



PREDICTION OF HUMAN BEHAVIOUR BASED ON CV

¹Arnold Joshua Sundar, ²Yash Solanki, ³Sonal Tiwari and ⁴Savita Sangam

^{1,2,3}Scholar, ⁴Professor

Department of Information Technology,
SSJCOE, Dombivli, India

Abstract: This study investigates the feasibility of predicting human behavior based on CV analysis. Using resume data, including education, work experience and skill sets, predictive models are developed to predict future behavior in a variety of contexts. Using machine learning techniques such as natural language processing and predictive analytics, this research aims to uncover patterns and correlations between recovery attributes and subsequent actions or decisions. Predictive models generated from this analysis offer valuable insights into potential career trajectories, job performance, and other behavioral outcomes. By leveraging resume data, organizations can make informed decisions regarding talent acquisition, workforce planning, and human resource management strategies. This study contributes to the advancement of our understanding of the prediction of human behavior and demonstrates the potential applications of CV-based analysis in various fields.

Index Terms - Machine Learning, Fare Prediction, Predictive analysis, Supervised Learning.

1. INTRODUCTION

In today's dynamic and competitive job market, understanding human behavior is pivotal for making informed decisions in recruitment, career planning, and talent management. With the advent of machine learning techniques, the analysis of Curriculum Vitae (CV) data offers a unique opportunity to delve deeper into the intricacies of human behavior. A CV serves as a comprehensive snapshot of an individual's professional journey, encapsulating their education, skills, experiences, and aspirations. By harnessing the power of advanced algorithms such as Random Forest Classifier, Logistic Regression, and Support Vector Classifier, this project aims to unlock valuable insights hidden within CVs to predict various facets of human behavior.

Moreover, resumes offer a glimpse into individuals' cognitive abilities and problem-solving skills through the articulation of achievements, projects undertaken, and challenges overcome. The manner in which candidates present their accomplishments can shed light on their capacity for innovation, collaboration, and leadership, crucial aspects for predicting future performance and behavior. The proliferation of digital platforms and online recruitment processes has led to an exponential increase in the volume and diversity of CV data available for analysis. This abundance of data presents a fertile ground for leveraging machine learning algorithms to extract meaningful patterns and trends. Through this project, we seek to bridge the gap between traditional HR practices and modern data-driven methodologies by developing predictive models that can anticipate human behavior based on CV attributes. By doing so, organizations can gain a competitive edge in talent acquisition, employee development, and workforce optimization.

Logistic Regression, on the other hand, provides interpretability and insight into the significance of individual features in predicting outcomes. Meanwhile, the Support Vector Classifier excels in handling high-dimensional data and is adept at identifying intricate decision boundaries within the feature space.

1.1 OBJECTIVES

- Use machine learning techniques to analyze historical behavioral patterns extracted from resumes to predict future job performance and cultural fit within the organization.
- Develop predictive models that use resume data to predict candidate behavior in various work scenarios, aiding in candidate selection and talent acquisition processes.
- Use Random Forest Classifier, Logistic Regression, Support Vector Classifier algorithms to identify personality traits, communication styles, and interpersonal skills from resume text, enabling better prediction of candidate behavior in team environments.
- Implement predictive analytics to predict candidates' responses to situational challenges or stressors based on past experience and qualifications listed on their resumes.

2. LITERATURE REVIEW

In recent years, there has been a growing interest in predicting human behavior based on resume data, driven by potential applications in recruitment, workforce management, and human resource analytics. Researchers have focused on using machine learning and natural language processing techniques to extract patterns and insights from resumes that can help predict individuals' future behavior and performance.

A fundamental aspect of this field revolves around how resume data is represented and processed. Various methods have been explored, including keyword extraction, semantic analysis, and topic modeling. Techniques such as word embedding and document embedding are commonly used to encode resume text into numerical representations for input into machine learning models.

The feature is another critical component in building effective predictive models from recovery data. Researchers have examined various features derived from resumes, including education, work experience, skills, certifications, and extracurricular activities. Additionally, features related to resume structure and formatting, such as section headings and font styles, were examined to capture additional information.

Several machine learning algorithms have been used to predict human behavior based on recovery data. These include logistic regression, random forests, neural networks, and ensemble methods. Each method offers unique advantages in processing different aspects of resume data and capturing underlying patterns.

Evaluating predictive models requires appropriate metrics to assess their performance. Commonly used metrics include accuracy, precision, recall, F1-score, and AUC-ROC. Researchers often use cross-validation techniques to ensure the generalizability of their models.

Predicting human behavior from resumes finds applications in a variety of fields, including recruiting, talent management, and career counseling. However, issues such as privacy concerns, biases in model predictions and the dynamic nature of labor markets need to be addressed. Future research directions may include exploring new data sources, incorporating additional contextual information, and developing interpretable models.

In conclusion, the prediction of human behavior based on CV analysis has enormous potential for improving decision-making processes in workforce management. By leveraging advances in machine learning and data analytics, researchers can develop predictive models that help employers identify suitable candidates, optimize team composition, and increase overall organizational performance. Further research is needed to address existing issues and explore new avenues for harnessing the predictive power of recovery data.

3. METHODOLOGY

Data Collection

- Sources of resume data: online job portals, company databases, recruitment agencies, etc.
- Data extraction methods: web scraping, API access, manual collection.
- Data preprocessing: cleaning, formatting, and structuring resume data for analysis.

Random Forest Classifier.

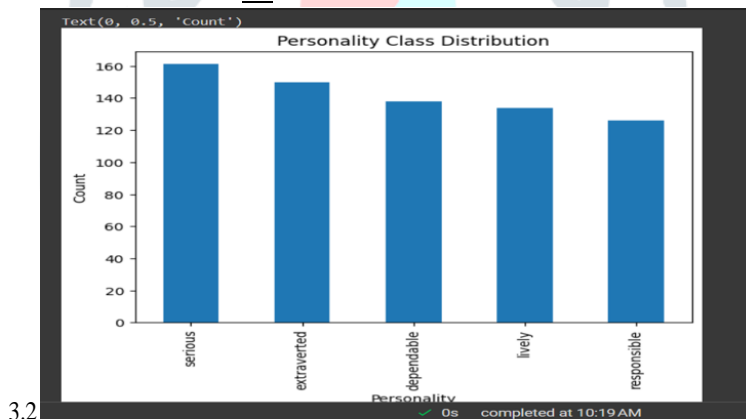
- High Accuracy
- Handles Large Datasets

Support Vector Classifier

- Effective for binary and multiclass classification tasks.
- Capable of handling high-dimensional data.

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
0	Male	17	7	4	7	3	2	extraverted
1	Male	19	4	5	4	6	6	serious
2	Female	18	7	6	4	5	5	dependable
3	Female	22	5	6	7	4	3	extraverted
4	Female	19	7	4	6	5	4	lively
	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (class label)
0	Female	20	7	9	9	5	5	dependable
1	Male	17	5	4	5	2	4	serious
2	Female	25	5	5	7	2	4	serious
3	Female	18	6	2	7	4	7	serious
4	Female	19	2	4	7	1	3	responsible

3.1: Screenshot of Dataset



3.2

Fig 3.2: Personality class distribution

3.1 Proposed Architecture

Proposing an architecture for predicting human behavior based on resumes involves several key components and steps. Firstly, the system would begin with data ingestion, where resumes in various formats are collected from job applicants. Next, a preprocessing stage would standardize and extract relevant information from the resumes, including educational background, work experience, skills, and any other pertinent details. Natural language processing (NLP) techniques would be employed to parse and interpret textual content, converting it into structured data for analysis.

Following preprocessing, feature engineering would play a crucial role in selecting and creating meaningful features from the resume data. This may involve techniques such as keyword extraction, sentiment analysis, and entity recognition to capture the essence of the applicant's qualifications, interests, and personality traits. Additionally, contextual features such as industry trends, job market demand, and organizational culture could be integrated to enrich the predictive model.

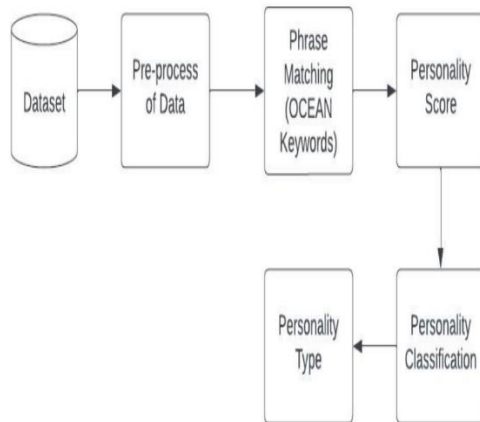


Figure 3.1.1: System Architecture Model

4. RESULTS

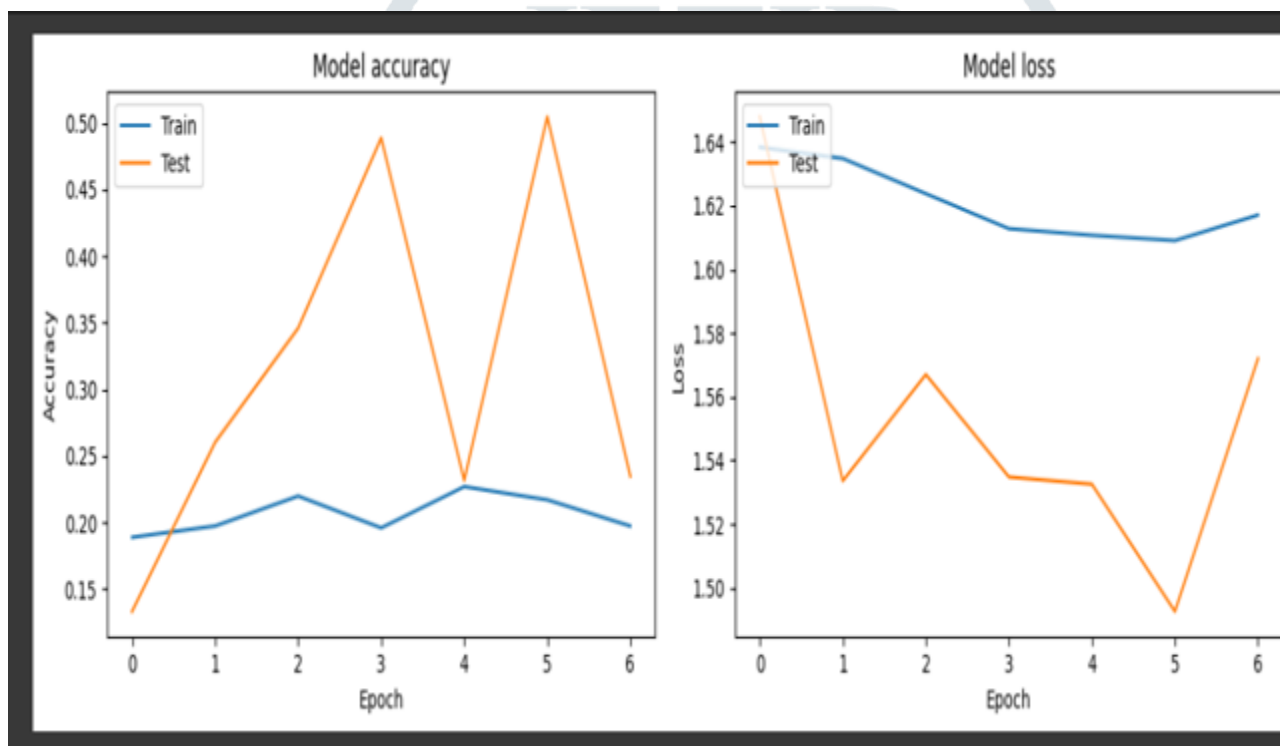


Figure 4.1: Model Accuracy Modal Loss

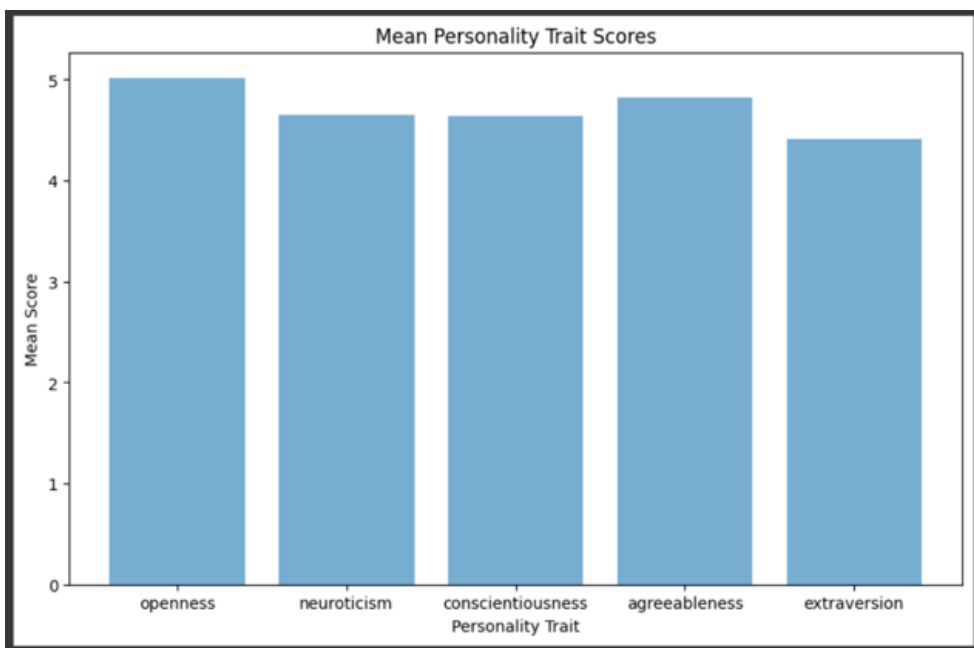


Figure 4.2: Mean personality trait score

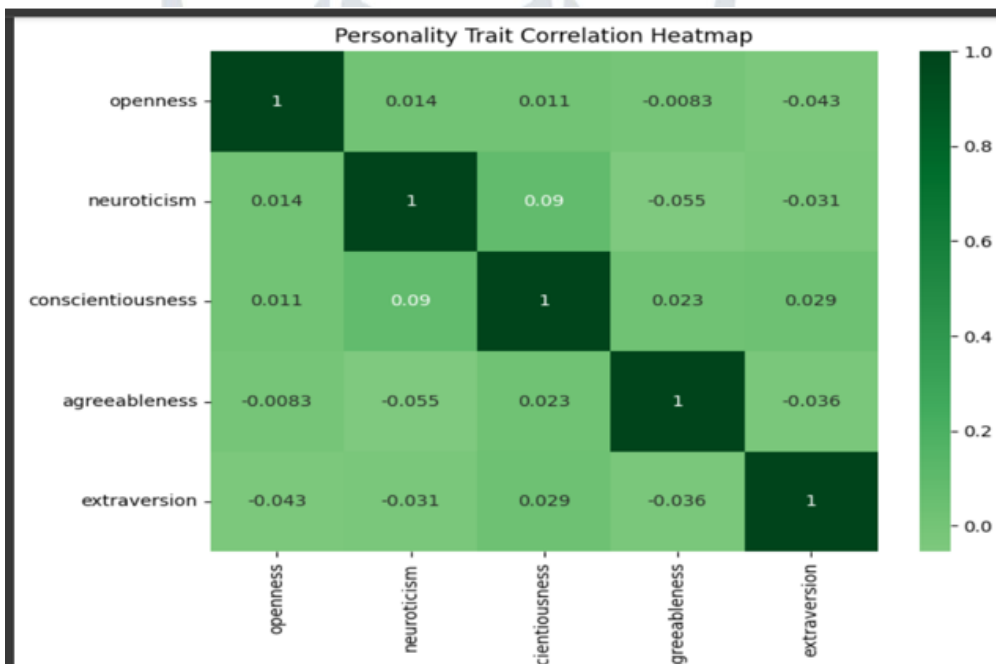


Figure 4.3: Personality trait heatmap

CONCLUSION

In conclusion, utilizing resumes for human behavior prediction presents a multifaceted landscape ripe with both opportunities and challenges. Through sophisticated data analysis and machine learning techniques, resumes can offer valuable insights into various aspects of human behavior, including work ethic, communication skills, and job performance potential. By leveraging this information, employers and recruiters can make more informed hiring decisions, leading to enhanced workforce productivity and organizational success. However, the use of resume data for behavior prediction also raises ethical concerns regarding privacy, fairness, and potential bias. It is imperative for stakeholders to prioritize transparency, accountability, and responsible data handling practices to mitigate these risks and ensure equitable outcomes for all individuals involved. Additionally, ongoing research and collaboration between academia, industry, and regulatory bodies are essential to address emerging challenges and advance the ethical and effective use of resume-based human behavior prediction in the evolving landscape of employment and recruitment.

FUTURE WORK

Future work in human behavior prediction based on resume data could explore several avenues to enhance predictive accuracy and applicability. Firstly, integrating natural language processing (NLP) techniques to extract deeper insights from textual resume content could provide a more comprehensive understanding of individual characteristics and tendencies. This could involve sentiment analysis, topic modeling, or entity recognition to capture nuanced aspects of personality, interests, and skillsets. Additionally, leveraging advanced machine learning models such as deep learning architectures could enable the discovery of intricate patterns and non-linear relationships within resume data, leading to more accurate predictions.

ACKNOWLEDGMENTS

We would like to express our gratitude to our college Principal, Dr. Pramod R. Rodge, for providing us with lab facilities and allowing us to pursue our project. We also extend our heartfelt thanks to our H. O. D., Dr. Savita S. Sangam, who provided us with the necessary computer facilities in our laboratory and played a significant role in making our project a success. Without their cooperation, our project would have come to a standstill. We would also like to thank our Project Guide, Dr. Savita S. Sangam, for her unwavering support throughout the project. Her expert guidance, kind advice and timely motivation have been invaluable in helping us develop our project, " Human Personality Prediction based on CV". We would like to acknowledge the contributions of our colleagues who supported us directly or indirectly during the project. In conclusion, we sincerely thank everyone who has contributed to our project and helped us achieve success.

REFERENCES

1. Matt James Francis, Mitja D Back, Eric Grunenber, Heinrich Peters (September 2023). Machine Learning in Recruiting: Predicting Personality from CVs and Short Text Responses.
2. Darpan Jain, Abhay Chanana, Nongmeikapam Thoiba Singh (January 2023). Personality prediction through CV analysis using machine learning techniques. 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies. (ICAECT).
3. Patel, S., Lee, M. (2019). Ethical Considerations in Predictive Modelling for Talent Acquisition 2019 Journal of Business Ethics.
4. Wang, L., Chen, H. (2022). Advances in Natural Language Processing for Behavioural Analysis in Recruitment. 2022 Expert Systems with Applications.
5. Atharva Kulkarni , Tanuj Shankarwar , Siddharth Thorat (September 2021). Personality Prediction Via CV Analysis using Machine Learning.