Decision Support System for Applicant's Qualifications and Personality using Machine Learning

Aldrin J. Diaz
San Carlos College
Pangasinan, Philippines

Aira Grace A. Esguerra
San Carlos College
Pangasinan, Philippines

Luijie C. Mangaliag
San Carlos College
Pangasinan, Philippines

Shyrelle Gresh DG. Ruan
San Carlos College
Pangasinan, Philippines

Jenniea A. Olalia
San Carlos College
Pangasinan, Philippines

Maynard Gel F. Carse
San Carlos College
Pangasinan, Philippines

ABSTRACT
The recruitment process is a critical phase for companies looking to hire competent and well-suited employees. However, traditional methods for evaluating applicants can be time-consuming, subjective and prone to bias. This study presents a decision support system that uses machine learning techniques to support the evaluation of job applicants based on qualifications and personality traits. The system processes input data from applicants' CVs and interview videos to extract relevant characteristics such as educational background, skills, certifications, work experience and the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism). Resume data is analyzed using Natural Language Processing and keyword matching to assess qualifications, while video features are processed using audio-visual analysis and ML models to predict personality traits. The extracted data is then matched against employer-defined criteria and applicants are ranked according to their overall suitability. The proposed system aims to streamline the recruitment process, reduce human bias and improve the objectivity and efficiency of applicant assessment. The results show the potential of integrating ML into recruitment workflows for more informed and data-driven decision making.

General Terms
Artificial Intelligence, Machine Learning, Decision Support System, Pattern Recognition, Natural Language Processing, Human Resource Management, Personality Analysis, Video Processing, Data Mining, Classification Algorithms.

Keywords
Decision Support System, Machine Learning, Applicant Evaluation, Recruitment, Natural Language Processing, Personality Traits, Audio-Visual Analysis, Big Five

1. INTRODUCTION
The recruitment process is an important part of organizational success. The analysis of a candidate's personality and qualifications in a recruitment process plays an important role in determining whether they are a good fit for the job for which they are applying. While qualifications demonstrate a candidate's technical skills and expertise, personality traits help determine job fitness, cultural alignment, teamwork, and overall workplace performance. The human resources (HR) department is mostly responsible during the hiring process as it is the one that reviews requirements, conducts interviews, processes all the necessary data, and selects candidates based on qualifications and personality. The most common ways to determine an applicant's traits are through traditional interviews or by answering a personality test, which most companies now require when hiring. A recruitment process like this takes a long time, especially in companies that have a large number of candidates waiting to be interviewed for each position. There was research developed to solve the problem, like the Integrated E-recruitment System for Automated Personality Mining and Applicant Ranking, for example, who was introduced (Faliagka et.al., 2014). The candidate's details would be extracted from the candidate's LinkedIn profile and their personality traits were automatically extracted from their social presence using linguistic analysis. However, significant limitations were noted, particularly in the inconsistent screening of major roles requiring specific training and credentials. Also, Personality Prediction Through CV Analysis using Machine Learning Algorithms for an Automated E-Recruitment Process was introduced (G. Sudha et. Al., 2021). In this work, a system that automates the task of segregating candidates based on eligibility criteria and personality evaluation in a recruitment process is proposed. To meet this requirement an online application is developed for the registration of candidate's details and analysis of candidate's personalities through an online Multiple-Choice Question (MCQ) test containing personality quizzes. Then the system analyzes professional eligibility by comparing the uploaded Curriculum Vitae trained datasets. This system employs a machine learning algorithm namely "Logistic Regression" which helps to make fair decisions to recruit a suitable candidate. With these observed limitations, the proponents wanted to improve existing studies to provide a fast and smooth recruitment process by analyzing each applicant's personality with their CVs and video introductions. This will enable a better and more effective way to shortlist many candidates applying for a particular position. Furthermore, the decision support system for applicant's qualifications and personality using machine learning will provide a user-friendly interface for HR professionals, allowing them to easily review and interpret the generated insights. The system will generate a personality profile for each applicant, highlighting key traits and skills that align with the specific requirements for the job. To sum it up, the proposed decision support system for applicant's qualifications and personality using machine learning aims to streamline and enhance the hiring process. By leveraging the power of artificial intelligence, we envision a future where recruitment is not only faster but also more accurate and insightful, ultimately leading to better hiring decisions and stronger, more cohesive teams within organizations.

2. RELATED LITERATURE

In the recruitment process, personality traits are important to ensure that the candidates who will be hired will not only meet the qualifications but also fit in well with the team and the culture of the company. The goal of this is to understand the complex web of applicants' traits and behavior. The Big Five Model (L. Goldberg, 1993), which provides an organized framework for classifying and comprehending individual variations across five broad dimensions, has a major contribution to this quest. The foundation for the Big Five Model is established by Goldberg's groundbreaking work, "The Development of Markers for the Big-Five Factor Structure". In this study, he identifies five fundamental aspects of personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism. This model, which has been adopted by psychologists and researchers worldwide as the most widely acknowledged theory of personality classification, will be important to this study since it will be used to describe the applicants' personalities. Along with the rapid growth of technology over the years, a notable development in recruitment technology is the Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process (Sudha et al., 2021). This system evaluates candidate eligibility and personality attributes via online tests and CV analysis. This also shortlists the candidates using Random Forest Classification and Logistic Regression and utilizes The Big Five Model to estimate a person's personality. Their test data accounts for 30% of the total dataset, allowing for adequate training and error correction, which will be adapted for this study along with the algorithms mentioned above. A machine learning-based method for predicting personality is present in Personality Prediction Using Machine Learning Classifiers, which uses classifiers including Support Vector Machines (SVM), Random Forest, and Gradient Boosting. Through the analysis of several factors, such as demographic data and textual data, the study proves that high-accuracy personality trait prediction was feasible. This study also demonstrates how machine learning classifiers can be used to increase the accuracy and dependability of personality assessments (X.Y. Chin., et. al., 2021). They employ several classification techniques, including Voting Classifiers, Logistic Regression, Naive Bayes, Support Vector Machines, and Ridge Algorithms. Based on the testing results, the Logistic Regression method performed better than the other approaches. To speed up hiring procedures and enhance candidate evaluation, academics have recently looked at cutting-edge techniques including personality and job role prediction utilizing text analysis and resumes. To anticipate job functions and personality qualities, Job Role and Personality Prediction Using CV and Text Analysis offers a fresh approach to candidate evaluation. The study shows that it is possible to automate candidate assessment procedures and improve recruitment efficiency by examining textual data from resumes and other sources. The study also demonstrates how machine learning algorithms and natural language processing methods can be used to increase the precision and dependability of candidate predictions (M. Goyal, S.S., et. al., 2022). Goyal's research examines how to improve candidate assessments and hiring outcomes by integrating social media data, linguistic analysis, and sophisticated algorithms.

In research entitled Personality Evaluation Through CV Analysis using Machine Learning Algorithm (A. Robey, K.S., et. al., 2019) that is published on International Journal for Research in Applied Science and Engineering Technology (IJRASET), a machine learning model is developed that can predict candidates' personality traits by analyzing textual information extracted from their CVs. By leveraging natural language processing techniques and advanced algorithms, the study demonstrates the feasibility of assessing candidates' personalities in a data-driven manner. The findings underscore the potential of integrating machine learning into recruitment processes to augment traditional assessment methods. New research also explores innovative approaches to enhance the candidate assessment process. One multimodal approach to assess candidates' personality traits based on video interviews is suggested by the research A Multi-modal Personality Prediction System, which combines visual and audio modalities in two distinct ways. These represent the average of the forecasts made from each modality separately and the concatenation of features in a multimodal environment (C. Suman and colleagues, 2023). Their method of using video to predict personality will be utilized for this study.

The use of facial image analysis to evaluate the Big Five personality traits has become more popular in recent years. The research Assessing the Big Five Personality Traits using Facial Images utilize computer vision techniques to analyze facial characteristics like morphology, symmetry, and expressions to derive personality traits from static photographs (A. Kanchur, et. al., 2020). Using a large- labeled dataset, they trained a chain of artificial neural networks (ANNs) to forecast Big Five self-reported scores. The result of this research is encouraging, indicating that particular facial cues may be connected to particular aspects of personality. For instance, extraversion and the intensity of a smile are tied, but neuroticism is linked to brow furrowing and tense facial muscles.

3. PRESENTATION OF ANALYSIS OF DATA

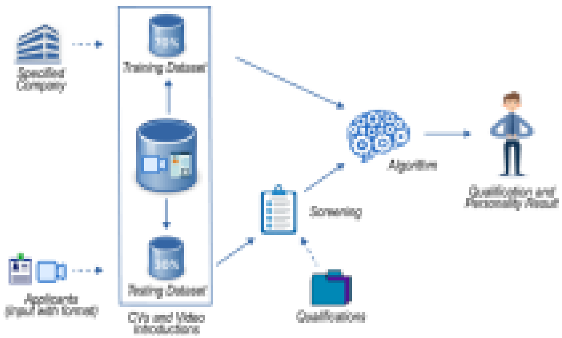

Figure 1. Operational Framework for the Proposed Project

The Decision Support System for Applicant's Qualifications and Personality Using Machine Learning helps companies optimize applicants for positions they are trying to fill. For the development of the proposed project, a dataset consisting of CVs and introduction videos was used, divided into a training dataset and a test dataset. The training dataset, which accounts for 70% of the total dataset, is used to train the machine learning model, while the test dataset is used to evaluate the performance of the trained model. We took inspiration from the dataset of a similar project called Comparative Analysis of Personality Prediction Systems of CV (P. Katyal et.al., 2023). After the system receives the required CV and introduction video, it will immediately review the collected data based on the qualifications specified by the employer. Missing values are checked, and the data is sorted out according to the desired criteria. Once the data has gone through the screening process, it is processed by an algorithm that analyzes and predicts the personality and qualification by comparing it to the trained data set using the model selected by the system.

As a result of the algorithm process, a shortlist of qualified candidates is displayed, showing the percentile of each applicant's traits (Openness, Conscientiousness, Extraversion/Introversion, Agreeableness, Neuroticism) and ranking according to compatibility to the job. The shortlist is only available for employers to review.

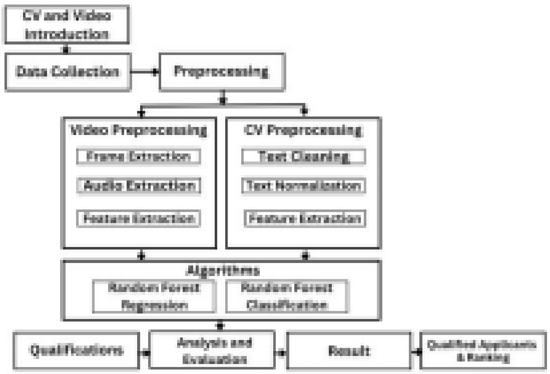

Figure 2. Overview of Experimental Design

In this methodology section, the systematic approach used to create a decision support system for the applicant's qualifications and personality with information taken from video introductions and CVs is described. Using machine learning and natural language processing (NLP) methods together, the objective is to predict personality traits with high accuracy.

To predict personality traits, the study used a mixed-methods approach, combining quantitative machine-learning approaches with qualitative data from video introductions and CVs. The basis for personality prediction is the Big Five personality traits model (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism).

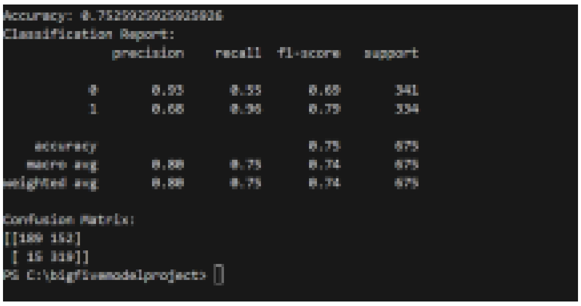

Figure 3: Visual Representation of CV Model Evaluation

The Random Forest Classification model is designed to predict categories such as whether a CV matches the employer's required skills, experience, education, and certifications. For this task, the model is trained using labeled data, where each CV is classified as either qualified or not qualified.

Evaluation Metrics:

In CV-based prediction, we are classifying applicants as either qualified or not qualified based on their CV attributes. The model's performance can be evaluated using the following metrics:

Accuracy:
Definition: The proportion of correct predictions out of all predictions made.
Formula:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$$

Explanation: Accuracy is a simple metric that gives the percentage of predictions that the model got right. It works well when the data is balanced (i.e., an equal number of qualified and unqualified applicants). However, in imbalanced datasets, it can be misleading.

Precision:
Definition: The proportion of true positive predictions among all positive predictions made by the model.
Formula:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Explanation: Precision answers the question: "Of all the applicants the model classified as qualified, how many were actually qualified?" This metric is important when the cost of false positives (misclassifying an unqualified applicant as qualified) is high.

Recall (Sensitivity):
Definition: The proportion of actual positive instances (qualified applicants) that were correctly predicted by the model.

Formula:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Explanation: Recall tells us how well the model is at identifying all qualified applicants. It is crucial when we want to minimize false negatives (missing out on qualified candidates).

F1-Score:
Definition: The harmonic mean of precision and recall, balancing the two metrics.
Formula:

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Explanation: F1-Score combines precision and recall into a single measure, providing a balance between the two. It is especially useful when the data is imbalanced, as it penalizes both false positives and false negatives.

Confusion Matrix:
Definition: A matrix that shows the number of correct and incorrect predictions, categorized by type (true positive, false positive, true negative, false negative). Explanation: The confusion matrix gives a detailed breakdown of the model's performance, helping to understand where it is making errors.

Cross-Validation:

Definition: A technique where the dataset is split into k subsets, and the model is trained and tested k times on different combinations of training and testing data. Explanation: Cross-validation helps to assess the model's generalization ability, ensuring that it performs consistently across different data splits.

CV Results:
Accuracy: 75%
This means that 75% of all CVs were correctly classified into their respective qualification categories.

Precision: 93%
A precision of 93% means that when the model predicted a CV as Qualified or Highly Qualified, it was correct 93% of the time.

Recall: 55%
A recall of 55% means that out of all truly Qualified or Highly Qualified candidates, the model correctly identified
55% of them.

F1-Score: 69%
An F1-score of 69% is the harmonic mean of precision and recall, providing a single score that balances both. These results indicate that the model is able to correctly classify the qualification of applicants with a high degree of precision and recall, while achieving a well-balanced F1- Score.

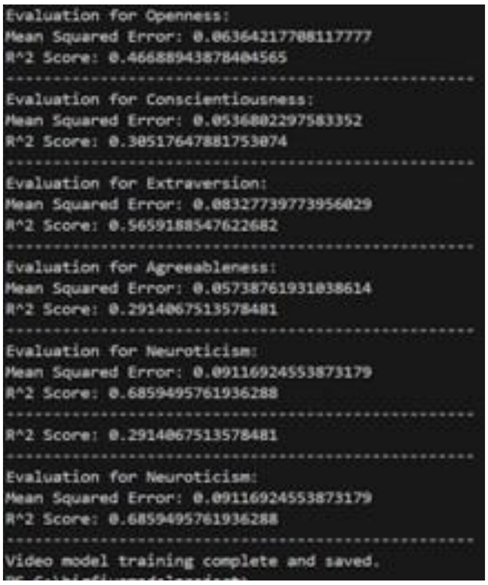

Figure 4: Visual Representation of Video Model

Evaluation

For Video Introduction-based predictions, where the model predicts continuous values (e.g., personality traits like Neuroticism), we use different regression metrics:

Mean Absolute Error (MAE):

Definition: The average of the absolute errors between predicted and actual values.
Formula:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{true},i} - y_{\mathrm{pred},i}\right|$$

Explanation: MAE measures the average magnitude of errors in predictions, with a focus on the absolute difference between predicted and true values. It provides a clear and interpretable metric for how far off predictions are on average.

Mean Squared Error (MSE):

Definition: The average of the squared differences between predicted and actual values.

Formula:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_{\mathrm{true},i} - y_{\mathrm{pred},i})^2$$

Explanation: MSE gives higher weight to larger errors due to the squaring of differences. This metric is useful for identifying outliers or large prediction errors.

Root Mean Squared Error (RMSE):

Definition: The square root of the MSE, which brings the error metric back to the original units of the predicted variable.

Formula:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

Explanation: RMSE gives a clearer sense of the error in the context of the original units, making it more interpretable than MSE. It is widely used for regression tasks and provides a sense of the average error in predictions.

R-squared (R2):

Definition: A measure of how well the model's predictions explain the variance in the target variable.

Formula:

$$R^2 = 1 - \frac{\sum(y_{\mathrm{true},i} - y_{\mathrm{pred},i})^2}{\sum(y_{\mathrm{true},i} - \bar{y}_{\mathrm{true}})^2}$$

Explanation: R2 represents the proportion of the variance in the dependent variable (e.g., personality trait scores) that is explained by the independent variables. A value close to 1 indicates a good fit, while values closer to 0 indicate that the model fails to explain much of the variance.

4. RESULTS, CONCLUSIONS, AND RECOMMENDATIONS

This study demonstrated that machine learning models can effectively analyze and predict applicant qualifications and personality traits using information from Curriculum Vitae (CVs) and video introductions. The CV-based model, which utilized Random Forest Classification, proved to be highly effective in categorizing applicants based on their qualifications. It performed best when analyzing well-structured CVs, where key features such as skills, education, work experience, and certifications were clearly defined. However, unstructured or inconsistently formatted CVs posed challenges, occasionally leading to misclassifications. To address this, improving text extraction techniques with advanced Natural Language Processing (NLP) methods could enhance the model's ability to process diverse CV formats more accurately.

For video-based personality prediction, the Random Forest Regression model showed a moderate level of accuracy in predicting Neuroticism scores. While the model successfully captured key audiovisual features from video introductions, external factors such as lighting conditions, audio clarity, and background noise affected its precision. As a result, more sophisticated video feature extraction techniques are needed to improve prediction reliability, ensuring that personality traits are assessed more accurately across varying recording conditions.

Looking ahead, integrating both CV and video data into a hybrid model presents an opportunity for a more holistic analysis of job applicants. By combining structured CV data with unstructured audiovisual information, a hybrid system could provide a more balanced and comprehensive evaluation of both qualifications and personality traits, allowing employers to make more informed hiring decisions. To further enhance the effectiveness and accuracy of the system, several key improvements are recommended. First, refining the data extraction process for CV-based predictions is essential. The current model performs well when processing structured CVs, but unstructured formats present challenges that can lead to misclassification. To address this, implementing advanced Natural Language Processing (NLP) techniques such as Named Entity Recognition (NER) and semantic analysis could improve the system's ability to capture nuanced information. These enhancements would allow the model to interpret a wider variety of CV formats more effectively, ensuring a more accurate classification of applicants. For video-based personality prediction, improving feature extraction is necessary to achieve higher precision. More advanced deep learning models, such as Convolutional Neural Networks (CNNs) for facial expression recognition and Recurrent Neural Networks (RNNs) for speech pattern analysis, could provide a more detailed and robust feature set. By extracting deeper insights from facial expressions, vocal tone, and other visual cues, the model would be better equipped to accurately predict personality traits, leading to a more reliable assessment of applicants.

Additionally, ensuring high-quality input data is critical for video-based predictions. Variability in video and audio quality—caused by differences in lighting, microphone clarity, and background noise—can negatively impact the model's accuracy. Standardizing the conditions under which applicants record their video introductions, such as ensuring consistent lighting and proper microphone use, would reduce inconsistencies and improve prediction reliability. Another significant improvement would be the integration of a hybrid model that combines CV and video introduction data. By leveraging both structured CV features and unstructured audiovisual traits, this approach would enable a more comprehensive evaluation of candidates. Such a model would provide a more balanced and holistic assessment, allowing employers to make more accurate hiring decisions based on both qualifications and personality attributes.

Lastly, further refinement of model performance can be achieved through cross-validation and hyperparameter tuning. Implementing these techniques would help identify the optimal model configuration, minimizing overfitting and ensuring better generalization to new, unseen data. By fine-tuning the system through systematic testing and optimization, the overall prediction accuracy and reliability of both CV-based and video-based models could be significantly improved. By addressing these areas for improvement, the system has the potential to provide a more accurate, efficient, and data-driven approach to evaluating job applicants, ultimately streamlining the hiring process and improving candidate selection outcomes.

In conclusion, while both CV-based and video introduction-based models offer valuable insights, further improvements in feature extraction and model training techniques are necessary to enhance the prediction accuracy and provide a more comprehensive assessment of applicants.

5. REFERENCES

[1] Chin, Xin Yee, et al. "Personality prediction using machine learning classifiers." Journal of Applied Technology and Innovation (e-ISSN: 2600-7304) 5.1 (2021): 1. Journal of Applied Technology and Innovation (e -ISSN: 2600-7304) vol. 5, no. 1, (2021)

[2] Goldberg, Lewis R. "The development of markers for the Big-Five factor structure." Psychological assessment 4.1 (1992): 26. DOI: https://psycnet.apa.org/doi/10.1037/1040-3590.4.1.26

[3] Goyal, Muskan, et al. "Job Role and Personality Prediction Using CV and Text Analysis." DOI: https://doi.org/10.22214/ijraset.2022.47201

[4] Kachur, Alexander, et al. "Assessing the Big Five personality traits using real-life static facial images." Scientific Reports 10.1 (2020): 8487. DOI: https://doi.org/10.3886/E109082V1

[5] Köchling, A., Riazy, S., Wehner, M.C. et al. Highly Accurate, But Still Discriminatory. Bus Inf Syst Eng 63, 39–54 (2021). https://doi.org/10.1007/s12599-020-00673-w

[6] Mali, Suraj, et al. "Personality Evaluation Through CV Analysis using Machine Learning Algorithm." International Journal for Research in Applied Science and Engineering Technology 10.6 (2022): 2060-2065. DOI: https://doi.org/10.22214/ijraset.2022.44124

[7] Nirmala, M., et al. "Personality Detection for Recruitment Using Machine Learning." International Conference on Communications and Cyber Physical Engineering 2018. Singapore: Springer Nature Singapore, 2024.

[8] Nirmala, M., Rajalakshmi, B., Sarvepally, S.R.M., Mandava, S., Megana, M. (2024). Personality Detection for Recruitment Using Machine Learning. In: Kumar, A., Mozar, S. (eds) Proceedings of the 6th International Conference on Communications and Cyber Physical Engineering. ICCCE 2024. Lecture Notes in Electrical Engineering, vol 1096. Springer, Singapore. https://doi.org/10.1007/978-981-99-7137-4_38

[9] Perera, Himaya, and Lakshan Costa. "MBTI Personality Classification through Analyzing CVs/Personal

Statements for e-Recruitment." DOI: 10.18178/ijml.2024.14.1.1153

[10] Samad, Devika. "A decision support system that provides an automated short listing of potential employees in recruitment and hiring of any organization." (2012).

[11] Sudha, G., et al. "Personality prediction through CV analysis using machine learning algorithms for automated e-recruitment process." 2021 4th international conference on computing and communications technologies (ICCCT). IEEE, 2021. DOI: https://doi.org/10.1109/ICCCT53315.2021.9711787.

[12] Suman, Chanchal, et al. "A multi-modal personality prediction system." Knowledge-Based Systems 236 (2022): 107715. DOI: https://doi.org/10.1016/j.knosys.2021.107715