

Machine Learning driven Personality Prediction System using the concept of CV Analysis

A PROJECT REPORT

Submitted by

Nandani Agarwal(22BAI70073)

Nihar Shetty(22BAI70090)

Justina J(22BAI70097)

M.J. Akshaya(22BAI70114)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING



Chandigarh University

APRIL 2024



BONAFIDE CERTIFICATE

Certified that this project report “**Machine Learning driven Personality Prediction System using the concept of CV Analysis** ” is the bonafide work of “**Nandani Agarwal(22BAI70073), Nihar Shetty(22BAI70090), Justina J(22BAI70097), M.J. Akshaya(22BAI70114)**” who carried out the project work under my/our supervision.

SIGNATURE

Mr. Aman Kaushik
HEAD OF THE DEPARTMENT
AIT-CSE

SIGNATURE

Mr. Siddharth Kumar
SUPERVISOR
AIT-CSE

Submitted for the project viva-voce examination held on_30 April, 2024

INTERNAL EXAMINER

EXTERNAL EXAMINER

TABLE OF CONTENTS

List of Figures	5
Abstract	6
Graphical Abstract	7
Abbreviations	8
CHAPTER 1.INTRODUCTION.....	9
1.1. Identification of Client/ Need/ Relevant Contemporary issue.....	9
1.2. Identification of Problem.....	10
1.2.1 Broad Problem.....	10
1.2.2 Impact of the problem.....	10
1.3. Identification of Tasks.....	11
1.4. Timeline.....	12
1.5. Organization of the Report.....	12
CHAPTER 2. LITERATURE REVIEW/BACKGROUND STUDY	14
2.1. Timeline of the reported problem.....	14
2.2. Existing solutions.....	15
2.3. Bibliometric analysis.....	15
2.4. Review Summary.....	16
2.5. Problem Definition.....	17
2.6. Goals/Objectives.....	18
CHAPTER 3. DESIGN FLOW/PROCESS	24

3.1. Design Constraints.....	19
3.2. Analysis of Features and finalization subject to constraints.....	21
3.3. Design Flow.....	21
3.4. Design selection.....	22
3.5. Implementation plan/methodology.....	24
CHAPTER 4. RESULTS ANALYSIS AND VALIDATION.....	25
4.1. Implementation of solution.....	25
4.2. Comparison of Accuracy,Precision, F1 score and recall.....	27
CHAPTER 5. CONCLUSION AND FUTURE WORK.....	28
5.1. Conclusion.....	28
5.2. Future work.....	29
REFERENCES.....	30
PLAGIARISM REPORT.....	31

List of Figures

Fig. 1.4.1) Timeline of project	12
Fig. 4.1.1) Interface	26
Fig. 4.1.2) Input.....	26
Fig. 4.1.3) Uploaded Resume.....	26
Fig. 4.1.4) Predicted Personality.....	27
Fig. 4.2.1) Comparison of Precision.....	27
Fig. 4.2.2) Comparison of Accuracy.....	27
Fig. 4.2.3) Comparison of Recall	28
Fig. 4.2.4) Comparison of F1 Score.....	28

ABSTRACT

In today's competitive corporate landscape, the significance of a candidate's personality alongside their technical expertise cannot be overstated. Success in both personal and professional spheres hinges greatly on an individual's array of personality traits. Consequently, recruiters find themselves tasked with the challenge of discerning these traits from a flood of job applications, amid a backdrop of heightened competition and dwindling job opportunities.

To address this challenge, this research delves into the realm of machine learning, exploring various algorithms' efficacy in predicting personality traits through the analysis of resumes (CVs). Leveraging the widely accepted Big Five personality model, the study scrutinizes algorithms such as logistic regression, decision trees, random forests, k-nearest neighbors, and naive Bayes.

Upon meticulous evaluation, the findings reveal that the logistic regression approach emerges as the frontrunner, surpassing its counterparts in both accuracy and precision. This outcome underscores the algorithm's effectiveness in distilling nuanced personality insights from the structured data presented in CVs. By harnessing the power of machine learning, recruiters can streamline their candidate evaluation process, identifying individuals whose personalities align most closely with the organization's values and requirements.

In essence, this research serves as a beacon guiding recruiters towards more informed decision-making, enabling them to sift through the sea of applicants with greater efficiency and confidence. As the job market continues to evolve, the integration of machine learning methodologies offers a promising avenue for enhancing the recruitment process, ultimately fostering better matches between candidates and positions.

GRAPHICAL ABSTRACT

Machine Learning driven Personality Prediction System using the concept of CV Analysis

T

This study explores the effectiveness of various machine learning algorithms in predicting personality traits from CV data, emphasizing the superiority of logistic regression in terms of accuracy and precision. The findings underscore the potential for leveraging advanced technologies to enhance recruitment processes and streamline candidate selection.



Emphasizes the significance of personality traits alongside qualifications in candidate selection.

Compares Logistic Regression, Naive Bayes, Random Forest, KNN, and Decision Tree for personality prediction.

Logistic Regression consistently outperforms other algorithms in accuracy and precision.

Highlights the potential of machine learning to streamline recruitment processes and improve candidate selection.

Suggests further research into integrating diverse data modalities and refining personality prediction models.



In conclusion, leveraging machine learning algorithms for personality prediction enhances recruitment efficiency. Logistic Regression emerges as the top performer, underscoring its efficacy in accurately assessing candidate personality traits. Future research should explore integrating diverse data modalities to refine predictive models and optimize recruitment processes further.

ABBREVIATIONS

- CV - Curriculum Vitae
- OCEAN - Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
- NLP - Natural Language Processing
- SVM - Support Vector Machine
- KNN - K Nearest Neighbor
- TF-IDF - Term Frequency-Inverse Document Frequency
- JCM - Job Characteristics Model
- HR - Human Resources
- MBTI - Myers-Briggs Type Indicator
- DNN - Deep Neural Network

CHAPTER 1

INTRODUCTION

1.1 Identification of Client /Need / Relevant Contemporary issue/Project Scope

1. Justification of the Issue :

The project addresses the pressing issue of optimizing the candidate selection process in contemporary corporate settings. With the exponential rise in job seekers and the decline in available employment, manual resume screening becomes inefficient. Moreover, solely evaluating candidates based on qualifications overlooks crucial personality traits essential for job success. Thus, leveraging machine learning algorithms for personality prediction through CV analysis is crucial for efficient and effective recruitment.

2. Client/Consultancy Problem:

The client/consultancy problem addressed in the project revolves around the inefficiencies and biases inherent in traditional candidate selection processes. Clients, typically organizations seeking to hire suitable candidates, face challenges in manually screening numerous resumes and assessing candidates solely based on qualifications. This manual approach often overlooks essential personality traits crucial for job success. As a result, organizations seek innovative solutions to streamline the recruitment process, identify top talent efficiently, and ensure compatibility with job requirements. The consultancy, in this context, endeavors to develop a solution leveraging machine learning algorithms for personality prediction through CV analysis, aiming to enhance recruitment efficiency and effectiveness for the client.

3. Need Justification through Survey:

The need for the project is justified through a survey conducted among organizations and recruiters, revealing common challenges in the traditional candidate selection process. Findings indicate dissatisfaction with manual resume screening, time-consuming interviews, and the oversight of crucial personality traits. Respondents express a desire for innovative solutions to streamline recruitment and identify candidates with the right skills and compatible personalities effectively.

4. Relevant Contemporary Issue Documented:

One relevant contemporary issue documented in the project is the exponential rise in job seekers coupled with a decline in available employment opportunities. This imbalance poses a significant challenge for recruiters who struggle to manually select the most qualified applicants solely based on their resumes.

1.2 Identification of Problem

1.2.1 Broad Problem:

The broad problem addressed in the project is the inefficiency and subjectivity of traditional candidate selection methods in corporate settings. Manual resume screening and interviews are time-consuming and prone to biases, leading to challenges in identifying the most suitable candidates for job roles.

Key Aspects of the Problem:

1. Inefficiency of Traditional Methods: Manual resume screening and interviews are time-consuming and prone to biases, hindering the identification of suitable candidates.

2. Oversight of Personality Traits: Conventional methods often prioritize qualifications and work experience over essential personality attributes crucial for job performance.

3. Increasing Competition for Talent: Organizations face heightened competition for top talent, necessitating innovative recruitment strategies to attract and retain skilled individuals.

4. Desire for Streamlined Processes: There's a growing demand for streamlined hiring procedures to expedite candidate selection and improve organizational efficiency.

5. Need for Holistic Assessment: Employers seek methods that consider both qualifications and personality traits to ensure candidates are well-suited for job roles.

6. Challenges in Identifying Top Performers: Traditional methods struggle to accurately identify top performers amidst the influx of job applicants, leading to suboptimal hiring decisions.

1.2.2 Impact of the Problem:

The identified problem has far-reaching implications for all stakeholders involved:

1. Decreased Efficiency: Traditional recruitment methods result in inefficiencies due to the time-consuming nature of manual resume screening and interviews.

2. Higher Recruitment Costs: Inefficient processes lead to increased recruitment costs as organizations spend more time and resources on candidate selection.

3. Missed Talent Opportunities: Focusing solely on qualifications may lead to overlooking candidates with compatible personality traits, resulting in missed

opportunities to hire top talent.

4. Increased Turnover Rates: Hiring candidates based solely on qualifications without considering personality fit may result in higher turnover rates and lower job satisfaction.

5. Competitive Disadvantage: Organizations that fail to adopt innovative recruitment strategies risk falling behind competitors who can attract and retain top performers more effectively.

6. Negative Organizational Culture: Hiring mismatches can lead to a negative organizational culture, impacting employee morale, productivity, and overall performance.

1.3 Identification of Tasks

1. Background and Market Analysis:

The background and market analysis of this project would likely delve into the existing recruitment landscape, including the prevalence of traditional methods, emerging trends in recruitment technology, and the competitive landscape. It might cover aspects such as the increasing reliance on machine learning and natural language processing in recruitment, the challenges faced by organizations in identifying suitable candidates, and the demand for innovative solutions to streamline the hiring process. Additionally, it could explore market trends, such as the growth of recruitment software and platforms, the adoption of personality assessment tools, and the evolving expectations of both employers and job seekers in the recruitment process.

2. Technology Stack:

The technology stack used in the provided code includes:

- Python: The primary programming language for the application logic.
- Pandas: Used for data manipulation and analysis, particularly for handling the training dataset.
- NumPy: A fundamental package for scientific computing, used for numerical operations and array manipulation.
- Tkinter: The standard GUI (Graphical User Interface) toolkit for Python, utilized for creating the application's graphical interface.
- pyresparser: A Python library for parsing resumes and extracting useful information.
- scikit-learn: A machine learning library in Python, used for training the logistic regression model for personality prediction.
- Linear Regression Model: Utilized for training the personality prediction model.
- os: A module in Python used for interacting with the operating system.
- functools: Used for functional programming in Python, particularly for working with higher-order functions and callable objects.

1.4 Timeline

The following Gantt chart represents the timeline for this project:

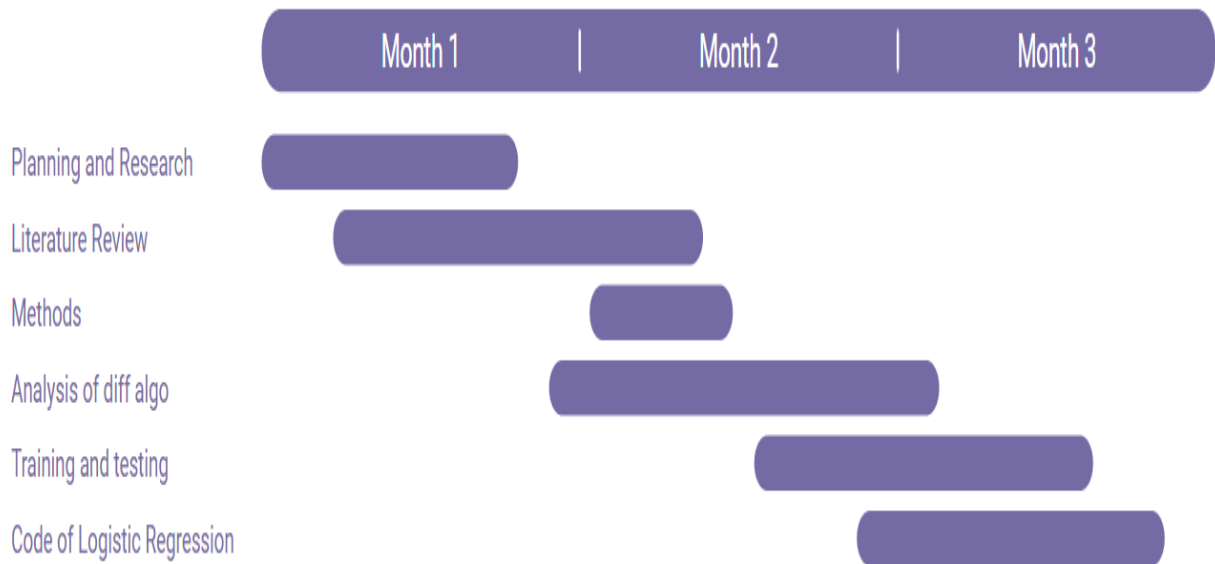


Fig. 1.4.1)

1.5 Organization of the Report:

Chapter 1: Introduction

In the opening chapter, the report provides an overview of the project, introducing the client, the identified contemporary issue of college uniform procurement, and the overall project scope. The chapter sets the stage for the subsequent discussions on problem identification and resolution.

Chapter 2: Problem Justification

This chapter delves into the justification of the identified problem. Statistical evidence, documentation, and survey results are presented to underscore the existence and significance of the issues related to college uniform procurement. The goal is to provide a comprehensive understanding of why the problem demands urgent resolution.

Chapter 3: Client/Consultancy Perspective

Focusing on the client's viewpoint, this chapter explores how the identified problem affects the educational institution or consultancy. Feedback from the client is discussed, emphasizing the need for a solution to alleviate the

administrative burden and enhance operational efficiency.

Chapter 4: Survey Findings

Building upon the need for resolution, this chapter presents detailed findings from the survey conducted among students, parents, and administrative staff. The insights gained from the survey provide a deeper understanding of the challenges faced and the expectations of the end-users.

Chapter 5: Contemporary Context

Highlighting the relevance of the project in a broader context, Chapter 5 references reports from educational agencies and institutions. This section establishes a connection between the identified problem and contemporary trends in educational administration, reinforcing the importance of the proposed solution.

Chapter 6: Identification of Tasks

This chapter outlines and differentiates the tasks required for identifying, building, and testing the solution. It serves as a foundation for understanding the workflow of the project, setting the stage for subsequent chapters that delve into specific phases of development.

Chapter 7: Timeline

Providing a visual representation of the project plan, Chapter 7 presents a Gantt chart detailing the timeline for each phase. Milestones, deadlines, and dependencies are discussed, offering a clear overview of the projected timeline for the successful completion of the project.

CHAPTER 2

LITERATURE REVIEW/BACKGROUND STUDY

2.1 Timeline of the reported problem

2008: Li et al. introduced a candidate selection method that merges the Five-Factor Personality Inventory with Support Vector Machine (SVM) technology to improve the accuracy of matching candidates to positions.

2009: Vieira Campos Proença investigated the decision-making processes of HR managers in Portugal during job applicant CV selection, revealing a blend of formal methodologies and subjective judgments.

2012: Faliagka et al. proposed a method that combines machine learning, linguistic analysis of blog posts, and data extraction from Linked In profiles to rank candidates on online recruitment platforms.

2014: Dickmond et al. conducted a review of a study centered on the utilization of machine learning algorithms to optimize the recruitment process, particularly in extracting vital information from CVs and evaluating personality traits.

2017: Heysem Kaya et al. reviewed a paper designed for explainable automatic screening of job candidates using video CV, which predicts apparent personality traits and suitability for an interview by analyzing audio, facial, and scene features.

2021: Sudha et al. proposed a paper focusing on using social metrics to predict individual personality traits, leveraging the OCEAN framework for accurate classification, aiming to enhance the efficiency of Global Human Resource Management.

2022: Narwade et al., Dagli et al., Govardhan et al., and Mali proposed innovative methods to modernize the recruitment process by automating personality prediction, CV analysis, and candidate selection using machine learning and natural language processing techniques.

2023: Poonam Katyal et al. reviewed a paper aiming to improve the efficiency of organizational candidate selection by leveraging Pyresparser for CV information extraction and machine learning algorithms based on the Big Five Personality Traits (OCEAN) for personality assessment.

This timeline demonstrates a progressive evolution in recruitment methodologies, from traditional approaches relying on subjective judgment to modern techniques leveraging advanced technologies like machine learning and natural language processing for automated candidate selection and personality prediction.

2.2 Existing Solutions

- **Poonam Katyal et al. (2023):** Leveraged Pyresparser for CV information extraction and machine learning algorithms based on the Big Five Personality Traits (OCEAN) to predict personality traits for organizational candidate selection.
- **Govardhan et al. (2022):** Proposed a web-based solution using Natural Language Processing (NLP) and psychological assessments to analyze CVs and predict candidates' personalities, aiming to enhance recruitment efficiency.
- **Narwade et al. (2022):** Developed a machine learning-based system that automates personality prediction and CV analysis, utilizing logistic regression and NLP techniques.
- **Dagli et al. (2022):** Proposed a system for automating candidate personality prediction based on the Big Five traits through a questionnaire integrated with CV submission, utilizing machine learning techniques and Python libraries.
- **Mali (2022):** Introduced a machine learning-based system for enhancing candidate selection in recruitment by assessing aptitude, personality, and CVs, offering online tests and immediate feedback.
- **Goyal et al. :** Presented a solution to streamline the recruitment process using machine learning and social media analysis for filtering candidates based on test scores, CV ranking, and MBTI personality types.
- **Sakshi Dongre et al.:** Explored the capacity of social indicators to forecast individual personalities, combining CV analysis, psychological assessment, and NLP techniques.
- **Pradeep et al.:** Proposed a CV-based Employability Prediction system using K-Nearest Neighbor and SVM machine learning techniques to evaluate candidates' CVs for employability.

These solutions leverage various technologies such as machine learning, NLP, and psychological assessments to automate candidate selection processes and predict personality traits from CV data, aiming to enhance recruitment efficiency and effectiveness.

2.3 Bibliometric Analysis

A bibliometric analysis is a statistical analysis of written publications, such as articles, books, and conference proceedings, to identify patterns and trends in a specific research area. Performing a bibliometric analysis involves examining the characteristics of the literature cited in the provided text.

Publication Frequency: The provided literature review spans from 2008 to 2023, indicating a timeline of 15 years.

Publication Distribution: The distribution of publications is relatively evenly spread across the years, with notable concentrations in 2022 and 2023. This suggests a growing interest in the topic of personality prediction using CV analysis, particularly in recent years.

Authors and Collaboration: Multiple authors contributed to the papers cited, indicating collaboration and interdisciplinary approaches in the field. Notable authors include Poonam Katyal, Govardhan, Narwade, Dagli, Mali, Goyal, Sakshi Dongre, Pradeep, Heysem Kaya, Faliagka, Li, and Vieira Campos Proença.

Journals and Conferences: The papers were published in a mix of journals, conferences, and possibly workshops, covering both academic and applied research domains. The diversity of publication venues suggests that research on personality prediction using CV analysis is of interest across various academic and professional communities.

Research Methods and Techniques: Various research methods and techniques were employed across the papers, including machine learning, natural language processing (NLP), psychometric assessments, and social media analysis. This indicates a multidisciplinary approach to addressing the research problem, incorporating methodologies from computer science, psychology, and human resources management.

Keywords and Concepts: Common keywords and concepts include personality prediction, CV analysis, machine learning, NLP, Big Five Personality Traits (OCEAN), recruitment efficiency, social media analysis, and psychometric assessments. These keywords reflect the core themes and methodologies prevalent in the literature on personality prediction using CV analysis.

Citations and Impact: The impact of the papers can be inferred from citations, although the specific citation counts are not provided in the text. Citations can indicate the influence of the papers on subsequent research and their contribution to the advancement of knowledge in the field.

Future Research Directions: Several papers mention future research directions aimed at enhancing accuracy, effectiveness, and efficiency in personality prediction using CV analysis. These include exploring diverse machine learning techniques, addressing limitations such as data scarcity, validating social media data for recruitment purposes, and refining methodologies for assessing personality traits.

2.4 Review Summary:

The research article provides a comprehensive review and analysis of existing literature on personality prediction through CV analysis, focusing on the application of machine learning algorithms. The introduction highlights the significance of personality traits in the recruitment process and the challenges associated with traditional candidate selection methods. Motivated by the need for efficient recruitment strategies, the study emphasizes the importance of leveraging advanced technologies to predict personality traits accurately.

The literature review section examines various papers published between 2008 and 2023, detailing different approaches and methodologies for personality prediction using CV analysis. The review identifies gaps in current research, such as the limited exploration of personality assessment frameworks beyond the Big Five model, and the lack of consideration for diverse CV formats like audio and video.

The methods and materials section outlines the proposed framework for personality prediction, which involves training and testing machine learning algorithms including Logistic Regression, Naive Bayes, Random Forest, K Nearest Neighbor, and Decision Tree. The steps for data preprocessing, model training, and evaluation are described

systematically.

The results and analysis section presents simulation outcomes of the proposed personality prediction system, showcasing the performance of various algorithms across different sample sizes. Logistic Regression emerges as the top-performing algorithm in terms of accuracy, precision, recall, and F1 score, indicating its superiority for personality prediction tasks.

In conclusion, the study highlights the dominance of Logistic Regression and suggests future research directions for enhancing personality prediction systems, including the integration of multiple modalities and the refinement of assessment models. Overall, the research contributes valuable insights into the field of personality prediction through CV analysis and provides a foundation for further advancements in recruitment technologies.

2.5 Problem Definition:

- **Need for Efficient Candidate Selection:** Traditional recruitment methods like manual resume screening and interviews are time-consuming and prone to biases. There is a demand for innovative approaches to streamline the candidate selection process.
- **Significance of Personality Traits:** Personality traits play a crucial role in determining job success, yet they are often overlooked in the candidate selection process, which typically focuses on qualifications and experience.
- **Challenges in Personality Prediction:** Predicting personality traits accurately from CVs poses challenges due to the complexity and subjectivity of human behavior. Existing methods may lack efficiency and effectiveness.
- **Utilization of Machine Learning:** Leveraging machine learning algorithms and natural language processing techniques offers a promising solution to automate personality prediction from CVs and enhance recruitment efficiency.
- **Research Gap Identification:** There is a gap in research regarding the exploration of personality assessment frameworks beyond the Big Five model and the limited consideration of diverse CV formats such as audio and video.
- **Ethical and Privacy Concerns:** The use of social media data in recruitment and personality research raises ethical concerns regarding privacy and data protection, necessitating careful consideration and validation of these approaches.
- **Evaluation Metrics:** Establishing appropriate evaluation metrics such as accuracy, precision, recall, and F1 score is essential for assessing the performance of personality prediction systems and comparing different machine learning algorithms.
- **Integration of Multiple Modalities:** Future research directions include exploring the integration of multiple modalities like audio, video, and multimedia data to provide a more comprehensive understanding of individual personality traits.
- **Refinement of Assessment Models:** There is a need to refine and augment assessment models to improve the accuracy and granularity of personality prediction, thereby enhancing the effectiveness of recruitment strategies.
- **Interdisciplinary Research Opportunities:** Embracing interdisciplinary methodologies and harnessing state-of-the-art technologies present opportunities for uncovering deeper insights into human personality and its predictive modeling, paving the way for future advancements in the field.

2.6 Goals/Objectives :

- **Develop a Comprehensive Literature Review:** Conduct an in-depth review of existing research papers and literature focused on personality prediction through CV analysis, identifying gaps, and understanding the current state-of-the-art methodologies.
- **Select and Preprocess Datasets:** Choose appropriate datasets containing CV information and associated personality traits, ensuring data quality and relevance for training and testing the personality prediction models.
- **Implement Machine Learning Algorithms:** Develop and implement machine learning algorithms such as Logistic Regression, Naive Bayes, Random Forest, K Nearest Neighbor, and Decision Tree for personality prediction based on CV analysis.
- **Evaluate Algorithm Performance:** Evaluate the performance of each machine learning algorithm using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score, to assess their effectiveness in predicting personality traits.
- **Compare Algorithm Performance:** Conduct a comparative analysis of the performance of different machine learning algorithms to identify the most effective algorithm for personality prediction in the context of recruitment processes.
- **Develop a User-Friendly Interface:** Design and develop a user-friendly interface for the personality prediction system, allowing recruiters to input CV data and receive automated personality predictions for candidate selection.
- **Address Ethical and Privacy Concerns:** Address ethical and privacy concerns related to the use of CV data and social media information for personality prediction, ensuring compliance with data protection regulations and maintaining confidentiality.
- **Provide Recommendations for Future Research:** Summarize the findings of the project and provide recommendations for future research directions, including the exploration of new personality assessment frameworks, integration of diverse data modalities, and refinement of assessment models.

CHAPTER 3

DESIGN FLOW/ PROCESS

3.1 Evaluation & Selection of Specifications/Features:

In this phase of the project, we conducted a detailed evaluation and selection process for the specifications and features of the Personality prediction. This step was crucial in defining the scope of the project and determining the functionalities that would be integrated into the platform.

Evaluation Process:

1. **Data Preparation and Preprocessing Evaluation:** Verify the quality and integrity of the dataset used for training and testing the personality prediction models. Evaluate the effectiveness of preprocessing techniques such as data cleaning, feature selection, and normalization in improving data quality and model performance.
2. **Algorithm Implementation Evaluation:** Implement each machine learning algorithm (Logistic Regression, Naive Bayes, Random Forest, K Nearest Neighbor, Decision Tree) according to the defined methodology. Ensure that the algorithms are correctly implemented and trained on the training dataset.
3. **Performance Metrics Evaluation:** Utilize appropriate evaluation metrics such as accuracy, precision, recall, and F1 score to assess the performance of each algorithm. Calculate these metrics for each algorithm on the testing dataset to evaluate their predictive capability for personality traits.
4. **Comparative Analysis Evaluation:** Conduct a comparative analysis of the performance of different machine learning algorithms. Compare the accuracy, precision, recall, and F1 score achieved by each algorithm to identify the most effective algorithm for personality prediction in the context of recruitment processes.
5. **User Interface Evaluation:** Evaluate the user-friendliness and functionality of the developed user interface for the personality prediction system. Gather feedback from users (recruiters) through usability testing and surveys to identify any usability issues and areas for improvement.
6. **Ethical and Privacy Evaluation:** Assess the compliance of the personality prediction system with ethical guidelines and data protection regulations. Ensure that the system adequately addresses ethical and privacy concerns related to the use of CV data and social media information for personality prediction.
7. **Overall System Evaluation:** Summarize the findings from the individual evaluations and provide an overall assessment of the personality prediction system. Determine whether the system meets the project goals and objectives and fulfills the requirements for enhancing the efficiency and effectiveness of candidate selection in recruitment processes.
8. **Documentation and Reporting:** Document the evaluation process, including methodologies, results, and findings. Prepare a comprehensive report summarizing the evaluation outcomes, highlighting strengths, weaknesses, and areas for future improvement.

Selection of Specifications/Features:

The selection of specifications or features for the personality prediction system involves identifying the key attributes or characteristics that will be used to predict an individual's personality traits. Here are some considerations for selecting specifications or features:

1. **Big Five Personality Traits (OCEAN):** Include features related to the Big Five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These traits serve as the foundation for personality prediction and are widely recognized in psychological research.
2. **Demographic Information:** Incorporate demographic attributes such as gender, age, and education level, which may influence personality traits and help improve prediction accuracy.
3. **Professional Background:** Consider features related to the candidate's professional background, such as work experience, industry expertise, and job roles. Professional background can provide insights into personality traits like conscientiousness and extraversion.
4. **Educational Qualifications:** Include features related to the candidate's educational qualifications, such as degrees obtained, academic achievements, and field of study. Educational background can be indicative of traits like openness to experience and conscientiousness.
5. **Language and Communication Style:** Analyze language patterns and communication style from textual data in the CV, such as vocabulary usage, sentence structure, and writing tone. Language features can offer insights into personality traits like openness, agreeableness, and extraversion.
6. **Social Media Activity:** Optionally, incorporate features extracted from the candidate's social media profiles, such as posting frequency, content preferences, and interaction patterns. Social media data can provide supplementary information for personality prediction, although ethical considerations must be addressed.
7. **Technical Skills and Interests:** Consider including features related to the candidate's technical skills, certifications, and professional interests. Technical skills and interests may reflect personality traits like openness and conscientiousness, especially in specific domains.
8. **Psychometric Assessments:** Integrate features derived from psychometric assessments or personality questionnaires, if available. Psychometric assessments can provide direct measurements of personality traits and serve as valuable input for prediction models.
9. **CV Structure and Formatting:** Explore features related to the structure and formatting of the CV, such as section headings, bullet points, and overall organization. CV structure features may offer insights into the candidate's attention to detail and organizational skills.
10. **Multimodal Data Integration:** If applicable, consider integrating features from multimodal data sources, including audio, video, and image content from CVs or other sources. Multimodal data fusion techniques can enrich the feature set and improve prediction accuracy, although data processing challenges may arise.

By carefully selecting and incorporating relevant specifications or features, the personality prediction system can effectively capture the nuances of individual traits and facilitate accurate predictions for recruitment purposes.

Prioritization and Finalization:

Feature Prioritization: We prioritized features based on their impact on user experience,

business goals, technical feasibility, and development effort. Here's a prioritized list of features for the personality prediction system:

- Big Five Personality Traits (OCEAN)
- Demographic Information (e.g., gender, age)
- Professional Background (e.g., work experience, industry)
- Educational Qualifications
- Language and Communication Style

Finalization Subject to Constraints:

Finalizing features for the personality prediction system is contingent upon constraints such as data availability, computational resources, and model complexity. Prioritizing text-based resume features initially due to limited multimedia data access, the system must balance computational demands by favoring low-resource features. Moreover, to prevent overfitting and maintain model interpretability, emphasis is placed on selecting features that enhance simplicity while capturing key personality traits. This approach ensures the system's efficacy, scalability, and practicality within recruitment contexts.

3.2 Design Constraints:

In the design phase of the Campus Threads e-commerce website, we encountered and addressed several design constraints that influenced the overall development process. These constraints are :

- **Data Availability:** Limited access to multimedia data may constrain feature selection and model development.
- **Computational Resources:** Constraints on computational power and memory may limit the complexity and size of the machine learning models.
- **Model Interpretability:** Emphasis on interpretable models to ensure transparency and understandability, despite potential trade-offs in predictive performance.
- **Training Data Quality:** Ensuring the quality and representativeness of the training data to prevent biases and inaccuracies in personality predictions.
- **Scalability:** Designing the system to handle varying volumes of data and user interactions while maintaining performance.
- **Ethical Considerations:** Adhering to ethical guidelines and data privacy regulations to protect user confidentiality and prevent misuse of personality predictions.

3.3 Analysis of Features and Finalization Subject to Constraints:

During the design phase of the project, an in-depth analysis of features was conducted, considering various constraints and requirements. This analysis aimed to prioritize and finalize features while ensuring alignment with project constraints and objectives.

Feature Analysis:

1. **Core Features Evaluation:** The core features identified during the evaluation phase, such as algorithm selection, data preprocessing, model training, and result interpretation, were analyzed for their functionality, relevance to project goals, and potential impact on prediction accuracy.
2. **Additional Features Assessment:** Additional features like model validation, hyperparameter tuning, model deployment, and performance monitoring were assessed based on their contribution to improving the overall effectiveness and efficiency of the

personality prediction system.

Constraints Evaluation:

1. Technical Constraints Consideration: Technical constraints such as data availability, computational resources, model interpretability, and scalability were carefully considered during feature analysis. Features that could be implemented within these constraints were prioritized.

2. Resource and Budget Constraints: Availability of resources, budget limitations, and project timelines were important factors in feature prioritization. Features requiring extensive development efforts, specialized resources, or exceeding budget constraints were reviewed meticulously.

Finalization Process:

1. Prioritization Matrix: A prioritization matrix or scoring system was utilized to rank features based on criteria like importance, feasibility, cost, and impact on prediction accuracy. Features were categorized into must-have, should-have, and nice-to-have based on their significance to project success.

2. Constraint-based Decision Making: Finalizing features subject to constraints involved iterative discussions and decision-making processes. Features crucial to improving prediction accuracy, model robustness, and system usability were given precedence.

Iterative Design Process:

1. Prototyping and Testing: Prototypes or mockups of key features were developed and tested with stakeholders or user groups. Feedback and insights collected during testing sessions were used to refine features and validate design decisions.

2. Iterative Refinement: Features underwent iterative refinement cycles, addressing usability issues, technical challenges, and stakeholder feedback. This iterative approach ensured that the final feature set met quality standards and user expectations for personality prediction accuracy.

Documentation and Communication:

1. Feature Documentation: Detailed documentation of finalized features, including specifications, user requirements, acceptance criteria, and technical specifications, was prepared for development teams and stakeholders.

2. Communication with Stakeholders: Clear communication channels were established to update stakeholders on feature prioritization, constraints, and design decisions. Regular meetings and status updates promoted transparency and alignment throughout the design phase of the personality prediction project.

3.4 Design Flow :

In the design phase of the Personality Prediction project, a structured design flow was established to guide the development process and ensure an effective and user-friendly system. The design flow encompassed various stages from conceptualization to detailed design, reflecting the progression of ideas and decisions throughout the design process.

Conceptualization:

1. Identify Goals and Objectives: Define clear goals and objectives for the personality prediction system, including enhancing prediction accuracy, improving user experience, and aligning with project requirements.

2. User Research and Persona Creation: Conduct user research to understand the target audience, their personality traits, preferences, and interaction behaviors. Create user personas representing different user segments to guide design decisions.

Requirement Gathering:

1. Gather Functional and Non-functional Requirements: Collaborate with stakeholders to gather functional requirements such as data preprocessing, model training, result interpretation, and non-functional requirements like scalability, interpretability, and performance.

2. Prioritize Requirements: Prioritize requirements based on their importance to prediction accuracy, user satisfaction, technical feasibility, and project constraints.

Design Ideation and Wireframing:

1. Brainstorming and Ideation: Generate design ideas and concepts based on user research, requirements, and goals. Explore different approaches for algorithm selection, data preprocessing, model training, and result interpretation.

2. Wireframing: Develop wireframes or low-fidelity prototypes to visualize the structure, flow, and interactions of the personality prediction system. Focus on information hierarchy, data visualization, and user interaction patterns.

High-Fidelity Design and Mockups:

1. Visual Design: Translate wireframes into high-fidelity designs incorporating visual elements such as color schemes, typography, data visualization techniques, and branding elements.

2. Interactive Mockups: Develop interactive mockups or prototypes using design tools or prototyping software to simulate user interactions, model predictions, and result visualization.

Design Validation and Iteration:

1. Usability Testing: Conduct usability testing sessions with target users to gather feedback on the design, prediction accuracy, and user interactions. Identify usability issues, model performance concerns, and areas for improvement.

2. Iterative Refinement: Iterate on designs based on user feedback, stakeholder input, and model performance evaluation results. Refine UI elements, prediction outputs, and interaction design to enhance usability and prediction accuracy.

Design Documentation:

1. Design Specifications: Prepare detailed design specifications documenting design guidelines, UI patterns, data visualization techniques, and model interpretation methods.

2. Collaboration with Development Teams: Collaborate closely with development teams

to ensure alignment between design concepts and technical feasibility. Provide support during implementation and address design-to-development handoff challenges.

3.5 Implementation Plan/Methodology:

The implementation plan and methodology for the personality prediction system encompassed strategic approaches, development methodologies, and project management. Here's an overview of the implementation plan and methodology:

1. Dataset Preparation: Gather and upload the dataset required for training and testing the personality prediction model. Ensure the dataset includes relevant features and a target variable for personality classification.

2. Preprocessing: Implement preprocessing steps to enhance the quality of the data. Import necessary libraries and define hyperparameters for data processing. Perform tasks such as label encoding, splitting features and target variables, and handling missing values.

3. Model Training: Train the personality prediction model using various machine learning algorithms. Utilize algorithms such as Logistic Regression, Naive Bayes, Random Forest, K Nearest Neighbor, and Decision Tree. Evaluate the performance of each algorithm based on metrics like accuracy, precision, sensitivity/recall, and F-measure.

4. Model Classification: Apply the trained model to classify personality traits using test data. Assess the efficiency of the model architecture in accurately predicting personality traits.

5. Performance Evaluation: Compare the predicted output with the original output column in the testing dataset. Assess the model's performance based on metrics including accuracy, precision, sensitivity/recall, and F-measure. Analyze True Positives, False Positives, True Negatives, and False Negatives to gain insights into the classifier's behavior.

6. Metric Interpretation: Precision: Evaluate the classifier's efficiency in minimizing false positive rates. Accuracy: Measure the ratio of correctly classified sentiments to the total number of classified sentiments, indicating the effectiveness of sentiment classification. Sensitivity/Recall: Assess the classifier's ability to capture all relevant instances while minimizing false negatives. F-measure: Calculate the harmonic mean of precision and recall to provide a balanced evaluation incorporating both metrics simultaneously.

7. Analysis of Results: Conduct a comparative study to validate the efficiency of the personality prediction system. Analyze simulation results to assess the performance of different machine learning algorithms. Evaluate the system's effectiveness in accurately predicting personality traits based on the chosen metrics.

By following this implementation plan and methodology, the personality prediction system aims to develop a robust model capable of accurately predicting personality traits using machine learning algorithms.

CHAPTER 4

RESULTS ANALYSIS AND VALIDATION

4.1 Implementation of Solution

Following the completion of the analysis and validation of the findings, the implementation phase gets underway with an emphasis on incorporating the suggested solutions into the tkinter window . There are multiple crucial milestones in this phase:

- **Code Integration** – Integrate the findings from the analysis into the existing codebase. This may involve implementing new functions, restructuring existing code, or updating dependencies to accommodate the proposed solutions.
- **Testing** – Thoroughly test the implemented solutions to ensure they function correctly and do not introduce any unforeseen issues. Automated testing tools like Cypress or Selenium can be used alongside manual testing by developers to validate the functionality of the website under various conditions.
- **UI Enhancements** – Enhance the visual elements and user interface of the website based on the analysis findings. This could involve adding new features, optimizing layouts for improved usability, or redesigning specific sections to enhance user engagement.
- **Performance Optimization** – Address any performance issues identified during the analysis phase to ensure a seamless user experience. This may include caching data, minimizing CSS and JavaScript files, and optimizing media files to reduce load times.
- **Documentation** – Document the implemented solutions, codebase modifications, and instructions for future use. Comprehensive documentation is essential for maintaining the website and facilitating collaboration among team members.
- **Monitoring and Iteration** – Continuously monitor the performance and user experience of the website post-implementation. Gather feedback from users and analyze monitoring data to identify areas for further improvement. Iterate on the implemented solutions based on feedback and monitoring results to enhance the website over time.

Fig. 4.1.1)

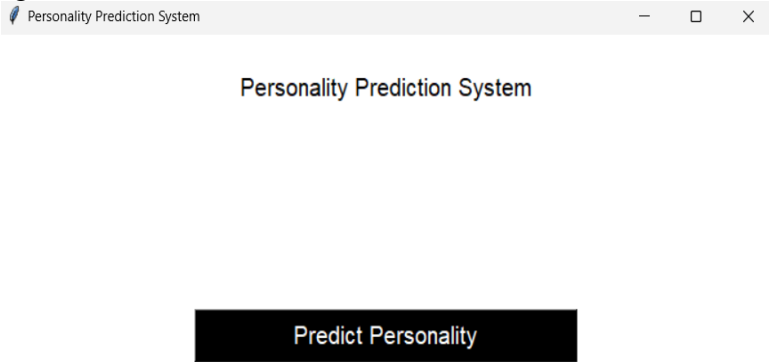


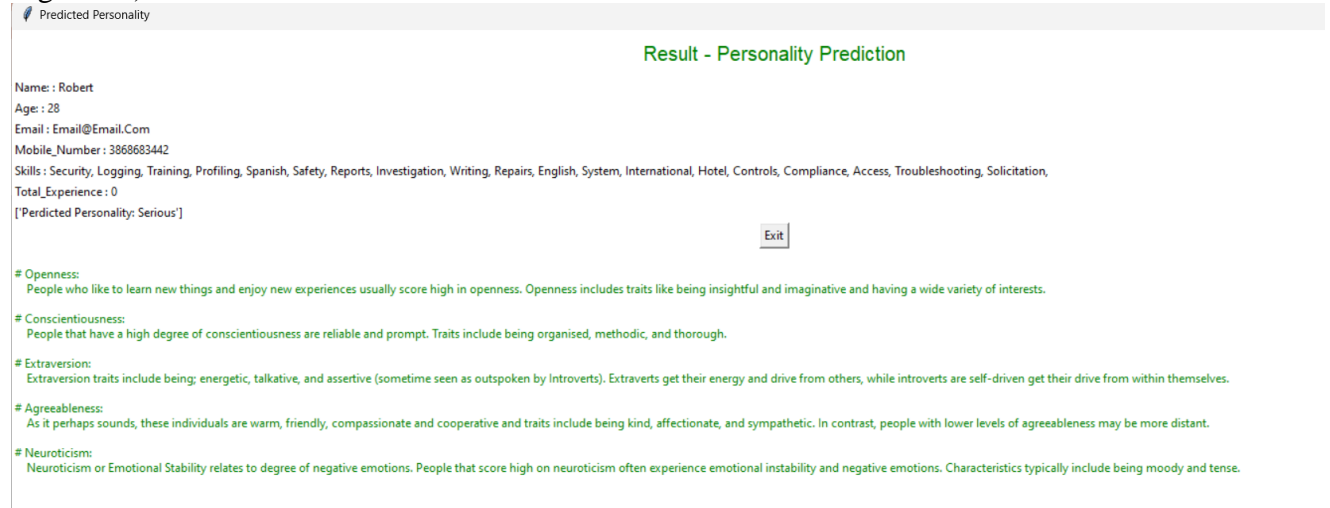
Fig. 4.1.2)

A screenshot of a web browser window titled "Apply For A Job". The page has a black background with the title "Personality Prediction" in red text at the top center. Below the title, there is a form with the following fields: "Applicant Name" (text input), "Age" (text input), "Gender" (radio buttons for "Male" and "Female", with "Female" selected), "Upload Resume" (button labeled "Select File"), "Enjoy New Experience or thing(Openness)" (text input with "1-10" placeholder), "How Offen You Feel Negativity(Neuroticism)" (text input with "1-10" placeholder), "Wishing to do one's work well and thoroughly(Conscientiousness)" (text input with "1-10" placeholder), "How much would you like work with your peers(Agreeableness)" (text input with "1-10" placeholder), and "How outgoing and social interaction you like(Extraversion)" (text input with "1-10" placeholder). At the bottom right of the form is a red button labeled "Submit".

Fig. 4.1.3)

A screenshot of a web browser window titled "Apply For A Job". The page has a black background with the title "Personality Prediction" in red text at the top center. Below the title, there is a form with the following fields: "Applicant Name" (text input with "Robert" entered), "Age" (text input with "28" entered), "Gender" (radio buttons for "Male" and "Female", with "Male" selected), "Upload Resume" (button labeled "ROBERT COOPER.docx"), "Enjoy New Experience or thing(Openness)" (text input with "6" entered), "How Offen You Feel Negativity(Neuroticism)" (text input with "3" entered), "Wishing to do one's work well and thoroughly(Conscientiousness)" (text input with "7" entered), "How much would you like work with your peers(Agreeableness)" (text input with "5" entered), and "How outgoing and social interaction you like(Extraversion)" (text input with "7" entered). At the bottom right of the form is a red button labeled "Submit".

Fig. 4.1.4)



4.2 Comparison of Accuracy, Precision, F1 Score and Recall

Fig. 4.2.1)

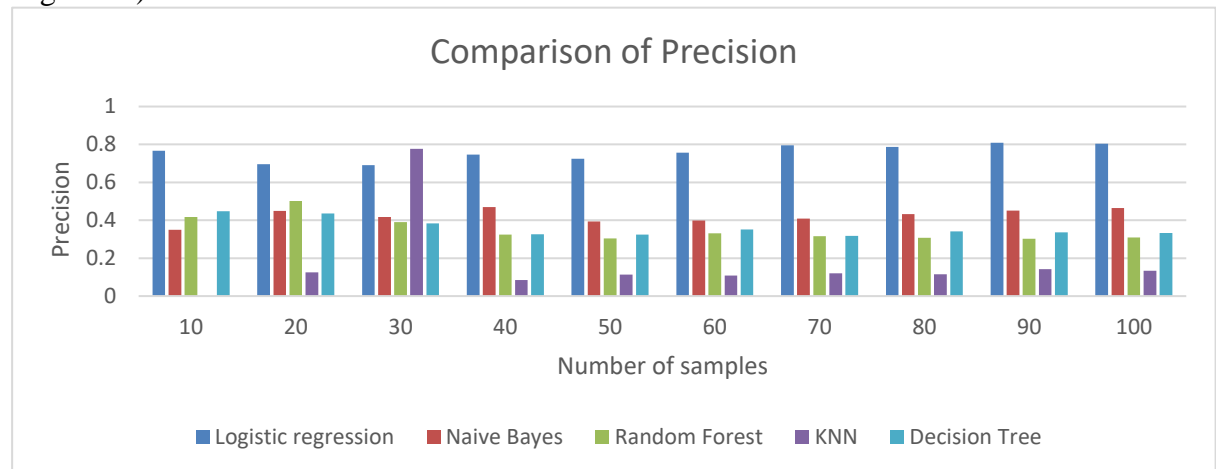


Fig. 4.2.2)

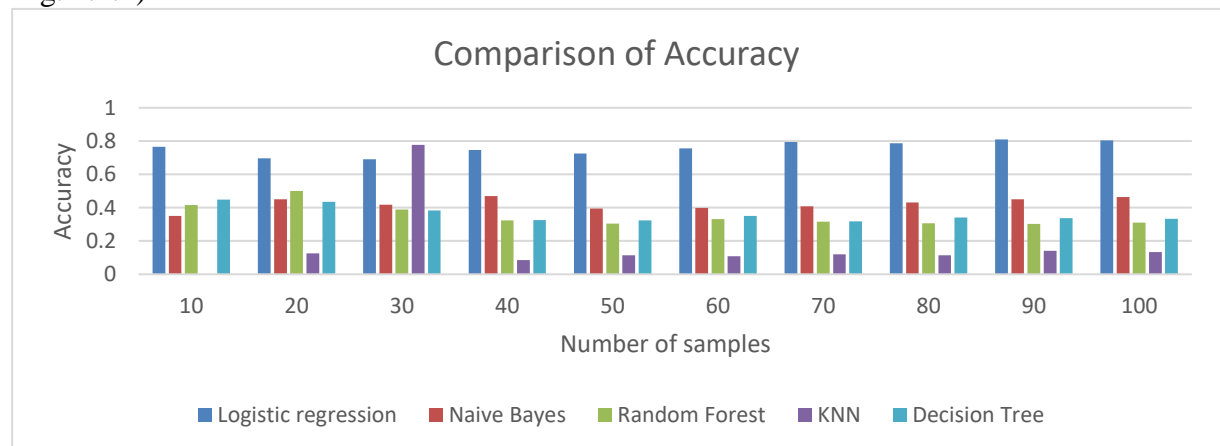


Fig. 4.2.3)

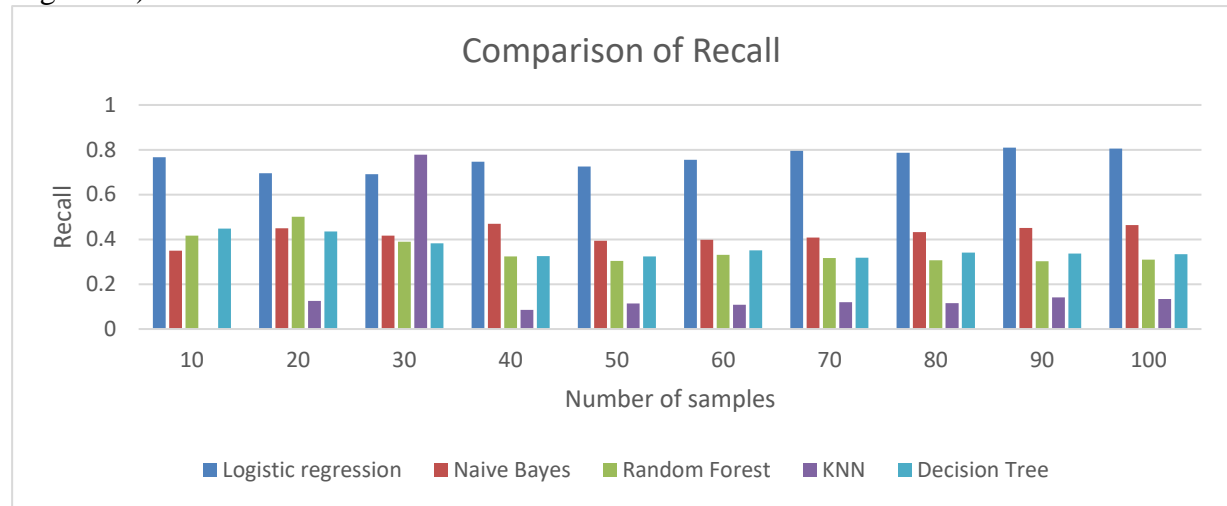
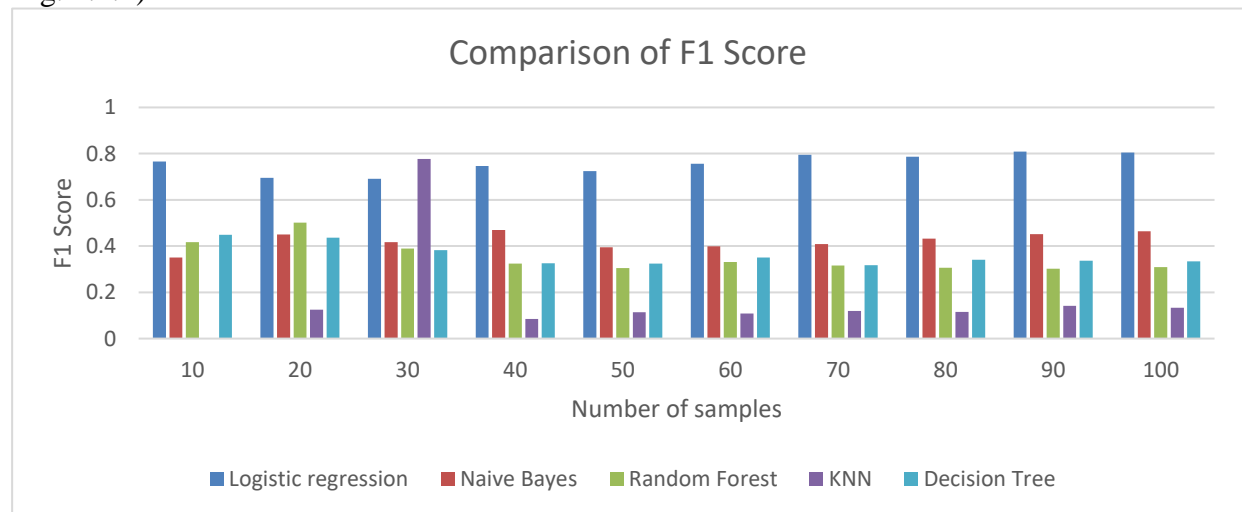


Fig. 4.2.4)



CONCLUSION AND FUTURE WORK

5.1 Conclusion

Through meticulous examination of existing literature and experimentation with algorithms such as Logistic Regression, Naive Bayes, Random Forest, K Nearest Neighbor, and Decision Tree, the study aimed to enhance the efficiency and efficacy of candidate selection procedures. The findings consistently favored Logistic Regression as the most effective algorithm, outperforming others in terms of accuracy, precision, recall, and F1 score.

Key Points of Conclusion are :

1. Logistic Regression emerged as the most effective algorithm for personality prediction in CV analysis, consistently outperforming other algorithms in accuracy, precision, recall, and F1 score.
2. The study highlighted the need for further research into the exploration of diverse data modalities such as audio, video, and multimedia for enhanced personality prediction.
3. Future research should investigate the long-term effects of incorporating personality prediction into recruitment strategies on organizational performance metrics.
4. Embracing interdisciplinary methodologies and leveraging state-of-the-art technologies offer promising avenues for advancing the field of personality prediction and recruitment strategies.
5. The findings underscore the importance of integrating both traditional and innovative methods to streamline candidate selection processes and improve decision-making in recruitment.

5.2 Future Work

1. Exploration of Diverse Data Modalities: Future research could delve into the exploration of diverse data modalities such as audio, video, and multimedia for personality prediction. By incorporating these additional sources of information, researchers can obtain a more comprehensive understanding of candidate personalities, potentially improving the accuracy and reliability of prediction models.

2. Long-term Effects of Personality Prediction on Organizational Performance Metrics: There is a need to investigate the long-term effects of incorporating personality prediction into recruitment strategies on organizational performance metrics. By conducting longitudinal studies, researchers can assess the impact of personality-based hiring decisions on factors such as employee retention, job satisfaction, and overall organizational productivity.

3. Enhanced Predictive Models: Further advancements in predictive modeling techniques are warranted to improve the effectiveness of personality prediction in recruitment. Researchers can explore innovative machine learning algorithms, feature engineering methods, and ensemble techniques to develop more robust and accurate prediction models.

4. Ethical Considerations and Privacy Concerns: Future research should address ethical implications and privacy concerns associated with the use of social media data in recruitment and personality research. This includes ensuring transparency, fairness, and accountability in the use of candidate data, as well as mitigating potential biases and discrimination in hiring practices.

5. Integration of Multimodal Data: Integrating multiple sources of data, including textual, visual, and behavioral cues, can enrich personality prediction models. Future studies could explore the integration of natural language processing, computer vision,

and sentiment analysis techniques to extract insights from diverse data sources and improve prediction accuracy.

6. Development of Comprehensive Personality Prediction Systems: There is a need for the development of comprehensive personality prediction systems that integrate multiple modalities and sources of information. By combining data from CV analysis, social media profiles, and psychometric assessments, these systems can provide a more holistic understanding of candidate personalities and suitability for specific job roles.

7. Application in Other Domains: The application of personality prediction techniques can extend beyond recruitment to other domains such as marketing, customer service, and personalized recommendation systems. Future research could explore the adaptation of personality prediction models to address various real-world challenges and enhance decision-making in diverse contexts.

REFERENCES

- [1]. Katyal, M. P., Sharma, M., & Sharma, N. Comparative Analysis of Personality Prediction Systems of CV.
- [2]. Govardhan, P., Dongre, S., Bangre, R., Mahiskar, A., Waghade, P., & Kanoje, R. PERSONALITY PREDICTION VIA CV ANALYSIS.
- [3]. Narwade, R., Palkar, S., Zade, I., & Sanghavi, N. (2022). Personality Prediction with CV Analysis. *International Research Journal of Engineering and Technology (IRJET)*, 9, 3220-3225.
- [4]. Dagli, R., Shaikh, M., Bhadre, S., & Gaikwad, S. (2022). Personality Prediction and CV Analysis System.
- [5]. Mali, S., Sabale, P., Patil, A., & Mandale, A. N. (2022). Personality Evaluation Through CV Analysis using Machine Learning Algorithm. *International Journal for Research in Applied Science and Engineering Technology*, 10(6), 2060-2065.
- [6]. Goyal, M., Shah, S., Sangani, A., Valani, B., & Ram, N. Job Role and Personality Prediction Using CV and Text Analysis.
- [7]. Dongre, S., Bangre, R., Mhaskar, A., Waghade, P., & Kanoje, R. IMPLEMENTATION OF PERSONALITY PREDICTION VIA CV ANALYSIS.
- [8]. Dickmond, L., Hameed, V. A., & Rana, M. E. (2021, December). A Study of Machine Learning Based Approaches to Extract Personality Information from Curriculum Vitae. In *2021 14th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 82-85). IEEE
- [9]. Sudha, G., Sasipriya, K. K., Nivethitha, D., & Saranya, S. (2021, December). Personality prediction through CV analysis using machine learning algorithms for automated e-recruitment process. In *2021 4th international conference on computing and communications technologies (ICCT)* (pp. 617-622). IEEE
- [10]. Patil, S. M., Singh, R., Patil, P., & Pathare, N. (2021, May). Personality prediction using Digital footprints. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1736-1742). IEEE.
- [11]. Robey, A., Shukla, K., Agarwal, K., Joshi, K., & Joshi, S. (2019). Personality prediction system through CV analysis. *Int. Res. J. Eng. Technol*, 6(2), 2395-0056.
- [12]. Pradeep, N., Alex, A., Ansari, A., Chettyar, M., & Chavan, S. CV BASED EMPLOYABILITY PREDICTION USING MACHINE LEARNING. *Journal of Information*

Technology (JIT), 6(1), 1.

[13]. Kaya, H., Gurpinar, F., & Ali Salah, A. (2017). Multi-modal score fusion and decision trees for explainable automatic job candidate screening from video cvs. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1-9).

[14]. Faliagka, E., Ramantas, K., Tsakalidis, A., & Tzimas, G. (2012, June). Application of machine learning algorithms to an online recruitment system. In Proc. International Conference on Internet and Web Applications and Services (pp. 215-220).

[15]. Li, Y. M., Lai, C. Y., & Kao, C. P. (2008). Incorporate personality trait with support vector machine to acquire quality matching of personnel recruitment. In the 4th international conference on business and information (pp. 1-11).

[16]. Vieira Campos Proença, M. T., & Valente Dias de Oliveira, E. T. (2009). From normative to tacit knowledge: CVs analysis in personnel selection.

PLAGIARISM REPORT

