho ou da escola, também podem desencadear crises de choro.

4. Aspectos Sociais e Familiares

Como é a interação da pessoa autista com família e outras pessoas.

4.1 Como ele/ela interage com familiares? *

Com a família ele é carinhoso, mas do jeito dele. Ele gosta de ficar perto, encostar, e às vezes abraça quando está confortável. Não gosta muito de contato visual, mas demonstra afeto ficando junto, sentado pertinho.

4.2 Como é a interação dele/dela com outras pessoas? *

Ele tem dificuldade de iniciar conversas e prefere brincar sozinho. Às vezes observa outras crianças brincar e até sorri, mas raramente entra na brincadeira sem ajuda.

4.3 Que dificuldades ele/ela geralmente apresenta ao estar em ambientes sociais? *

Ambientes muito barulhentos deixam ele sobrecarregado. Também tem dificuldade em esperar sua vez ou entender algumas regras de brincadeira, o que gera frustração.

4.4 Como ele/ela costuma reagir à rotina da casa (tarefas, horários, mudanças)? 52 *

O Lucas funciona muito melhor com rotina. Quando tudo segue como ele espera, ele fica tranquilo. Se algo muda, ele fica ansioso, começa a andar de um lado para o outro e pergunta várias vezes o que vai acontecer.

5. Estereotipias / Manias

Comportamentos repetitivos e padrões que aparecem no dia a dia.

5.1 Quais estereotipias, manias ou comportamentos repetitivos ele/ela apresenta?(ex.: * Balançar o corpo, tapar os ouvidos, repetir frases (ecolalia), andar de um lado para o outro, girar em torno de si mesmo, balançar as mãos(flapping), etc...)

Ele balança o corpo para frente e para trás quando está empolgado ou nervoso. Também bate as mãos levemente quando fica muito feliz. Às vezes repete frases de desenhos animados que gosta (ecolalia), principalmente antes de dormir.

6. Aspectos Tecnológicos (Uso de Software/App)

Informe como a pessoa autista utiliza a tecnologia no dia a dia.

6.1 Ele/ela usa algum app com frequência? Quais?(ex.: Youtube, Netflix, Instagram, * etc...)

YouTube Kids e Netflix. De vez em quando joga no tablet jogos bem simples.

53

6.2 Que tipo de conteúdo ele/ela mais consome? *

- ☒ Vídeos curtos
- ☐ Jogos simples
- ☒ Sons/músicas repetidas
- ☒ Desenhos
- ☐ Conteúdo educacional
- ☐ Filme
- ☐ Receitas culinárias
- ☐ Experimentos científicos
- ☐ Videoclipe
- ☐ Outro:

6.3 Há algo em tecnologia que estressa ele/ela? *

(Ex.: barulho do vídeo, esperar carregar, travar, notificações etc.)

Ele fica irritado quando o vídeo trava ou demora para carregar. Também não gosta de propagandas repentinas.
.....

7. Características Gerais

Aspectos importantes para entender o perfil da pessoa autista.

54

7.1 Ele/ela tem alguma rotina fixa? Qual? *

im. Ele acorda todos os dias no mesmo horário, toma o café com as mesmas comidas e só troca de roupa depois de ver um episódio do desenho preferido. Depois da escola, sempre tem um tempo de descanso na rede antes de fazer qualquer outra coisa.

7.2 Quais são os principais interesses dele/dela? *

Ele ama astronomia, fala de planetas o tempo todo. Também gosta de puzzles e de brinquedos de montar. Músicas repetitivas também chamam muito a atenção dele.

7.3 Há algo importante que você gostaria de acrescentar? *

O Lucas é muito inteligente e sensível aos detalhes. Ele tem dificuldade com mudanças, mas quando ganha confiança em alguém se torna extremamente afetuoso. Sempre que é respeitado no ritmo dele, ele floresce.

Este formulário foi criado em Universidade do Estado do Amazonas - UEA.

Google Formulários

Referências Bibliográficas

AGGARWAL, C. C.; ZHAI, C. An introduction to text mining. In: _____. *Mining Text Data*. Boston, MA: Springer US, 2012. p. 1–10. ISBN 978-1-4614-3223-4. Disponível em: <https://doi.org/10.1007/978-1-4614-3223-4_1>.

ALMEIDA, J. d.; RASSI, A. P. Linguistic ambiguity analysis in large language models (LLMs). *Trabalhos em Linguística Aplicada*, SciELO, v. 63, n. 3, p. e24132, 2024. Disponível em: <<https://www.scielo.br/j/tl/a/Z6nCFZLTHX6XVmmhJsKvWxd/>>.

ALVES, G. S. *Agilidade em projetos de ciências de dados: Um estudo sobre a adoção do método ágil XP com modelo de processo CRISP-DM*. Dissertação (Mestrado) — Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT), São Paulo, 2023. Disponível em: <https://repositorio.usp.br/directbitstream/7fab9faa-9102-4418-8473-e4ea81df7350/3275880/_-_Integrating_CRISP-DM_and_eXtreme_programming_in_agile_data_science_projects.pdf>.

American Psychiatric Association. *Manual Diagnóstico e Estatístico de Transtornos Mentais - DSM-5*. Porto Alegre: Artmed, 2014. Tradução de Maria Inês Corrêa Nascimento et al.

BORGES, B. R. Trabalho de Conclusão de Curso, *Análise Comparativa de Algoritmos de Classificação de Texto*. Uberlândia, MG: [s.n.], 2023. Disponível em: <<https://repositorio.ufu.br/handle/123456789/43701>>.

CHAPMAN, P. et al. *CRISP-DM 1.0: Step-by-step Data Mining Guide*. Chicago, USA: SPSS Inc., 2000.

CHEN, W. et al. Designing assistive technologies for the autistic spectrum: A systematic review. *Journal of Autism and Developmental Disorders*, v. 46, n. 1, p. 1–16, 2016.

CHIEN, M. E. et al. ican: a tablet-based pedagogical system for improving communication skills of children with autism. *International Journal of Human-Computer Studies*, v. 73, p. 79–90, jan. 2015.

INDURKHYA, N.; DAMERAU, F. J. *Handbook of Natural Language Processing*. 2. ed. [S.l.]: Chapman & Hall/CRC, 2010.

JURAFSKY, D.; MARTIN, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd draft. ed. Stanford University / Prentice Hall, 2024. Disponível em: <<https://stanford.edu>>.

- KHURANA, D. et al. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, Springer Science and Business Media LLC, v. 82, n. 3, p. 3713–3744, 2022. ISSN 1573-7721. Disponível em: <<http://dx.doi.org/10.1007/s11042-022-13428-4>>.
- LADEIRA, F.; AMARAL, I. *A educação de alunos com multideficiência nas Escolas de Ensino Regular*. Lisboa: Ministério da Educação, Departamento da Educação Básica, 1999. (Coleção Apoios Educativos).
- MELO, Á. H. d. S. et al. Empathyaut: an empathy map for people with autism. In: *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*. [S.l.: s.n.], 2020. p. 1–6.
- MELO, H. d. S. et al. Personaut: A personas model for people with autism spectrum disorder. In: *Anais do XIX Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação, 2020. p. 466–471. Disponível em: <<https://sol.sbc.org.br/index.php/ihc/article/view/13851>>.
- MELO Áurea Hiléia da S. *ProAut: um processo para apoiar a prototipação de interfaces de aplicações para autistas*. 289 p. Tese (Doutorado) — Universidade Federal do Amazonas, Instituto de Computação, Manaus, 2021.
- NORMAN, D. A.; DRAPER, S. W. *User Centered System Design: New Perspectives on Human-Computer Interaction*. [S.l.]: Lawrence Erlbaum Associates, 1986.
- PASSERINO, L. M.; MONTARDO, S. P. Inclusão social via acessibilidade digital: proposta de inclusão digital para pessoas com necessidades especiais. *E-Compós*, v. 8, 2007.
- PRUITT, J.; ADLIN, T. *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. [S.l.]: Morgan Kaufmann, 2006.
- SHIMAOKA, A. M.; FERREIRA, R. C.; GOLDMAN, A. The evolution of crisp-dm for data science: Methods, processes and frameworks. *SBC Computing Reviews*, v. 4, n. 1, p. 28–43, Oct. 2024. Disponível em: <<https://journals-sol.sbc.org.br/index.php/reviews/article/view/3757>>.
- SOUSA, A. F. d. S.; NETO, P. M. d. C.; PESSOA, A. N. L. d. S. Inclusão digital: Os obstáculos a serem enfrentados na busca pela dignidade na sociedade conectada. *RECIMA21 - Revista Científica Multidisciplinar*, v. 4, n. 6, p. e463353, 2023.
- SOUZA, A. C. de; BENITEZ, P.; CARMO, J. dos S. Diretrizes de acessibilidade de interfaces digitais para pessoas com transtorno do espectro autista: uma revisão integrativa de literatura. *Revista Educação Especial*, 2021. [Em linha].
- VANDENBUSSCHE, N. et al. Using natural language processing to automatically classify written self-reported narratives by patients with migraine or cluster headache. *The Journal of Headache and Pain*, v. 23, n. 1, p. 129, 2022.
- World Wide Web Consortium (W3C). *Web Content Accessibility Guidelines (WCAG) 2.2*. [S.l.], 2023. Accessed: 2025-10-05. Disponível em: <<https://www.w3.org/TR/WCAG22/>>.