

Personality Prediction using Machine Learning

¹ Jyothika K R, ² Vaishnavi A Dixit

¹Assistant Professor, Department of MCA, BIET, Davanagere

²Student, 4th Semester MCA, Department of MCA, BIET, Davanagere

Abstract

Personality traits significantly influence human behavior, social interactions, and individual potential across various domains including psychology and business. This paper presents a Flask-based web application that predicts personality traits by leveraging psychological and demographic features through advanced machine learning models: LightGBM and CatBoost. The system accepts user inputs such as gender, age, and the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) via an intuitive web interface, and provides real-time personality type predictions. Utilizing a large-scale dataset with efficient categorical encoding, the models perform multi-class classification with high accuracy and consistency. The application also integrates administrative functionalities for user management, health tips dissemination, and query handling, fostering interactive communication between users and administrators. This work demonstrates the effective combination of data science and web development to build scalable, user-friendly tools for personality analysis, with potential applications in personal development, recruitment, and psychological research.

Keywords: *Personality, LightGBM, CatBoost, Personality traits*

I. INTRODUCTION

Personality is a complex and multifaceted construct that reflects the distinctive patterns of thoughts, emotions, and behaviors that define an individual. It has been extensively studied in psychology due to its profound influence on how people perceive their environment, interact socially, and respond to a variety of situations. The Five-Factor Model, commonly known as the Big Five personality traits—openness, conscientiousness, extraversion, agreeableness, and neuroticism—has emerged as a widely accepted framework for understanding personality. These traits provide a comprehensive

yet concise representation of human personality, capturing core dimensions that are relatively stable over time and predictive of diverse life outcomes. The ability to accurately assess and predict personality traits has significant implications across multiple fields, including psychology, education, healthcare, and business.

In practical terms, personality analysis aids in understanding individual differences that affect behavior and decision-making. Organizations leverage personality insights to optimize team dynamics, select appropriate leadership candidates, and improve employee placement, thereby

enhancing productivity and job satisfaction. Similarly, in educational settings, personality assessments can inform tailored learning approaches, while in healthcare, they contribute to better mental health diagnosis and personalized treatment plans. Despite the importance of personality assessment, traditional methods such as self-report questionnaires and clinical interviews often face challenges related to time consumption, subjective bias, and limited scalability. These limitations have motivated the exploration of data-driven approaches that harness the power of machine learning to provide faster, more objective, and scalable personality predictions.

Machine learning techniques have revolutionized many domains by enabling the analysis of large and complex datasets to uncover hidden patterns and relationships. In the context of personality prediction, machine learning models can process diverse input features—including demographic data and psychological characteristics—to classify personality types with high accuracy. Among the array of machine learning algorithms, gradient boosting frameworks such as Light Gradient Boosting Machine (LightGBM) and Categorical Boosting (CatBoost) have shown exceptional performance in classification tasks. LightGBM is known for its speed and efficiency, particularly with large datasets, while CatBoost excels in handling categorical variables and minimizing overfitting. Both models are capable of multi-class classification, making them well-suited for predicting personality traits that often involve multiple categories or levels.

The integration of these sophisticated machine learning models into a web-based application offers a powerful platform for delivering personality predictions to a broad audience. Web applications provide accessibility, user-friendly interfaces, and real-time interaction, enabling users to input their data and receive immediate feedback on their personality profiles. Flask, a lightweight Python web framework, facilitates the rapid development and deployment of such applications by allowing seamless integration of machine learning models with web technologies. By combining Flask with LightGBM and CatBoost, it is possible to build an interactive system that not only predicts personality traits accurately but also supports additional functionalities such as user management, health tips dissemination, and query handling between users and administrators.

This paper presents the development of a Flask-based web application designed to predict personality traits using psychological and demographic inputs processed through LightGBM and CatBoost models. The application accepts user inputs including gender, age, and the Big Five personality dimensions, then feeds this data into the machine learning models to generate real-time personality predictions. The system displays the results from both models side-by-side, allowing users to compare different machine learning approaches and gain deeper insights into their personality profiles. Beyond prediction, the platform incorporates administrative features that enable management of user accounts, posting of health-related tips, and communication via a query-response mechanism. This dual-user architecture

fosters an interactive environment where users can engage with administrators, enhancing the overall user experience and providing additional value beyond personality assessment.

The motivation behind this project stems from the growing need for scalable, accurate, and accessible personality prediction tools that can be utilized across various sectors. By leveraging advanced machine learning algorithms and web development frameworks, the system addresses common challenges associated with traditional personality assessment methods. It offers a scalable solution that can handle large datasets and diverse user inputs while maintaining ease of use through a clean and responsive web interface. The comparative analysis of LightGBM and CatBoost models within the application also contributes to the understanding of their respective strengths and limitations in the domain of personality prediction.

The remainder of this paper delves into the related research in personality prediction and machine learning applications, details the system architecture and implementation, presents experimental results evaluating model performance, and discusses the implications and future directions of this work. Through this project, we demonstrate how the synergy of data science and web development can produce practical, insightful tools that enhance psychological assessment and promote personal development.

II.RELATED WORK

Personality Prediction Using ML and NLP,
Authors: Adinath Mahandule, Gaurav Fiske, Vedang Malusare, Omkar Bhambure, Rajashree Shedge

The paper introduces a personality prediction system for students, leveraging machine learning (ML) and natural language processing (NLP) techniques based on responses to a structured questionnaire. The system classifies personality traits using the Myers-Briggs Type Indicator (MBTI) model. This approach holds substantial value in educational and recruitment domains, enabling educators to tailor learning strategies and helping recruiters assess personality fit for job roles. By offering insights into student behavior, the system promotes personalized learning and informed hiring decisions, demonstrating the practical application of AI in understanding human behavior.[1]

Personality Prediction Using Deep Learning,
Authors: Indira Dutta, R. Athilakshmi, Amulya,

This paper presents a deep learning-based personality prediction model that uses a Semi-Supervised Deep Embedded Clustering (DEC) approach to extract and classify personality traits. Targeting applications in mental health monitoring and criminal behavior analysis, the system effectively learns feature representations while performing clustering. Achieving 95.90% accuracy on a Kaggle personality dataset, the model outperforms traditional deep learning methods like RNN, LSTM, and GRU. The study highlights the growing impact of AI in understanding human

behavior and envisions future applications in smart mental health care systems.[2]

1. Applicant Personality Prediction System Using Machine Learning, Authors: M. Karnakar, Haseeb Ur Rahman, A B Jai Santhosh, NageswaraRao Sirisala

This paper introduces the Applicant Personality Prediction System (APPS), a machine learning-based tool designed to streamline the recruitment process for HR departments by evaluating candidate resumes and personality traits. The system uses the Spacy NLP module to compare resumes with job descriptions and collects personality test data rated on traits such as openness and conscientiousness. By applying multinomial logistic regression to these inputs, APPS predicts the candidate's personality and suitability for specific roles, significantly reducing the time and effort required for candidate shortlisting.[3]

Prediction of Personality Trait using Machine Learning on Online Texts, Authors: Raj Kishan Cherukuru, Ashish Kumar, Sumit Srivastava, Vivek Kumar Verma.

learning-based personality prediction system that analyzes user-generated text from social media to classify individuals based on MBTI personality traits. Using XGBoost, an advanced gradient boosting algorithm, the system predicts four key personality dimensions: Judging–Perceiving, Intuition–Sensing, Feeling–Thinking, and Introversion–Extraversion. To overcome dataset imbalance issues, the study employs resampling techniques, enhancing accuracy and model reliability. Pre-processing methods like tokenization, stemming, stop-word removal, and TF-IDF feature selection further refine the input

data. The model achieves 99% precision, demonstrating its potential in recruitment and personalized service delivery.[4]

2. Smart-Hire Personality Prediction Using ML, Authors: Isha Gupta, Manasvi Jain, Prashant Johri.

This paper presents a machine learning-based system for personality prediction using the Big Five (OCEAN) personality model, aimed at enhancing recruitment and self-assessment processes. By applying algorithms like KNN, CNN, and Logistic Regression, the system classifies personality traits and evaluates technical proficiency. The model incorporates a phrase frequency algorithm to strengthen prediction accuracy. This tool can be beneficial in competitive environments and hiring scenarios by helping users understand and improve their personality traits based on analytical feedback.[5]

3. Using Textual Data for Personality Prediction: A Machine Learning Approach, Authors: Aditi V. Kunte, Suja Panicker.

This research explores the use of machine learning techniques for predicting personality traits based on textual data from social media, specifically Twitter. By eliminating the need for traditional personality questionnaires, the proposed approach saves time and enhances credibility. The study employs algorithms like Linear Discriminant Analysis, Multinomial Naive Bayes, and AdaBoost on a standard Twitter dataset. The work demonstrates the feasibility of leveraging large-scale social media text to classify individuals into personality categories, making it valuable for applications in hiring, mental health analysis, and personalized marketing.[6]

4. Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process, Authors: G. Sudha, Sasipriya K K, Sri Janani S, Nivethitha D, Saranya S, Karthick Thyagesh G.

This paper presents a machine learning-based system designed to automate the e-recruitment process by analyzing candidates' resumes and personality traits. The system leverages logistic regression to match job seekers with company requirements by evaluating both professional qualifications from the uploaded CV and personality traits through an online MCQ-based quiz. By integrating the Job Characteristics Model (JCM) into a digital HR framework, the proposed approach enhances the efficiency, fairness, and accuracy of recruitment decisions, helping organizations streamline candidate selection in the digital age.[7]

5. Machine Learning based Two Stage Personality Prediction System, Author: Suja Panickar.

This paper introduces a two-stage machine learning system for predicting negative personality traits based on behavioral analysis using textual inputs from social media, specifically Twitter. By combining Natural Language Processing techniques with classifiers such as Multinomial Naïve Bayes, SVM, kNN, and AdaBoost, the system effectively analyzes user-generated text to identify personality traits without relying on traditional questionnaires. The approach shows promising results and offers a valuable direction for future research in psychological profiling and personality prediction through online behavior analysis.[8]

6. A Framework for Personality Prediction for E-Recruitment Using Machine Learning Algorithms, Authors: Lakshyajit Thapa, Aniket Pandey, Deepanshu Gupta, Ayush Deep, Rakesh Garg.

This paper introduces a machine learning-based framework for predicting personality in e-recruitment to improve candidate-job fit. By analyzing diverse online data sources such as resumes, social media, and psychometric test responses, the system uses psychological models like Big Five, MBTI, HEXACO, and Enneagram for personality assessment. Feature extraction is performed using modern NLP tools like BERT, and classification is handled through algorithms including SVM, XGBoost, and LSTM. The framework aims to support recruiters in identifying candidates whose personalities align with organizational needs, enhancing both hiring effectiveness and workplace productivity.[9]

7. Machine Learning for Predicting Personality Traits from Eye Tracking, Authors: Meftah Salima, Sahnoun M'hammed, Mourad Messaadia, Sidi Mohamed Benslimane.

This research presents a novel approach to predicting the Big Five personality traits using machine learning models trained on eye-tracking data. By analyzing individuals' visual behavior, the study reduces reliance on traditional self-reported questionnaires, which are often biased. The proposed system uses Random Forest, Gradient Boosting, and Extreme Gradient Boosting algorithms and shows improved accuracy in personality prediction. The results support the integration of personality insights into human-centered systems, such as smart industries and

intelligent human-machine interfaces, enabling more adaptive, personalized interactions.[10]

III. METHODOLOGY

The methodology for the Flask-based personality prediction web application centers on collecting user demographic and psychological data, preprocessing it, and utilizing advanced machine learning models—LightGBM and CatBoost—to predict personality traits. Users interact with a web interface to input features such as gender, age, and responses related to the Big Five personality traits. The system preprocesses this data by encoding categorical variables and normalizing numerical values to ensure optimal model performance. Both LightGBM and CatBoost models are trained on a large-scale dataset, with hyperparameter tuning and validation to maximize predictive accuracy. The trained models are integrated into the Flask web application, enabling real-time predictions based on user input. The application also incorporates administrative features, allowing admins to manage users, provide health tips, and respond to queries, thereby creating an interactive platform for both personality assessment and user support.

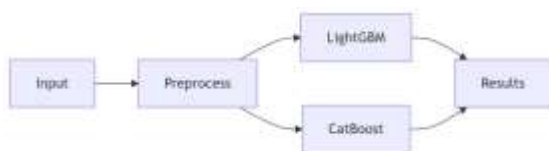


Fig 1. Proposed Methodology

1. Data Collection and Preprocessing:

User demographic and psychological information, including gender, age, and Big Five trait responses, is collected via the web interface. The data is preprocessed through encoding and normalization to prepare it for machine learning models.

2. Model Development:

LightGBM and CatBoost, both robust gradient boosting algorithms, are trained on the processed dataset. Hyperparameter tuning and validation ensure the models achieve high predictive accuracy for multi-class personality classification.

3. Web Application Integration:

The trained models are deployed within a Flask web framework. Users can input their data and receive instant personality predictions, while results from both models are displayed for comparison.

4. Administrative Features:

Admins can log in to manage users, provide health tips, and respond to queries, enhancing the platform's interactivity and user engagement.

This methodology ensures the system is accurate, scalable, and user-friendly, combining machine learning with practical web development for effective personality prediction.

IV. TECHNOLOGIES USED

The Flask-based personality prediction web application utilizes a modern technology stack that integrates machine learning, web development, and database management to deliver accurate and interactive personality assessments:

Python: The core programming language for backend development, machine learning model implementation, and data preprocessing.

Flask: A lightweight Python web framework used to build the web application, handle HTTP requests, manage user sessions, and integrate machine learning models for real-time predictions.

LightGBM: An advanced gradient boosting framework employed for efficient and scalable multi-class personality classification using decision tree algorithms.

CatBoost: A gradient boosting library optimized for handling categorical data, used alongside LightGBM for robust personality trait prediction.

HTML/CSS/JavaScript: Standard web technologies for building responsive and user-friendly front-end interfaces, enabling users to input data and view results interactively.

Jinja2: Flask's templating engine, used to dynamically render HTML pages based on user data and model predictions.

SQLite/MySQL: Relational databases used to securely store user information, prediction results, health tips, and queries for both users and administrators.

Pandas/Numpy: Python libraries for data manipulation, analysis, and preprocessing before feeding data into machine learning models.

Scikit-learn: Used for additional data preprocessing, feature encoding, and evaluation metrics.

Chart.js or Similar Visualization Libraries: For visualizing personality trait results and health tips within the web interface.

Admin Dashboard Tools: Custom Flask views and templates for administrative management, including user account management, health tip posting, and query response.

This combination of technologies ensures the application is scalable, secure, and user-friendly, providing both advanced machine learning capabilities and a seamless web experience

V Result



Fig 5.1 Traits input page.

Result page



Fig 5.2 Result page

VI. CONCLUSION

In conclusion, the Flask-based personality prediction web application successfully integrates advanced machine learning models—LightGBM and CatBoost—with an intuitive web interface to provide accurate, real-time personality assessments based on psychological and demographic features. By combining robust data preprocessing, scalable model deployment, and interactive user and administrative functionalities, the system demonstrates how data science and web development can be leveraged to deliver practical, user-friendly tools for psychological analysis and personal development. This approach not only

streamlines personality assessment but also fosters meaningful user engagement and support, making it valuable for educational, organizational, and self-improvement contexts.

REFERENCES

- [1]. Personality Prediction Using ML and NLP, Authors: Adinath Mahandule, Gaurav Fiske, Vedang Malusare, Omkar Bhambure, Rajashree Shedge, DOI: 10.1109/ICCUBEA58933.2023.10392087, Publisher: IEEE, Conference: 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, Date Added to IEEE Xplore: 22 January 2024
- [2]. Personality Prediction Using Deep Learning, Authors: Indira Dutta, R. Athilakshmi, Amulya, DOI: 10.1109/ICAECT57570.2023.10117573, Publisher: IEEE, Conference: 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, Date Added to IEEE Xplore: 15 May 2023
- [3]. Applicant Personality Prediction System Using Machine Learning, Authors: M. Karnakar, Haseeb Ur Rahman, A B Jai Santhosh, NageswaraRao Sirisala, DOI: 10.1109/GCAT52182.2021.9587693, Publisher: IEEE, Conference: 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, Date Added to IEEE Xplore: 09 November 2021
- [4]. Prediction of Personality Trait using Machine Learning on Online Texts, Authors: Raj Kishan Cherukuru, Ashish Kumar, Sumit Srivastava, Vivek Kumar Verma, DOI: 10.1109/ICONAT53423.2022.9725910, Publisher: IEEE, Conference: 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, Date Added to IEEE Xplore: 10 March 2022
- [5]. Smart-Hire Personality Prediction Using ML, Authors: Isha Gupta, Manasvi Jain, Prashant Johri, DOI: 10.1109/ICDT57929.2023.10151367, Publisher: IEEE, Conference: 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, Date Added to IEEE Xplore: 19 June 2023
- [6]. Using Textual Data for Personality Prediction: A Machine Learning Approach, Authors: Aditi V. Kunte, Suja Panicker, DOI: 10.1109/ISCON47742.2019.9036220, Publisher: IEEE, Conference: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, Date Added to IEEE Xplore: 16 March 2020
- [7]. Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process, Authors: G. Sudha, Sasipriya K K, Sri Janani S, Nivethitha D, Saranya S, Karthick Thyagesh G, DOI: 10.1109/ICCCT53315.2021.9711787, Publisher: IEEE, Conference: 2021 4th International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, Date Added to IEEE Xplore: 18 February 2022
- [8]. Machine Learning based Two Stage Personality Prediction System, Author: Suja Panickar, DOI: 10.1109/ICICT60155.2024.10544512, Publisher: IEEE, Conference: 2024 International Conference on Inventive Computation Technologies (ICICT),

Lalitpur, Nepal, Date Added to IEEE Xplore: 07 June 2024

[9]. A Framework for Personality Prediction for E-Recruitment Using Machine Learning Algorithms, Authors: Lakshyajit Thapa, Aniket Pandey, Deepanshu Gupta, Ayush Deep, Rakesh Garg, DOI: 10.1109/Confluence60223.2024.10463354,

Publisher: IEEE, Conference: 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, Date Added to IEEE Xplore: 21 March 2024

[10]. Machine Learning for Predicting Personality Traits from Eye Tracking, Authors: Meftah Salima, Sahnoun M'hammed, Mourad Messaadia, Sidi Mohamed Benslimane, DOI: 10.1109/DASA59624.2023.10286789, Publisher: IEEE, Conference: 2023 International Conference on Decision Aid Sciences and Applications (DASA), Annaba, Algeria, Date Added to IEEE Xplore: 23 October 2023