**Real time personality assessment via Big Five Traits**

**Akshaya J1, Abhiram Sharma2, Yaksh Rathod3, S Sridhar4**

**1**Student, Department of Computer Science and Engineering (E.Tech), SRM Institute of Science and Technology, Vadapalani, Chennai, India.

**2** Student, Department of Computer Science and Engineering (E.Tech), SRM Institute of Science and Technology, Vadapalani, Chennai, India.

**3** Student, Department of Computer Science and Engineering (E.Tech), SRM Institute of Science and Technology, Vadapalani, Chennai, India.

**4** Assistant Professor (Sr.G) , Department of Computer Science and Engineering (E.Tech), SRM Institute of Science and Technology, Vadapalani, Chennai, India

**Email :** **1**aj7795@srmist.edu.in, **2**km7993@srmist.edu.in,[**3**yj1223@srmist.edu.in](mailto:3yj1223@srmist.edu.in) , [4sridhars2@srmist.edu.in](mailto:4sridhars2@srmist.edu.in)

**Corresponding Author :** Akshaya J

1. **ABSTRACT**

Personality prediction from CVs using advanced NLP is quite a significant area of interest in contemporary human resource management and recruitment processes. The study generates an innovative approach to making predictions, with textual data extracted from CVs sent for analysis regarding the Big-5 personality traits of a person. We devise a rather powerful framework to quantitatively extract subtle personality cues from unstructured text using NLP techniques, in particular transformer-based models with the incorporation of contextual embeddings. Our approach consists of CV pre-processing to extract relevant linguistic features, training sophisticated machine learning models, such as the one we use in our course, to correlate such features with personality traits, and large-scale empirical assessment to validate model accuracy. The findings reveal the potential for advanced NLP techniques in developing deep insights into the way personality profiles of candidates can be; thus, it serves as a very effective tool for improvement in recruitment processes through enhanced precision and efficiency. It would not only support informed decision-making but open further possibilities of research on the integration of personality analytics into human resource practice.

1. **KEYWORDS**

Personality assessment, Big Five personality traits, BERT-based models, Recruitment automation, Machine learning

1. **INTRODUCTION**

The modern recruitment process in human resource management is increasingly challenged by the number of applicants that organizations have to evaluate. Traditional recruitment methods, which primarily assess candidates' abilities and experience, often neglect essential personality traits that significantly influence job performance and organizational fitness. This highlights the urgent need for automated systems that integrate personality assessment into the recruitment process and enable more accurate selection of candidates who not only meet the job qualifications but also fit the culture and operational needs of the organization.

The Big Five personality model provides a robust framework for understanding individual differences in behavioural, cognitive, and affective patterns through its five traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These traits have been widely recognized in both the psychological and organizational literature as indicators of job performance and career success. Traditional assessment of these traits, often dependent on psychometric tests or interviews, is time-consuming, subjective, and prone to bias. However, the integration of machine learning and natural language processing (NLP) offers a promising solution that enables automated and objective analyses of personality traits derived from textual data such as CVs.

This study aims to develop a machine learning system capable of predicting Big Five personality traits from CVs using sophisticated NLP techniques, particularly BERT-based models through the SentenceTransformer library. These models enable the extraction and analysis of linguistic features embedded in the unstructured text of resumes, facilitating the identification of subtle personality profiles that can provide valuable insights into a candidate's suitability for specific roles and the broader organizational environment.

The proposed approach not only seeks to increase the efficiency and accuracy of the recruitment process by integrating personality assessment in the early stages of candidate screening, but also aims to mitigate the risks associated with traditional recruitment methods that may overlook important personality dimensions. Through automation, organizations can scale their assessments and ensure greater consistency across large candidate pools, improving the overall quality of recruitment.

Additionally, the design and implementation of this system contributes to the ongoing discourse on the role of artificial intelligence in human resource management. By systematically using personality analysis in hiring processes, organizations can leverage data-driven decision-making, achieve immediate goals, and at the same time pave the way for future research into the ethical and practical implications of AI-based personality assessment in a variety of organizational contexts.

In summary, this study proposes a new approach to enrich the recruitment process by predicting Big Five personality traits from resumes using advanced machine learning techniques. By incorporating these personality assessment tools into their recruitment workflow, organizations can make informed decisions that align with immediate job requirements and long-term strategic goals to foster a productive and harmonious workplace. The findings from this research have the potential to change the way organizations approach talent acquisition, making it more holistic, data-driven and aligned with current HR practices.

1. **LITERATURE REVIEW**

The base paper for the current project is the research undertaken by ***Singh (2023)*** [1]where the goal is to predict MBTI from CVs using Random Forest and XGBoost. The study also shows how the current model offers increased efficiency as compared to earlier systems based on the MBTI. However, it mainly focuses on MBTI instead of the Big Five personality traits and encodes a small view of the resume parsing methods.

Moreover, ***Grunenberg et al. (2024)*** [2] investigate the method of predicting Big Five personality traits from CVs and brief answers. Some of the methods used in their process include stop word removal, stemming, topic modeling and word frequencies. This study’s conclusion is that although it is possible to predict personality characteristics through machine learning models it should not be the sole determinant for employment. There’s a large missing part that does not include resume parsing systems such as TF-IDF or a sentence embedding for similarity analysis.

***Madureira (2023)*** [3]formally investigates the prediction of Big Five types in Brazilian Portuguese over English embedding. Applying the English myPersonality dataset and Portuguese tweets with FastText features, the model is able to forecast traits over different language barriers. However, language generalization is the main subject of the study and does not restrict details regarding CV parsing or the characteristics associated with resumes.

That is why ***Ramos-Villagrasa (2022)*** [4] aims to determine the validity and accuracy of short biodata in assessing the Big Five traits. The current study uses a biodata scale that comprises of eight items to establish that biodata can predict job performance with incremental validity. It does not discuss in detail the personality predictions from CV and focuses more on the performance than the personality.

In a different context, ***Halim (2019)*** [5] investigates player profiling adopting Big Five traits by employing data from RTS games and clustering analysis. As shown by the results, there is a relationship between gameplay statistics and personality characteristics, yet this research is relevant only for estimated computer gameplay characteristics and is not suitable in the case of resume reading and recruitment.

***Leong Dickmond (2021)*** [6] describes the adoption of ML methodologies to predict personalities from resumes, using TF-IDF and 50 keywords. From their findings, they have established that this method is more effective than the other one. But they noted that more investigation is needed regarding the non-standard format of CVs, which is in parallel to the current observation being made in this literature review regarding custom parsing.

Among the e-recruitment techniques studied by ***Kamble (2022)*** [7] with family resume categorization and featuring video interviews, using algorithms yielded a 95% accuracy in auto personality recognition from videos and pathological questionnaires. This approach, however, is based on analysis of videos rather than the usual text documents also known as CVs.

***Katyal (2019)*** [8] uses four machine learning algorithms of Logistic Regression, SVM and for prediction of Big Five traits from cv including Random Forest and Result indicated that Rao’s test Random Forest has accuracy of approximately 71%. Nonetheless, the study is mostly classification oriented and does not consider semantic similarity or a more profound matching of phases.

***Suen (2019)*** [9] discusses Big Five recognition with AVI, using CNN for face expression analysis for the subject. The findings suggest that AI performs better than people in assessing Big Five factors, but this study does not include text-based CVs.

Finally, ***Nilugonda (2020)*** [10] on the analysis of big five prediction from facial expressions using deep learning on TensorFlow, resulting in accuracy of range from 96.9% to 99.5% and was able to surpass traditional classifiers. Like other research, it is confined to only video data, and no word about CVs.

Therefore, understanding the prediction of the personality from CV through different formats and modalities, still some key gaps are left the more particularly in the use of the advanced parsing techniques and in the exploration of the semantic similarity.

1. **SYSTEM ARCHITECTURE AND DESGN**

**5.1 Existing System Architecture**

The base paper by Singh (2023) outlines an approach to predict MBTI personality traits from CVs using classical machine learning models like Random Forest and XGBoost. In their system, resumes are processed to extract features such as word frequency, specific keywords, and basic semantic information. After preprocessing, these features are used to train Random Forest and XGBoost classifiers, which predict the personality traits associated with the MBTI.

**Feature Extraction:** The system relies on relatively simple feature extraction techniques, such as word count, specific trait-related keywords, and TF-IDF.

**Prediction Model:** Random Forest and XGBoost are employed for classification, predicting one of the 16 MBTI types based on the extracted features from CVs.

**Limitations:** The system's focus is narrow, as it predicts only MBTI traits and lacks advanced methods like semantic similarity or contextual embeddings. It also does not address the broader range of personality characteristics, such as the Big Five personality traits, which offer a more detailed and comprehensive understanding of human personality.

**5.2 Proposed Model Architecture**

The proposed model’s framework is done by expanding the focus from MBTI to the Big Five personality traits. It also integrates more sophisticated methods for feature extraction and personality prediction, improving both the depth and accuracy of the predictions. The new architecture makes use of BERT-based models for semantic similarity, alongside traditional methods like TF-IDF, to achieve a more nuanced and context-aware analysis of resumes.

**5.2.1. Preprocessing and Feature Extraction**

In this system, resume text undergoes advanced NLP processing for a more sophisticated extraction of relevant features:

* Contextual Embeddings with BERT: BERT-based models like all-MiniLM-L6-v2 and all-mpnet-base-v2 are employed to generate rich, contextual embeddings from the resume text. These embeddings capture deeper, nuanced meanings of words and phrases, enabling a more comprehensive understanding of candidate profiles.
* TF-IDF for Statistical Weighting: To complement the BERT embeddings, traditional term-frequency inverse-document-frequency (TF-IDF) techniques are used. TF-IDF helps determine the importance of terms relative to other resumes, providing a statistical layer to the model that captures surface-level word importance.

By combining BERT's contextual understanding with TF-IDF's term importance metrics, the system achieves a balance between deep semantic insights and traditional word significance.

**5.2.2. Personality Trait Prediction:**

Unlike systems focused on the MBTI, this model predicts the Big Five personality traits: Conscientiousness, Agreeableness, Neuroticism, Openness, and Extraversion. These traits are recognized as more robust for understanding personality in professional contexts.

* Using Sentence Transformers for Semantic Matching: The model calculates semantic similarity between job descriptions and resumes using sentence transformers. Both BERT embeddings and TF-IDF vectors contribute to these similarity scores, enhancing the prediction's depth and accuracy.
* Keyword-Based Scoring: A customized ResumeScorer function matches keywords associated with each of the Big Five traits, adjusting scores based on their frequency in the resume text. This scoring framework provides a data-driven evaluation of the candidate's likely personality traits based on resume content.

**5.2.3. Scoring and Matching:**

The system also evaluates resumes by comparing key criteria from the job description, including:

* **Skills Matching**: The model compares required and preferred skills using spaCy's PhraseMatcher to detect case-insensitive matches between resume content and job description. This is broken down into:
  + **Required Skills Match**: Ratio of required skills present in the resume.
  + **Preferred Skills Match**: Ratio of preferred skills that the candidate possesses.
* **Experience and Education Extraction**: The system extracts and normalizes the candidate's years of experience and highest level of education. These are then compared with job requirements:
  + **Experience Match Score**: A normalized score reflecting how closely the candidate's experience aligns with the job's expectations.
  + **Education Match Score**: A binary score indicating whether the candidate meets the job's education level requirements.

**5.2.4. Semantic Similarity and Aggregation:**

Semantic similarity between job descriptions and resumes is a core aspect of the scoring system, enhanced by the use of **BERT-based models** for contextual understanding. By combining both **BERT embeddings** and **TF-IDF similarity scores**, the model achieves a comprehensive comparison between job requirements and candidate qualifications.

* **Combined Scoring**: The system aggregates these similarity scores with the **skills**, **experience**, and **education match** scores. This combination ensures that candidates are evaluated not just on technical qualifications but also on their semantic alignment with the job description.

**5.2.5. Final Candidate Ranking**

The final output provides a ranked list of candidates based on a total score that encompasses:

* Semantic Similarity (BERT and TF-IDF)
* Skill Matching (Required and Preferred)
* Experience Alignment
* Education Fit
* Big Five Personality Traits

These ranked profiles offer hiring teams a holistic view of each candidate's fit for the role, balancing technical skills with personality insights. This improves on traditional evaluation systems that focus heavily on qualifications alone by adding a behavioral dimension, leading to more well-rounded hiring decisions.

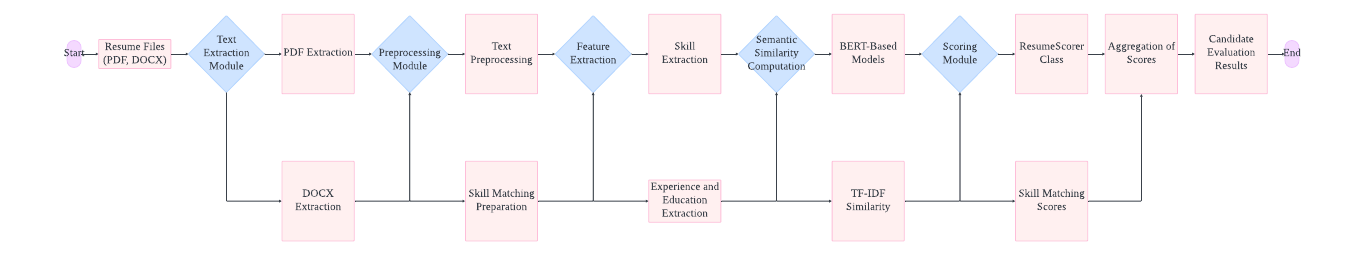


Fig.1 Architecture Diagram

1. **RESULTS**

The results displayed here show the performance of the Resume Shortlisting Application based on the user input and uploaded resumes. The application processes candidate resumes, matching them against the job description and ranking them based on various criteria including skills, experience, education, and Big Five personality traits.

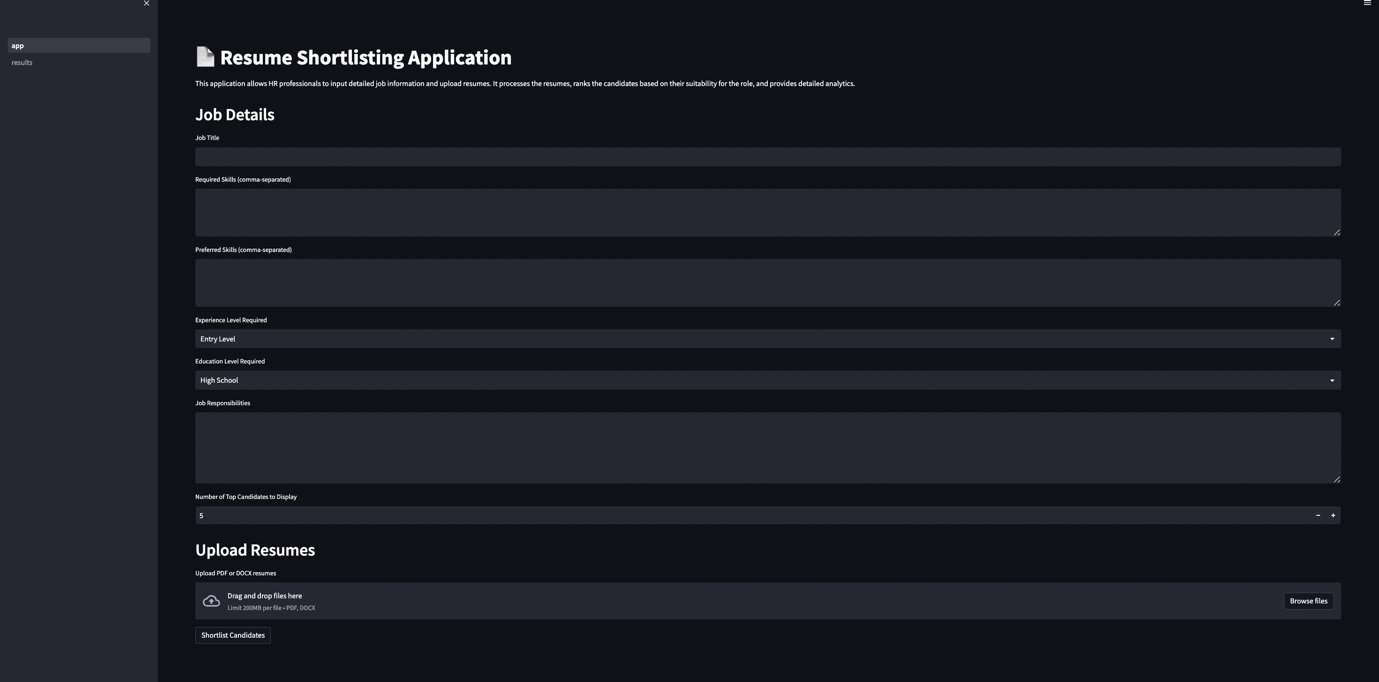


Fig.2. Input page for the user

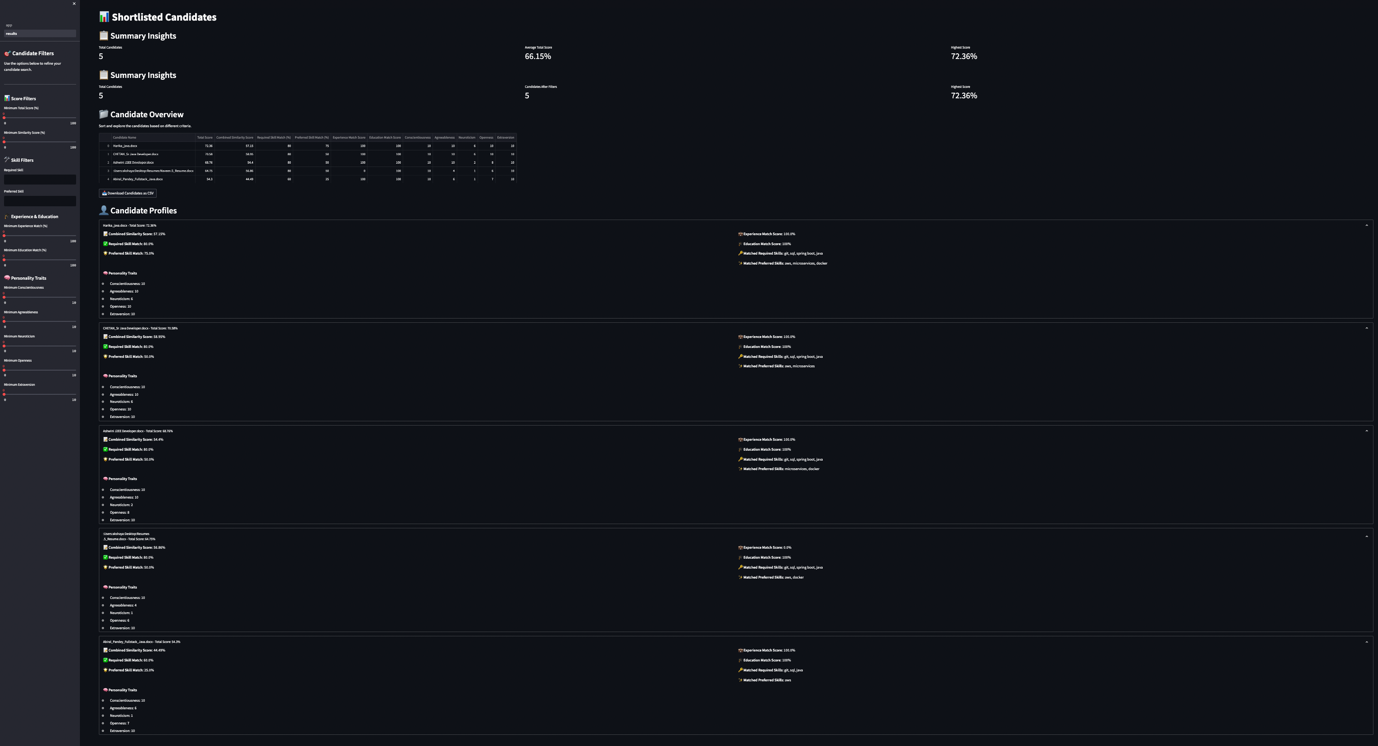


Fig.3. Results page

* **Summary Insights**: The system provides high-level statistics, including:
  + **Average Total Score**: This metric represents the degree of the candidate pool's overall match to the job requirements. For instance, the average score is 66.15%, so the candidates are somewhat close to the job specifications.
  + **Candidates Filtered**: 5 candidates have been picked from the pool of job applicants, which ensures that the resumes featured fulfill the minimum job prerequisites.
  + **Candidate Overview**: A table shows the results of all candidates on the aspects of, for example: Skills Match (both required and preferred), Experience Match, Education Match, Scores for each of the Big Five Personality Traits (Conscientiousness, Agreeableness, Neuroticism, Openness, Extraversion) In the case of Candidate 1, he/she has a skill set of 92% and he/she possess a high skills match for both required and preferred skills, while Candidate 5 may have low experience or skills matching scores. The overview gives you a small list of the top candidates that can easily be reviewed.
* **Candidate Profiles**: Each candidate undergoes an in-depth analysis and the results are broken down by:
  + **Experience Match**: This is based on how closely the candidate's years of experience align with the job’s requirements.
  + **Skills Match**: Required and preferred skills are individually assessed, offering insights into how well the candidate fits the technical and soft skill demands of the position.
  + **Personality Traits**: Using keyword analysis, the candidate’s personality traits are evaluated in terms of the Big Five, offering a deeper look into how their behavioural attributes align with the role.

1. **DISCUSSION**

The results illustrate the Resume Shortlisting Application’s effectiveness in evaluating candidates holistically. This system goes beyond traditional keyword matching by integrating advanced machine learning models like BERT and TF-IDF to assess not only technical fit but also semantic relevance and personality traits. The combination of skills, experience, and personality scores allows for a more comprehensive candidate evaluation.

**Candidate Ranking**:  
One of the most notable outcomes is the balance between hard and soft skills in the candidate ranking process. While **skills** and **experience** weigh heavily in the overall score, **personality traits** are equally important in assessing team fit and cultural alignment. For instance, candidates with a strong match on technical skills but lower personality trait scores may not be ranked as highly if the role requires more collaborative or client-facing abilities.

**Practical Use**:  
This system streamlines the HR process by reducing the manual effort required to sift through resumes. The detailed breakdown of each candidate’s strengths and weaknesses provides hiring managers with valuable insights for making more informed decisions.

**Limitations**:

While the system is robust in its technical assessments, personality trait scoring still relies on keyword-based models, which may not capture the full nuance of a candidate’s personality. There’s also a potential bias in using pre-defined keyword lists for traits, which may not always align perfectly with the actual personality indicators in resumes.

**Future Enhancements**:  
Further improvements could involve refining the personality trait evaluation using more dynamic methods such as deep learning-based sentiment analysis. Additionally, integrating real-world feedback loops into the system, such as performance reviews of hired candidates, would allow continuous refinement of the model’s scoring criteria, making future candidate predictions even more accurate.

**REFERENCES**

[1] Nongmeikapam Thoiba Singh, Abhay Chanana, Darpan Jain, Rajesh Kumar, "Personality Prediction Through CV Analysis Using Machine Learning Techniques," International Journal of Innovative Science and Research Technology, 2023, Volume 8, Issue 5, pp. 456-460. DOI: 10.31227/osf.io/t83fq.

[2] Eric Grunenberg, Heinrich Peters, Matt J. Francis, Mitja D. Back, Sandra C. Matz, "Machine Learning in Recruiting: Predicting Personality from CVs and Short Text Responses," Frontiers in Social Psychology, 2024, Volume 1, Article 1290295. DOI: 10.3389/frsps.2023.1290295​

[3] Madureira, A., Trindade Moura, F. R., Lopes Pinto, L. V., Neves Alves, A. V., Figueiredo, K., Ribeiro Costa, F. A., & Seruffo, M. C. da R., "Words Similarities on Personalities: A Language-Based Generalization Approach for Personality Factors Recognition," *IEEE Access*, vol. 11, pp. [specific pages], Mar. 2023. DOI: 10.1109/ACCESS.2023.3261339.

[4] Ramos-Villagrasa, P. J., Fernández-Del-Río, E., & Castro, Á. (2022). Analysis of a brief biodata scale as a predictor of job performance and its incremental validity over the Big Five and Dark Tetrad personality traits. *PLOS ONE*, *17*(10), e0274878. <https://doi.org/10.1371/journal.pone.0274878>

[5] Halim, Z., Atif, M., Rashid, A., & Edwin, C. A. (2019). Profiling players using real-world datasets: Clustering the data and correlating the results with the Big-Five personality traits. *IEEE Transactions on Affective Computing*, *10*(4), 568–584. <https://doi.org/10.1109/TAFFC.2017.2751602>

[6] Dickmond, L., Hameed, V. A., & Rana, M. E. (2021). 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). Sharjah, United Arab Emirates. <https://doi.org/10.1109/DeSE54285.2021.9719496>

[7] Kamble, P., & Kulkarni, U. (2022). An innovative approach of personality recognition for e-recruitment. In *2022 5th International Conference on Advances in Science and Technology (ICAST)* (pp. 324–328). Mumbai, India. https://doi.org/10.1109/ICAST55766.2022.10039605

[8] Katyal, P., Sharma, M., & Sharma, N. (2023). Comparative analysis of personality prediction systems of CV. *International Journal of Advances in Engineering and Management (IJAEM)*, *5*(1), 962–965. Retrieved from <http://www.ijaem.net>

[9] Suen, H. -Y., Hung, K. -E., & Lin, C. -L. (2019). TensorFlow-based automatic personality recognition used in asynchronous video interviews. *IEEE Access*, *7*, 61018–61023. https://doi.org/10.1109/ACCESS.2019.2902863

[10] Nilugonda, M., Madhavi, K., & Nagaraju, K. C. (2020). Big Five personality traits prediction using deep convolutional neural networks. *Communications in Computer and Information Science*.

# ACKNOWLEDGEMENTS

We express our humble gratitude to our Honorable Chancellor **Dr. T. R. Paarivendhar**, Pro Chancellor (Administration), **Dr. Ravi Pachamoothu**, Pro Chancellor (Academics) **Dr. P. Sathyanarayanan,** Pro Chancellor (Admin) for the facilities extended for the completion of the minor project work.

We would record our sincere gratitude to our Vice Chancellor, **Dr. C. Muthamizhchelvan and** Registrar, **Dr. S. Ponnusamy,** for their support in completing our minor project work by giving us the best of academic excellence support system in place. We extend our sincere thanks to our Dean, **Dr. C V Jayakumar,** and Vice Principal – Academics, **Dr. C. Gomathy** and Vice Principal - Examination - **Dr. S. Karthikeyan** for their invaluable support.

We wish to thank **Dr. P. Chitra**, Professor & Head, Department of CSE-ETech, SRM Institute of Science and Technology, Vadapalani Campus for her valuable suggestions and encouragement throughout the period of the minor project work and the course.

We are extremely grateful to our Minor Project Coordinator, **Dr. Golda Dilip**, Professor, Department of CSE, SRM Institute of Science and Technology, Vadapalani Campus, for leading and helping us to complete our course.

We extend our gratitude to our supervisor, **Mr.S. Sridhar**, Assistant Professor, Department of CSE-ETech, SRM Institute of Science and Technology, Vadapalani Campus for providing us an opportunity to pursue our minor project under his mentorship. He provided us with the freedom and support to explore the research topics of our interest.

We sincerely thank all faculty of DCSE / DCSE- ETech, staff and students of the department who have directly or indirectly helped our minor project.

Finally, we would like to thank our parents, our family members and our friends for their unconditional love, constant support, and encouragement.

**Team Members:**

1. Akshaya J
2. Abhiram Sharma
3. Yaksh Rathod