Personality Profiling Using CV Analysis

Arya Raj S
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
cs.aryarajs@sbcemail.in

R Gopika Krishnan
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
gopikark2002@gmail.com

Drishya Das
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
drishyabrahmadas@gmail.com

Rohith R
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
rohithrchrkd@gmail.com

Jocelyn Ann Joseph
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
jocelyndaniel3@gmail.com

Abstract— **Human personality has been crucial to the growth of both organizations and individuals. Standard questionnaires and Curriculum Vitae (CV) analysis are two methods used to assess human personality. So, a personality prediction system that combines CV analysis and MBTI model questionnaires to accurately predict an individual's personality traits based on their uploaded CV is introduced. The system utilizes advanced Natural Language Processing (NLP) techniques to extract relevant information from the CV, including work experience, education, skills, and achievements. By analysing the textual content, the system identifies keywords and phrases associated with different personality traits, laying the foundation for precise predictions. MBTI model questionnaires are integrated to further enhance the accuracy of personality prediction. User responses to the questionnaires are carefully analysed and mapped to the corresponding personality traits using established psychological theories and models. A machine learning algorithm is then employed to create a predictive model, learning from a pre-labelled dataset of CVs and their associated personality traits. The system's performance is evaluated using metrics such as accuracy and precision, ensuring its effectiveness in capturing the nuances of individual personality traits. The developed system has significant applications in recruitment and team composition, aiding employers in making informed hiring decisions by evaluating candidates whose personalities align with specific job requirements. Additionally, individuals can benefit from gaining insights into their own personality traits, enabling them to make informed career choices and pursue tailored personal development opportunities. Overall, the proposed system provides an efficient and accurate approach for personality prediction based solely on CV analysis and questionnaire responses.**

Keywords—***Personality prediction, CV analysis, MBTI model, NLP, Machine Learning, Recruitment, Team Composition***

I. INTRODUCTION

The study presents a personality prediction system that uses computer vision (CV) analysis to predict personality traits based on uploaded CVs and MBTI Model questionnaires. This method offers a more objective and automated approach to personality assessment than traditional self-report questionnaires, interviews, or observations. The system extracts meaningful information from CVs and MBTI Model questionnaires, predicting various personality traits based on the Myers-Briggs Type Indicator (MBTI) model.

The system combines textual information from CVs and responses to MBTI Model questionnaires to uncover patterns, correlations, and indicators associated with specific personality traits. Machine learning algorithms and natural language processing techniques are employed to analyze the data and make predictions.

The system has potential applications in recruitment, hiring, education, and marketing. It can help identify candidates with desired personality traits for job roles, customize teaching approaches, and tailor marketing strategies based on predicted personality traits. However, ethical considerations such as privacy protection, informed consent, and potential biases must be considered.

In conclusion, the CV analysis system offers an innovative approach to understanding and predicting personality traits, with the potential to revolutionize industries, improve decision-making processes, and provide valuable insights into human behaviour.

II. BACKGROUND

Numerous assessments are available to determine personality types, including the MBTI, Rorschach, and Big Five tests. In this paper, MBTI model is utilized to predict.

A. MBTI Test

The Myers-Briggs Type Indicator (MBTI) is a commonly used personality assessment model that divides people into 16 personality types according to how they choose to be classified in four major dichotomies - Extraversion Vs. Introversion, Sensing Vs. Intuition, Thinking Vs Feeling, Judging Vs. Perceiving.

The MBTI is not a diagnostic tool, but it can be used to help people understand their own personality and the personalities of others. The MBTI is widely used in business, education, and personal development settings.

- *Extraversion (E) vs. Introversion (I):* This dimension reflects how individuals gain energy – either from external interactions and stimulation (extraversion) or from introspection and solitary activities (introversion).
- *Sensing (S) vs. Intuition (N):* This dimension pertains to how people gather and process information. Sensors depend on actual information and data, while Intuitives focus on patterns, possibilities, and abstract thought.
- *Thinking (T) vs. Feeling (F):* This dimension relates to decision-making processes. Thinkers make decisions based on logical analysis and objective criteria, while Feelers consider values, emotions, and the impact on relationships.
- *Judging (J) vs. Perceiving (P):* This dimension refers to a person's approach to the outside world. Judgers prefer structure, planning, and closure, while Perceivers are more adaptable, spontaneous, and open-ended in their approach.

These dichotomies result in 16 possible combinations, each represented by a four-letter code. Here are the 16 MBTI personality types:

1) *ISTJ - Introverted, Sensing, Thinking, Judging:*
   - Practical, organized, and meticulous.
   - Prefers structure and order.
   - Relies on facts and evidence.
   - Values efficiency and responsibility.
2) *ISFJ - Introverted, Sensing, Feeling, Judging:*
   - Warm, nurturing, and supportive.
   - Focuses on harmony and cooperation.
   - Values tradition and stability.
   - Attentive to the needs of others.
3) *INFJ - Introverted, Intuition, Feeling, Judging:*
   - Insightful, creative, and idealistic.
   - Values deep, meaningful connections.
   - Driven by personal values and vision.
   - Seeks to make a positive impact on the world.
4) *INTJ - Introverted, Intuition, Thinking, Judging:*
   - Strategic, visionary, and independent.
   - Analytical and future-oriented.
   - Prefers efficiency and results.
   - Focuses on long-term goals.
5) *ISTP - Introverted, Sensing, Thinking, Perceiving:*
   - Adaptable and hands-on problem solvers.
   - Enjoys exploring and experimenting.
   - Values practical solutions.
   - Prefers flexibility and spontaneity.
6) *ISFP - Introverted, Sensing, Feeling, Perceiving:*
   - Artistic, sensitive, and gentle.
   - Values personal expression and authenticity.
   - Appreciates aesthetics and beauty.
   - Prefers a flexible and adaptable lifestyle.
7) *INFP - Introverted, Intuition, Feeling, Perceiving:*
   - Idealistic, creative, and empathetic.
   - Values authenticity and individuality.
   - Seeks personal growth and meaning.
   - Empathizes deeply with others.
8) *INTP - Introverted, Intuition, Thinking, Perceiving:*
   - Analytical, curious, and inventive.
   - Values logical reasoning.
   - Enjoys exploring new ideas.
   - Prefers flexibility and open-endedness.
9) *ESTP - Extraverted, Sensing, Thinking, Perceiving:*
   - Energetic and goal-oriented.
   - Enjoys hands-on problem solving.
   - Prefers a dynamic and fast-paced environment.
   - Adaptable and spontaneous.
10) *ESFP - Extraverted, Sensing, Feeling, Perceiving:*
   - Sociable, outgoing, and lively.
   - Enjoys engaging with others.
   - Values experiences and spontaneity.
   - Prefers flexibility and variety.
11) *ENFP - Extraverted, Intuition, Feeling, Perceiving:*
   - Enthusiastic, creative, and empathetic.
   - Values exploration and possibilities.
   - Prefers variety and adaptability.
   - Motivated by personal values.
12) *ENTP - Extraverted, Intuition, Thinking, Perceiving:*
   - Inventive, curious, and strategic.
   - Enjoys exploring new ideas and possibilities.
   - Values innovation and debate.
   - Prefers flexibility and adaptability.
13) *ESTJ - Extraverted, Sensing, Thinking, Judging:*
   - Efficient, organized, and practical.
   - Values structure and order.
   - Takes a systematic and results-oriented approach.

- Prefers clear goals and plans.

14) *ESFJ - Extraverted, Sensing, Feeling, Judging:*

- Warm, sociable, and nurturing.
- Values harmony and cooperation.
- Enjoys helping and supporting others.
- Prefers structure and tradition.

15) *ENFJ - Extraverted, Intuition, Feeling, Judging:*

- Charismatic, empathetic, and persuasive.
- Values connection and collaboration.
- Motivated by a vision for the future.
- Seeks to inspire and lead.

16) *ENTJ - Extraverted, Intuition, Thinking, Judging:*

- Strategic, assertive, and goal-oriented.
- Values efficiency and results.
- Takes charge and enjoys leadership roles.
- Prefers structured plans and organization.

B. Logistic Regression Algorithm

Logistic regression is a notable machine learning approach used to solve binary classification problems. It simulates the link between the input attributes and the binary outcome variable, estimating the likelihood that an instance belongs to a given class. Despite its title, logistic regression is actually a classification algorithm rather than a regression algorithm.

The algorithm maps the input features to a probability value between 0 and 1 by using a logistic function, commonly referred to as the sigmoid function. The linear combination of the input features and their associated weights is transformed into a probability score by the logistic function.

The logistic regression algorithm follows these steps:

- *Input:* Get a collection of dependent binary labels (y) that correspond to a set of input attributes (X) for a training dataset.
- *Feature Scaling:* Perform any necessary pre-processing steps, such as feature scaling or normalization, to ensure the input features are on a similar scale.
- *Model Training:* Initialize the weights (coefficients) and bias term (intercept) randomly or with some predefined values. The logistic regression model seeks to identify the best possible values for these parameters.
- *Hypothesis Calculation:* Compute the weighted sum of the input features and apply the logistic function to obtain the predicted probability of the positive class:

  $z = w1*x1 + w2*x2 + ... + wn*xn + b$ probability = sigmoid(z)

  Here, w1, w2, ..., wn are the weights, x1, x2, ..., xn are the independent input features, b is the intercept term, and sigmoid(z) is the logistic function.
- *Loss Function:* Describe a loss function to measure the difference between predicted probabilities and true labels. The binary cross-entropy loss is the most widely used loss function in logistic regression.
- *Gradient Descent Optimization:* Minimize the loss function using optimization algorithms like gradient descent. The goal is to update the coefficients and bias in the direction that reduces the loss and improves the model's predictions.
- *Model Evaluation:* Using evaluation metrics like accuracy, precision, recall, or F1 score, evaluate the trained logistic regression model's performance. This evaluation is typically done on a separate validation or test dataset that the model hasn't seen during training.
- *Prediction:* Use the trained logistic regression model to predict the probability of the positive class for new instances based on their input features. A threshold can be applied to convert the probability scores into binary class predictions.

C. Natural Language Toolkit

A Python package called Natural Language Toolkit (NLTK) is used to carry out natural language processing (NLP). It was developed as a tool to facilitate the easy implementation of NLP in Python-based apps. NLP has become an essential component of contemporary technology due to the rise of unstructured data from social media, online reviews, and voice-based human-computer interface, among other factors. For numerous NLP applications, including morphological processing, syntactic analysis, semantic analysis, pragmatic analysis, automatic text summarization, etc., NLTK provides a helpful toolbox. Modules and sub-packages make up NLTK. In an NLTK processing pipeline, modules are called sequentially. This library is used to transmit Python data structures from one module to another. NLTK is a rich and versatile library that facilitates various aspects of natural language processing, from basic text processing tasks to more advanced language understanding and generation tasks. It's widely used in both academia and industry for research, development, and education in the field of natural language processing.

III. Related Study

Jayashree Rout et al. [1] proposed a system for automating the eligibility checking and aptitude assessment of candidates in the recruiting procedure. In order to address this, a candidate's resume and results from an aptitude or personality evaluation are analyzed using an internet program. Based on the supplied CV, the system evaluates a user's eligibility for employment. The TF-IDF algorithm is used by the system as part of a machine learning strategy. So the system's output provides a suggestion for an ideal candidate.

Atharva Kulkarni et al. [2] investigated various machine learning approaches for accurately predicting personality using CV analysis while also including Natural Language Processing (NLP) techniques. According to the findings from the study, the Random Forest algorithm outperformed the other algorithms including kNN, Logistic Regression, SVM, and Naive Bayes in terms of precision and accuracy.

Rutuja Narwade et al [3] proposed a system that makes use of Logistic Regression, Natural Language Toolkit, Flask, and Firebase for smooth and efficient deployment. The main methodology behind the personality questionnaire is the Big Five Personality Model which makes use of the OCEAN keywords namely, "Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism" for predicting an individual's personality.

In [4], the paper uses k-means clustering machine learning technique to classify personality types, a crucial application of machine learning algorithms. This classification can provide significant benefits, such as helping individuals choose careers and interests, and increasing efficiency in shortlisting candidates. Organizations also rely on personality predictions for shortlisting candidates.

IV. PROPOSED SYSTEM

A. Methodology

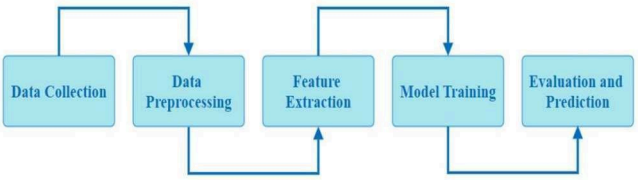


Fig. 1. System Architecture

The methodology for the personality prediction system via CV analysis involves several key steps as shown in figure-1. First, a dataset of CVs is collected, encompassing information such as work experience, education, skills, and achievements. Furthermore, MBTI Model questionnaires are created, which contain statements based on four types of preferences: extroverted versus introverted, sensing versus intuiting, thinking versus feeling, and judging versus perceiving. These questionnaires require respondents to rate their agreement on a numerical scale.

Next, the CVs undergo pre-processing to clean and normalize the textual data, including tasks like tokenization and removal of stop words. Simultaneously, the questionnaire responses are encoded into numerical values representing the ratings.

Following pre-processing, features are extracted from both the CVs and questionnaire responses. For the CVs, relevant features can be derived, such as term frequencies, TF-IDF scores, or word embeddings, which capture the quantitative representation of the textual information. The numerical ratings from the questionnaires serve as features themselves.

To create a comprehensive representation, the extracted features from both the CVs and questionnaires are combined. This fusion step allows for a holistic view of an individual's profile, integrating textual information and personality trait ratings.

With the combined feature representation, a machine learning model is trained using an appropriate algorithm i.e., logistic regression. To accurately assess the model's performance, the dataset is divided into training and testing sets during the training process.

Once the model is trained, it is used to predict personality traits for new individuals based solely on their uploaded CVs and questionnaire responses. The model takes the combined features as input and generates predictions based on MBTI model.

Finally, the predicted personality traits are interpreted and analysed within the context of the MBTI Model framework. The system uses information from the submitted CV and questionnaire to provide insights into a person's personality.

The methodology ensures that the personality prediction system leverages CV analysis and MBTI Model questionnaires to predict personality traits accurately, enabling valuable applications in various domains such as recruitment, education, and marketing.

B. Modules

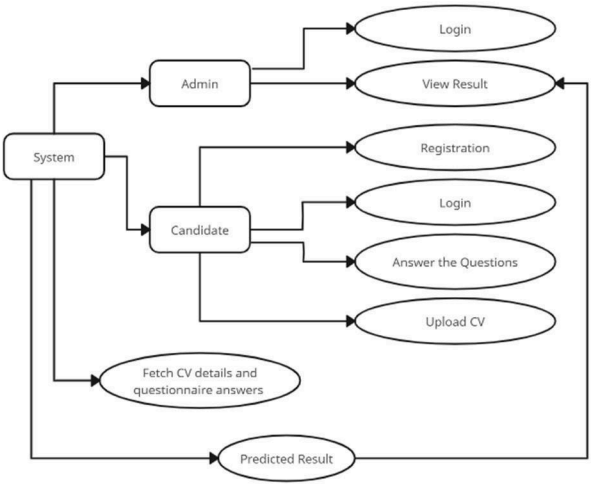


Fig. 2. System Flow Diagram

The two main modules of the proposed system are the admin login and the user login, each of which is made up of numerous smaller modules. For both the instructor and the candidate to access these sub-modules, they must log in with working credentials. Applying candidates would use their user login, while the recruiting company would use their admin login.

1) Module 1-Admin Page:

- *Login*: In order to modify the system's parameters and gain access to Module 1's sub-modules, the administrator must first log in with their admin ID and password.
- *Personality Questions:* The admin can include entire personality-related question in this sub-module to determine the applicant's personality. It would be based on the MBTI personality model. It comprises of four key characteristic features– Extraversion(E) Vs. Introversion(I), Sensing(S) Vs. Intuition(N), Thinking(T) Vs Feeling(F), Judging(J) Vs. Perceiving(P).

- *View Candidates:* All of the registered candidates' details are visible to the admin.
- *View Outcomes:* The admin has access to each candidate's results along with their predicted personality type which makes it easier for the admin to choose the preferred applicant.

2) *Module 2-Applicant's Page*

- *Registration and CV upload:* To access the user module listed below, the applicant must first complete the registration form and then generate login credentials.
- *Login:* To access the submodules below, the applicant must enter valid credentials.
- *Test:* The candidate can move on to the online personality test after successfully logging into the system.
- *CV upload:* The candidate has to upload their CV along with the personality test and submit it.
- *View outcomes:* Once the test has been successfully completed, the candidate can see the results, which will display their predicted personality.

V. Implementation

A. Imports:

Importing necessary libraries such as pandas, numpy, tkinter, sklearn, spacy, etc., for data handling, GUI creation, model training, and natural language processing.

B. train_model Class:

- This class is responsible for training the logistic regression model using data from the provided CSV file (mbti.csv).
- The train method reads the data from the CSV file, preprocesses it, and trains a logistic regression model on specific columns.
- The logistic regression model is used to predict the target variable ('Personality (Class label)') based on features like gender, age, and questionnaire responses and extracted features from CVs.
- The test method takes a test data array as input and uses the trained model to predict the output.

C. check_type Function:

- This function takes a data point and checks its type.
- If the data is a string or a tuple, it returns the title-cased string.
- If the data is a list or a tuple, it concatenates its items into a string.

D. Event Handling Functions:

- Functions like confirm_signup, admin_sign_in, candidate_sign_in, and confirm are bound to button clicks.
- These functions handle user authentication, sign-up confirmation, and form submission.

E. prediction_result:

Gathers input data from the GUI, processes it, and uses the trained model's test function to predict personality traits based on the input and displays the result in the RESULT tab.

F. File Upload and Text Processing:

- The user can upload a PDF file containing a CV, and the important features are extracted and analysed based on which the prediction is made along with other features and file path is also displayed.
- The user's responses to personality-related questions are processed and used for personality prediction.

G. Confirm:

Validates if all necessary fields are filled before making predictions or displaying success/error messages.

H. display_personality_traits(personality):

This function takes the predicted personality traits in the form of a dictionary and displays them in a readable format. It's meant to be used after a personality prediction is made and serves to provide feedback to the user about the predicted traits.

I. Mainloop:

Starts the GUI and runs the application.

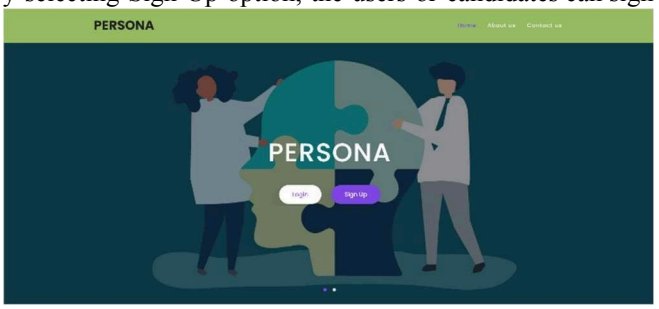
VI. Results And Discussion

This section presents an analysis of the results gathered. We will describe the outcomes as snapshots of the acquired output is given below.

A web-based platform is implemented, with two main modules: Instructor Panel and Applicant Panel. In (Fig.3) shows the Home Panel.

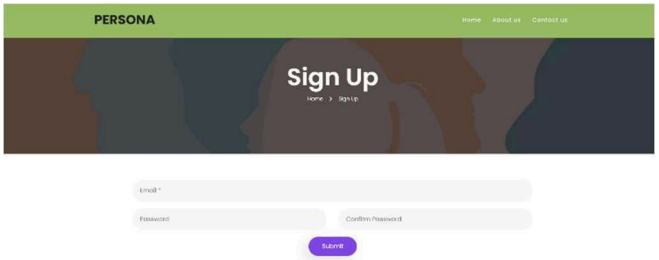
Fig. 3. Home Panel

By selecting Sign-Up option, the users or candidates can sign



up into the system using their credentials (Fig.4).

Fig. 4. Sign-Up Tab-1



In the Sign-Up Panel (Fig.5), the candidates can register onto

the portal by adding their personal details and setting a Username and Password along with the browse CV option to upload their CV and parse necessary details that are needed to predict the personality.

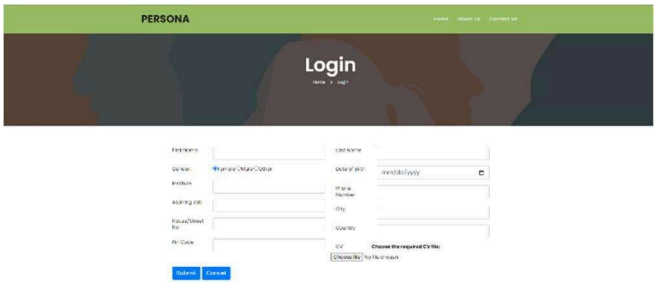


Fig. 5. Sign-Up Tab-2

By selecting Login option (Fig.6), the users or candidates can login into the system.

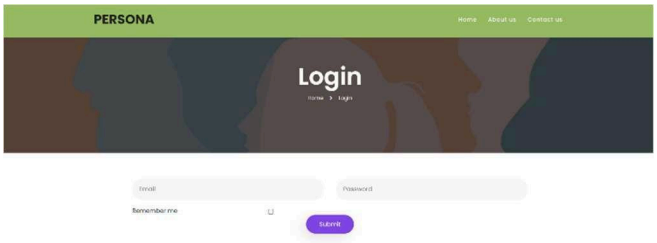


Fig. 6. Login Panel

After logging into the portal, the User will be able to attempt the MBTI test with a rating scale of 1-7 in the personality assessment tab (Fig.7).

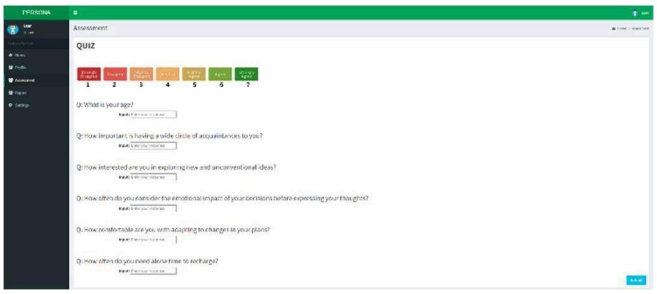


Fig. 7. Personality Assessment Tab

So, after successful submission, the system gathers input data from the questionnaire, processes it, and uses the trained model's test function to predict the personality of the candidate as shown below (Fig.8).

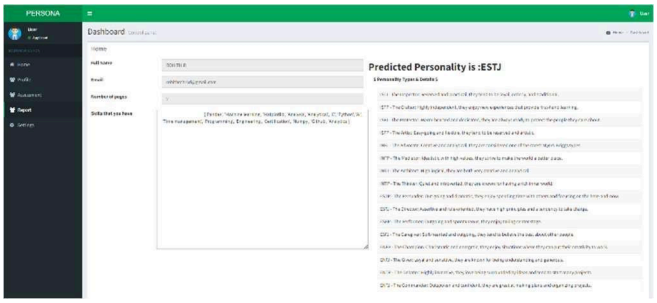


Fig. 8. Result Panel

After logging into the Instructor Panel, the Admin will now be able to see the finalized result i.e., the predicted personality of the candidates who have registered through the Candidate Panel (Fig.9).

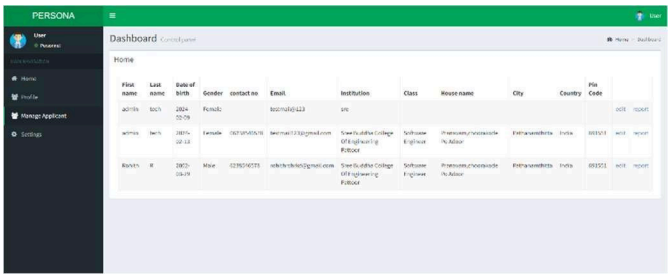


Fig. 9. Report Panel

VII. Experimental Results

After feeding the data to the models, we were able to infer that the Logistic Regression model delivers the highest accuracy, followed by SVM, kNN, and Random Forest, as shown in (Fig.10).

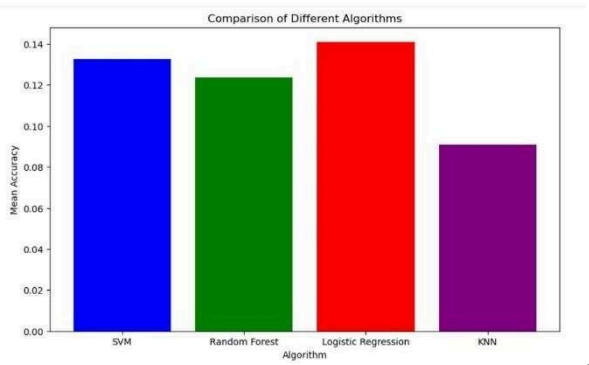


Fig. 10. Analysis

VIII. Conclusion And Future Scope

For predicting personality traits, the system analyzes CVs and conducts quizzes. It makes use of the MBTI framework and machine learning technologies like logistic regression. By converting CV text into numerical characteristics using TF-IDF feature extraction, the system predicts traits with a reasonable level of accuracy. Adding more questionnaires improves prediction accuracy and offers an in-depth understanding of personality. The system predicts personality traits through CV analysis, aiding recruiters, career counselors, and HR professionals in hiring and providing personalized guidance. However, it relies solely on CV and questionnaire information and does not consider facial expressions, body language, or visual content. Therefore, it should be considered part of a comprehensive evaluation process, with human judgment and intuition also considered. Further research suggests that in the future, the dataset can be improved such that the model produces more relevant results. Further research is needed for improved robustness and wider contextual information. Further study can be undertaken to increase the efficacy and effectiveness of the suggested system, hence increasing the accuracy of personality prediction using CV analysis.

In conclusion, the personality prediction system via CV

analysis presents a valuable tool for understanding an individual's personality traits.

REFERENCES

[1] Jayashree Rout, Sudhir Bagade, Pooja Yede, Nirmiti Patil "Personality Evaluation and CV Analysis using Machine Learning Algorithm", International Journal of Computer Sciences and Engineering (IJCSE), May 2019.

[2] Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat "Personality Prediction System Via CV Analysis Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), April 2021.

[3] Rutuja Narwade, Srujami Palkar, Isha Zade, Asst Prof. Nidhi Sanghavi "Personality Prediction System Via CV Analysis Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), April 2022.

[4] Hima Vijay, Neenu Sebatian, "Personality Prediction using Machine Learning", International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 2022

[5] Manasi Ombhas, Prajakta Gogate, Tejas Patil, Karan Nair "Automated Personality Classification Using Data Mining Techniques", April 2017.

[6] Aditi Chandrashekar, Santhini Nadar, Shreya Ganesh, Mrinal Khadse "Online Recruitment System and Personality Prediction" International Journal for Research in Engineering Application and Management (IJREAM), April 2021.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[7] Aleksandar Kartelj; Vladimir Filipović; Veljko Milutinović." Novel Approaches to Automated Personality Classification: Ideas and their Potential"".

[8] Vivian LAI, Kyong Jin SHIM, Richard J. OENTARYO, Philips K. PRASETYO, Casey VU, Ee-peng LIM, and David LO:" Career Mapper: An Automated Resume Evaluation Tool".

[9] Allan Robey, Kaushik Shukla, Kashish Agarwal, Keval Joshi, Professor Shalmali Joshi, "Personality Prediction System through CV Analysis", International Research Journal of Engineering and Technology (IRJET), Feb 2019