

16º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2025

MindParse: Classificação de Personalidade Utilizando Aprendizado de Máquina

JOÃO PEDRO MACHADO SILVA¹, LEONARDO SOLOVIJOVAS SANTOS², ARTHUR
EMANUEL DE OLIVEIRA CAROSIA³

¹ Graduando em Bacharelado em Ciência da Computação, IFSP, Campus São João da Boa Vista, jp.dausi@hotmail.com.

² Graduando em Bacharelado em Ciência da Computação, IFSP, Campus São João da Boa Vista, leo.solovijovas@gmail.com.

³ Docente do Instituto Federal de São Paulo, Campus São João da Boa Vista, arthuremanuel.carosia@ifsp.edu.br.

Área de conhecimento (Tabela CNPq): 1.03.00.00-7 Ciência da Computação

RESUMO: Este trabalho apresenta um sistema de classificação de personalidade com foco nos traços de extroversão e introversão, utilizando técnicas de aprendizado de máquina. A análise da personalidade humana, especialmente sob a ótica da introversão e extroversão, é um tema amplamente estudado na psicologia e possui grande relevância para compreender padrões de comportamento e interação social. Assim, a escolha desse objeto de pesquisa busca integrar aspectos psicológicos e computacionais, explorando como algoritmos podem reconhecer tendências comportamentais de forma automatizada e objetiva. A partir da base de dados simulada *Extrovert vs. Introvert Behavior Data*, foram testados modelos como MLP, SVM e Random Forest para prever a personalidade com base em variáveis como tempo sozinho, medo de palco e tamanho do círculo social. Os modelos MLP e SVM apresentaram acurácia superior a 92%. Essa abordagem tem potencial para auxiliar profissionais de psicologia e áreas correlatas na identificação de perfis comportamentais, reduzindo vieses de autorrelato e ampliando o uso de dados quantitativos em avaliações de personalidade. A proposta, portanto, une psicologia e inteligência artificial, promovendo um avanço na integração entre tecnologia e ciências humanas.

PALAVRAS-CHAVE: personalidade; extroversão; introversão; inteligência artificial; aprendizado de máquina.

MindParse: Personality Classification Using Machine Learning

ABSTRACT: This work presents a personality classification system focused on the traits of extroversion and introversion, utilizing machine learning techniques. The analysis of human personality, especially through the lens of introversion and extroversion, is a widely studied topic in psychology and is highly relevant for understanding behavioral and social interaction patterns. Thus, the choice of this research topic aims to integrate psychological and computational aspects, exploring how algorithms can recognize behavioral tendencies in an automated and objective manner. Based on the simulated dataset *Extrovert vs. Introvert Behavior Data*, models such as MLP, SVM, and Random Forest were tested to predict personality based on variables like time spent alone, stage fear, and social circle size. The MLP and SVM models achieved an accuracy above 92%. This approach has the potential to assist psychology professionals and related fields in identifying behavioral profiles, reducing self-report bias, and expanding the use of quantitative data in personality assessments.

Therefore, the proposal unites psychology and artificial intelligence, promoting progress in the integration between technology and the human sciences.

KEYWORDS: personality; extroversion; introversion; artificial intelligence; machine learning.

INTRODUÇÃO

A personalidade humana é uma área crucial da psicologia desde o século 20, sendo uma das áreas a introversão-extroversão. Em 1921, Carl Gustav Jung propôs uma definição para introversão e extroversão. Extroversão seria quando o indivíduo carrega sua energia em ambientes sociais e a introversão, quando ele prefere a solidão para manter seu equilíbrio emocional. Tradicionalmente, testes psicológicos como o MBTI (Myers & Briggs, 1980) ou o método conhecido como cinco Grandes Fatores de Personalidade (Costa & McCrae, 1992) são realizados para identificar esses traços. Porém, esses métodos são dependentes de autorrelatos e podem apresentar erros contínuos.

Os avanços da Inteligência Artificial recentemente, principalmente no aprendizado de máquina (Machine Learning), tem tornado possível automatizar e otimizar a análise de dados comportamentais, dados sobre o comportamento das pessoas. Do mesmo modo, métodos supervisionados de classificação foram melhorados e utilizados com sucesso em muitas áreas, como em exames clínicos, análise de emoções e recomendação personalizada (Goodfellow et al., 2016; Pedregosa et al., 2011). Em paralelo a tudo isso, surge a possibilidade de usar métodos de aprendizado de máquina para categorizar a personalidade das pessoas com base na identificação de características comportamentais mais fáceis de serem percebidas.

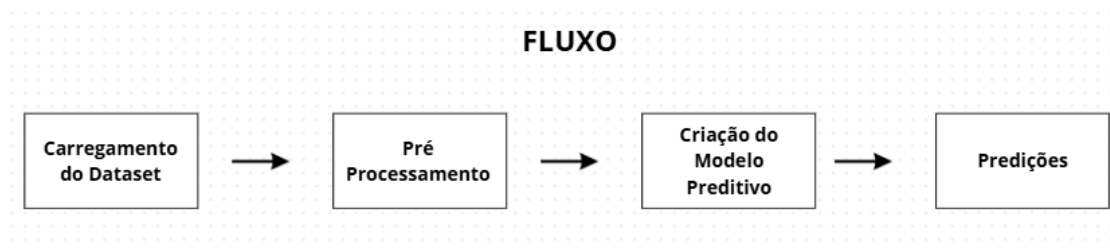
Neste contexto, o presente projeto propõe a criação de um sistema capaz de prever e classificar indivíduos como introvertidos ou extrovertidos com base em características comportamentais observáveis. Para isso, utilizou-se a base de dados *Extrovert vs. Introvert Behavior Data* (Kapilavayi, 2025) e modelos de aprendizado de máquina MLP (Multi-Layer Perceptron), SVM (*Support Vector Machine*) e Random Forest. A escolha desses modelos se justifica por sua complementaridade: a MLP representa uma abordagem de redes neurais capaz de capturar relações não lineares complexas entre variáveis; o SVM é amplamente reconhecido por sua eficiência em conjuntos de dados pequenos e boa separação entre classes; e o Random Forest oferece robustez contra ruído e sobreajuste, além de fornecer interpretabilidade por meio da importância das variáveis. Essa combinação visa equilibrar precisão, generalização e interpretabilidade no processo de classificação da personalidade.

Dessa forma, o estudo busca integrar conceitos psicológicos e computacionais, explorando o potencial das técnicas de aprendizado de máquina na análise de traços humanos. A proposta reforça a convergência entre psicologia e inteligência artificial, contribuindo para a construção de modelos mais objetivos e acessíveis na avaliação da personalidade humana.

MATERIAL E MÉTODOS

O processo metodológico abrangeu diversas etapas interconectadas (Figura 1), sendo elas:

Figura 1. Etapas realizadas no projeto.



O conjunto de dados usado foi “extrovert-vs-introvert-behavior-data”(disponível em: <https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data>), oferece dados comportamentais e sociais oferecida por alunos de uma escola através de um

questionário, como tempo gasto sozinho, frequência em eventos sociais, medo de palco, e uso de mídias sociais. Esses dados nos permitem fazer Análise de padrões sociais, e prever a personalidade com base em suas respostas.

As etapas de pré-processamento foram as seguintes:

- **Tratamento de Valores Ausentes:** Todas as linhas que continham valores ausentes (NaN) foram removidas do dataset.
- **Codificação de Variáveis Categóricas:** A coluna 'Personality' foi mapeada para valores numéricos: 'Introvert' = 0 e 'Extrovert' = 1 e as colunas 'Stage_fear' e 'Drained_after_socializing', foram transformadas em representações numéricas: 'No' foi codificado como 0 e 'Yes' como 1.
- **Separação de Features e Target:** A coluna 'Personality' foi definida como a variável target, enquanto as demais colunas: 'Time_spent_Alone'(0-10), 'Stage_fear'(Yes ou No), 'Social_event_attendance'(0-10), 'Going_outside'(0-7), 'Drained_after_socializing'(Yes ou No), 'Friends_circle_size'(0-15), 'Post_frequency'(0-10) constituíram o conjunto de features.
- **Normalização dos Dados:** Para garantir que todas as variáveis contribuíssem igualmente para o aprendizado dos modelos, foi aplicado o MinMaxScaler do Scikit-learn, que transforma os dados para o intervalo [0,1], reduzindo a influência de escalas diferentes entre as features. Essa técnica foi escolhida por preservar a distribuição original dos dados e ser adequada para algoritmos sensíveis à escala, como MLP e SVM. O objeto MinMaxScaler treinado foi salvo para uso futuro em inferências com novos dados.
- **Divisão em Conjuntos de Treino e Teste:** O dataset pré-processado e normalizado foi dividido em conjuntos de treino (80%) e teste (20%).

Modelos de Classificação:

Três modelos de aprendizado de máquina foram selecionados e implementados para a tarefa de classificação binária (Extrovertido vs. Introvertido), sendo eles:

1. Multi-Layer Perceptron (MLP):

- **Arquitetura:** Criada a partir da API Sequential do Tensorflow/Keras. A arquitetura consistiu em uma camada de entrada com 64 neurônios seguida de camadas ocultas com 32 e 16 neurônios, todas utilizando a função de ativação ReLU (Rectified Linear Unit). A camada de saída possui 1 neurônio com função de ativação Sigmóide. Valores definidos de forma a equilibrar a complexidade do modelo e a capacidade de generalização. A quantidade moderada de neurônios (64–32–16) permite capturar relações não lineares sem causar sobreajuste, enquanto a função de ativação *ReLU* acelera o aprendizado nas camadas ocultas.
- **Compilação:** O modelo foi compilado com o otimizador Adam, utilizando uma taxa de aprendizado de 0.001, o que proporcionou convergência estável, e a função de perda *binary_crossentropy*.
- **Treinamento:** O modelo foi treinado no conjunto normalizado de treino por 30 épocas com um tamanho de batch de 16 por se tratar de um dataset relativamente pequeno.

2. Support Vector Machine (SVM):

- **Parâmetros:** Para o SVM, foi utilizado o Support Vector Classifier (SVC) com kernel RBF, o padrão do Scikit-learn, adequado para capturar relações não lineares. Os hiperparâmetros de regularização ($C = 1.0$) e de influência do kernel ($\gamma = 'scale'$) foram mantidos nos valores padrão, equilibrando complexidade e capacidade de generalização.

3. Random Forest:

- **Parâmetros:** foi utilizado um classificador *RandomForestClassifier* da biblioteca Scikit-learn. O número de estimadores foi definido como 100, valor escolhido por equilibrar desempenho e custo computacional, garantindo robustez e redução de

sobreajuste.

As previsões finais para as métricas de acurácia, precisão, recall e F1-score foram obtidas através da binarização das probabilidades (*Threshold* de 0.5).

RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados e analisados os resultados obtidos na avaliação dos modelos MLP, SVM e Random Forest, considerando métricas de desempenho como acurácia, precisão, recall, F1-Score e AUC-ROC. Os testes realizados indicam que todos os modelos alcançaram performances elevadas, com destaque para a MLP e a SVM, que apresentaram métricas praticamente idênticas e ligeiramente superiores à Random Forest. Além das métricas globais, também são discutidas as matrizes de confusão e a análise de correlação, que evidenciam a capacidade dos modelos em capturar de forma consistente os padrões associados às características de introversão e extroversão. Na avaliação dos modelos, os resultados e métricas de validação foram:

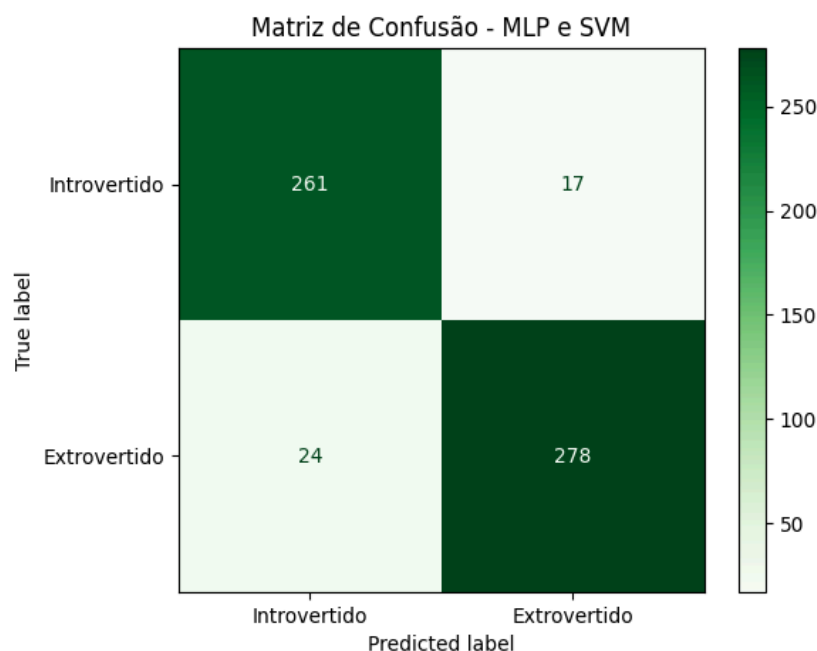
TABELA 1. Resultados da aplicação das métricas de validação para cada modelo.

	Acurácia	Perda	Precisão	Recall	F1-Score	AUC-ROC
MLP	0.9293	0.2412	0.9424	0.9205	0.9313	0.9531
SVM	0.9293	-	0.9424	0.9205	0.9313	0.9441
Random Forest	0.9224	-	0.9356	0.9139	0.9246	0.9488

Apesar do alto desempenho preditivo dos modelos, é crucial realizar uma análise crítica sobre as limitações e possíveis vieses do dataset '*Extrovert vs. Introvert Behavior Data*'. Primeiramente, a base de dados é descrita como sendo originada de um questionário aplicado a 'alunos de uma escola', o que introduz um viés de amostragem significativo. Os resultados, portanto, podem ter sua generalização limitada a uma faixa etária e contexto sociocultural específicos, não sendo diretamente aplicáveis a populações mais amplas. Em segundo lugar, a coleta de dados via questionário depende do autorrelato, o que pode acarretar imprecisões, pois a percepção que os indivíduos têm de si mesmos pode divergir de seu comportamento real. Por fim, a natureza potencialmente simulada ou idealizada dos dados pode simplificar a complexidade das interações humanas. Futuros trabalhos deveriam buscar validar esta abordagem com dados coletados de múltiplas fontes demográficas e, se possível, cruzando informações de autorrelato com dados observacionais para mitigar esses vieses.

Quanto a Matriz de Confusão, MLP e SVM tiveram os mesmos resultados, Random Forest performou ligeiramente pior, segue Matriz de Confusão:

Figura 2. Matriz de Confusão MLP e SVM.



Uma análise mais profunda, por meio de uma matriz de correlação, revelou padrões bastante lógicos, os números nos dizem que os resultados do modelo estão em perfeita sintonia com as definições clássicas de introversão e extroversão, o que valida tanto a qualidade dos dados quanto a inteligência do modelo em aprender com eles.

A fim de evitar previsões com confiança baixa, foi criada uma nova classe a partir da previsão, se refere aos “Ambivertidos” (pessoas que possuem características tanto de introvertidos quanto de extrovertidos), para confiança > 0.5 e < 0.7 são classificados como “Ambivertidos com tendência a Extrovertido”, para confiança > 0.3 e < 0.5 são classificados como “Ambivertidos com tendência a Introvertido”, para confiança > 0.7 são classificados como “Extrovertidos”, por fim, aqueles com confiança < 0.3 foram classificados como “Introvertidos”.

Analisando as previsões geradas pelos modelos que apresentaram os melhores resultados (MLP e SVM) em relação ao conjunto de testes, foi notado que MLP possui previsões mais próximas aos extremos (0 e 1), já o SVM possui previsões mais próximas a 0.5, visto que ambos os modelos possuem os mesmos resultados de acurácia e matriz de confusão, pode-se inferir que SVM é mais preciso para prever “Ambivertidos”, enquanto MLP é mais preciso para os extremos de “Extrovertido” e “Introvertido”.

CONCLUSÕES

O estudo demonstrou a viabilidade e eficácia da utilização de técnicas de aprendizado de máquina para a classificação de traços de personalidade, especificamente extroversão e introversão. A partir da base de dados simulada “Extrovert vs. Introvert Behavior Data”, os modelos MLP e SVM apresentaram desempenho superior, com acurácia acima de 92%, comprovando a capacidade desses algoritmos em capturar padrões comportamentais relevantes. A inclusão de uma categoria intermediária (“Ambivertidos”) permitiu refinar a interpretação dos resultados, evidenciando diferenças no comportamento preditivo de cada modelo: MLP se mostrou mais assertivo para os extremos, enquanto SVM se destacou na detecção de perfis intermediários.

Os resultados reforçam o potencial de integração entre psicologia e inteligência artificial, oferecendo uma abordagem rápida, objetiva e acessível para análise de personalidade. Além disso, o estudo sugere que a combinação de dados comportamentais bem estruturados e modelos supervisionados avançados pode servir de base para futuras pesquisas e aplicações práticas, como em contextos educacionais (para orientar métodos de ensino personalizados), clínicos (apoio em avaliações psicológicas) e organizacionais (processos de seleção e desenvolvimento de equipes).

CONTRIBUIÇÕES DOS AUTORES

L.S.S. e J.P.M.S. contribuíram para a análise dos dados coletados, bem como para a concepção e implementação da metodologia do projeto, incluindo a aplicação e validação dos modelos preditivos. A.E.O.C. contribuiu por meio da orientação durante todas as fases do projeto. Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Agradecemos primeiramente ao Instituto Federal de São Paulo (IFSP), campus São João da Boa Vista, pela estrutura e suporte oferecidos para a realização deste trabalho. Um agradecimento especial ao nosso orientador, Prof. Arthur Emanuel de Oliveira Carosia, por sua orientação e contribuições em todas as etapas deste projeto. Agradecemos também aos revisores anônimos deste congresso, cujas críticas e sugestões construtivas contribuíram significativamente para a melhoria da qualidade final deste artigo.

REFERÊNCIAS

Costa, P. T.; MCCRAE, R. R. *NEO PI-R: Professional manual*. Odessa: Psychological Assessment Resources, 1992.

Jung, C. G. *Psychological types*. Princeton University Press, 1921.

Kapilavayi, R. **Extrovert vs introvert behavior data**. Kaggle, 2023. Disponível em: <https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data>. Acesso em: 3 jul. 2025.

Myers, I. B.; BRIGGS, P. B. *Gifts differing: Understanding personality type*. Nicholas Brealey Publishing, 1980.