**CHAPTER 1**

**Introduction**

Now a days the growth of industries is too fast. The requirements of the industries are increasing day by day also they need a better locality for working purpose. For that we have to detect the persons personality and an overall performance in an early stage. The analysis of human behavior is quite difficult in this era. Personality is the quality of person which separates him from the bunch of people. If we know the behavior of the person in advance, then it will be easy to communicate with him and also to make predictions about him. The study of personality falls into Psychological Study. The prediction of personality is tedious work as we have to check all the aspects of a particular person, but by using machine learning it becomes easy and accurate too. We can also check with the skill set of the requested candidate using the submitted CV. The whole CV can be analyzed with the help of a machine learning algorithm which is TF-IDF. We can get a personal profile of the required candidate using TF score. Thus, we can evaluate the requested candidate based on these two criteria.

In this seminar report, we discuss about the various psychological techniques that are used along with machine learning algorithm to predict personality. Mostly we focused on MBIT, Big-Five tests that are widely used in this field. Also, we are going to look into TF-IDF, Logistic Regression machine learning algorithm used to predict personality. Finding personality not only help others but also helpful for the person to find its personal interest. HR department handles this selecting process, for reducing their work as well as time we can use this software and will give the selected applications of candidates to HR department to make a choice of the right candidate. This reduce departments work by reviewing the applications of an employee who is most eligible for their requirements.

In this system, we are predicting the personality of a particular candidate by using MBTI test and Big five personality model. These are open-source resources available.

After a candidate passes a certain limiting score, then he can submit his respective CV. The submitted CV and requirements of the company are Inputs to the TF-IDF machine learning algorithm. The algorithm processed data and gives TF score and the weight. Then company have to compare the weights and make a threshold score for evaluation of candidate. The system is more useful for HR department as these automated the process of requirement specification of the candidate. We can implement this in future as a real time application. As the whole system is focusing on the requirements of the company so it is more beneficial to use for a particular job requirement profile. Also, the persons traits are been determined using the personality models, so it can be also beneficial for an individual to do his personal development. For enable a more effective way to shortlist submitted candidates CVs from a large number of applicants providing a consistent and fair CV ranking policy, which can be legally justified to job requirements. The industry can use this system to choose expert candidate.

**CHAPTER 2**

**Literature Survey**

**2.1 Analysis of Human Traits Using Machine Learning [1]**

The Prediction of personality is treated as a classification problem in computer science as the people are classified into the different classes of the personality types. There are a number of psychological tests that yield different types of personality classes. Popular tests include MBTI, Big Five, DISC. The Myers-Briggs Type Indicator (MBTI) is one of the most famous and widely used personality tests or descriptors. They describe the way people behave and interact with the world around them with four binary categories and 16 total types.

They processed of the data using the nltk library which provides the inbuilt list of stop words available in the English language. The regular expressions are used to remove the http links, symbols, numbers. Using the TF-IDF vectorizer feature vectors are created with the low computational overhead. The proposed system for the personality prediction using the XGBoost yielded the best performance. The XGBoost model outperformed the logistic regression model, Naïve bayes Classifier and Decision Tree Classifier. The performance of the XGBoost is increased after the hyper parameter tuning which further increased the accuracy difference between the XGBoost and the other models. The highest of the XGBoost model is 86 for the classifier Intuition vs Sensing.

**2.2** **Personality Evaluation and CV Analysis using Machine Learning Algorithm [2]**

The prediction of human personality using standard questionnaires that is provided by the HR Department according to job selection criteria. Candidates fill online Curriculum Vitae which can be later on viewed by Admin. Candidates are provided with separate set keys for attempting the aptitude and personality-based tests. CV analysis is

performed using the CV filled by the candidate in the website (Refer Fig. 5.2). A machine learning approach has been used in analysis of data through content and collaborative filtering. Further the test scores help in deciding the qualities in the candidates.

**2.3** **Personality Prediction through Curriculum Vitae Analysis Involving Password Encryption and Prediction Analysis [3]**

An efficient and effective approach is used to rank and evaluate candidates through psychometric analysis for calculating emotional quotient. Technical eligibility criteria from the online CVs and emotional aptitude by leveraging responses in evaluations are processed by the proposed system. The OCEAN model performs the linguistic and personality analysis of the candidates. 65% and 87% average accuracy levels were obtained by the algorithm dependent and independent approaches respectively.

OCEAN Model is used to assess emotional quotient and predict the personality traits. Machine learning techniques such as Logistic Regression are used for modelling the personality predictor. The details of the candidates are kept secure by using a password encryption algorithm. The passwords are only known to the required individuals.

**2.4** **Personality Prediction System Through the CV Analysis [4]**

The framework will work the human asset division with selecting right contender for specific employment profile which thusly gives master workforce to association. Administrator or the worry individuals can without much of a stretch waitlist a hopeful in light of their online test checks and can choose a fitting contender for wanted occupation profile.

This framework will naturally decide the key expertise trademark by characterizing every master’s inclination and positioning choices. The exhibited framework computerizes the procedures of necessities particular and candidate’s positioning. The proposed framework produces positioning choices that were moderately

exceptionally reliable with those of the human specialists. This framework will empower a more powerful approach to short rundown submitted hopeful CVs from an extensive number of candidates giving a steady and reasonable CV positioning strategy.

**2.5** **A Novel Approach to Evaluate and Rank Candidates in A Recruitment Process by Estimating Emotional Intelligence through Social Media Data [5]**

They have presented an efficient approach to evaluate and rank candidates in a recruitment process by estimating their emotional intelligence through the social media to data. The proposed system processes the technical eligibility criteria based on the entries made by the users in their online resumes and the applicants’ emotional aptitude by leveraging their presence in the social media, which have become a huge repository of the collective interests and opinions of its users, who are motivated to share information and receive feedback from friends and colleagues.

The Big Five personality model was used to perform the linguistic analysis. Since their method works by extracting meta-attributes from the text, this is less dependent on any particular language, and thus could be extended to regional languages as well. The approach was tested using data from three different datasets available in the literature and presented an average accuracy level of around 87% when using the algorithm independent approach, while the algorithm dependent approach showed a vastly inferior average accuracy of about 45%.

**2.6** **Personality Prediction with Social Behavior [6]**

Social behavior in online social networking sites can be used to predict User’s big five personality traits. Psychologist used to follow the personality questionnaire approach. The process that they proposed is costly and impractical at times. With the popularity of online social networks, researches envisaged to predict the personality automatically. Researches tried to assess the personality based on internet and social network site usage.

However only some of the personality traits like Extraversion and emotional stability could be assessed using this approach. Through linkage and content-based analysis of these online social networking sites data, researchers were able to predict personality traits quite accurately. Based on Facebook “Likes”, Network Structure like number of friends and groups, Status update, Photo upload, Tags and then using various regression and machine learning algorithms researchers were able to correlate these features with personality trait.

**2.7 Summary of all Papers**

Personality Prediction software can be built by using machine learning approach to make a decision of best candidates among all the applicants. This can be done by implementing a prediction algorithm which can predict the Best suited applicant for the given requirements of the company. This prediction algorithm may use TF-IDF algorithm, Ocean model for finding the best applicant. This approach matches the requirements of the company to the particular candidate’s CV. Thus, we can get the perfect candidate for our required work.

Another approach for prediction is the company can also form the online personality tests for checking the applicant’s interest which make their work easy for short-listing the candidates. The tests include Emotional aptitude test, Evaluation tests, Personality Test. The results of the test conclude the required candidates from huge pool of applicants. Perform psychometric analysis on the selected candidates for further evaluation. We can also use Big Five Personality model which predicts personality traits of the candidate. This approach detects persons personality and classifies in categories as Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism

**CHAPTER 3**

**Motivations and Objectives**

**3.1 Motivation**

1. Looking towards industrial growth, for enable a more effective way to shortlist submitted candidates CVs from a large number of applicants providing a consistent and fair CV ranking policy, which can be legally justified to job requirements.
2. The industry can use this system to choose expert candidate. By using this system, we can reduce time of HR department.
3. HR department handles this selecting process, for reducing their work as well as time we can use this software and will give the selected applications of candidates to HR department to make a choice of the right candidate.
4. This reduce departments work by reviewing the applications of an employee who is most eligible for their requirements.

**3.2 Objectives**

1. The system provides effective way to shortlist candidates CVs from a large number of applicants.
2. Framework reduces workload of Human Recourse department.
3. These Automate the process of Requirement specifications of the industry.
4. Reduces time requires for evaluation.
5. The personality traits and TF score generated gives candidates personal profile which helps further for personal Development.

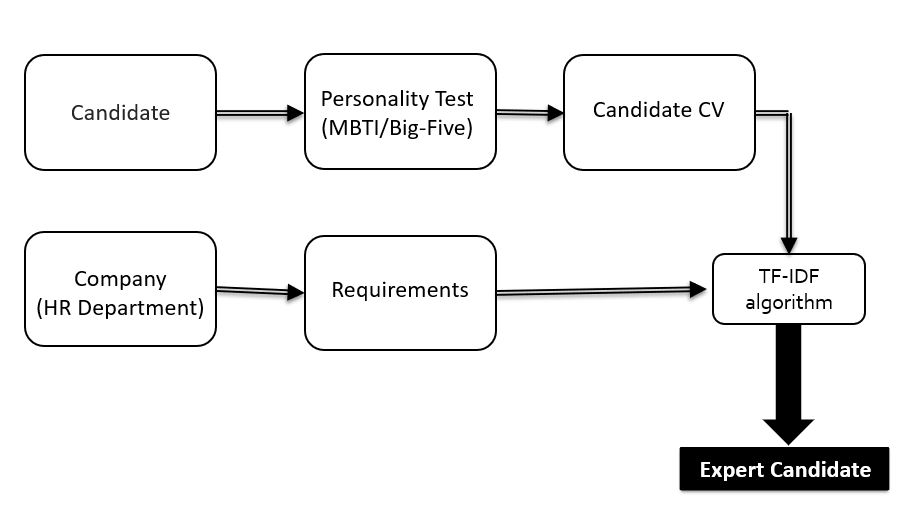
**3.2 Proposed Approach**

In our system, we proposed a system which can check the personality of the candidate and also the system can evaluate the candidates CV based on organization

requirement. This system provides with expert workforce for the organization which will help the HR department to select the expert candidate for the particular job profile.

In our society intelligence is high appreciated. If you have a high IQ, you have a better chance of being successful at professional life. For prediction of personality, psychometric question used. There are various personality prediction models are available. In this paper, we use MBTI test and Big-Five Personality Model to detect the personality of the candidate. This model gives some required score. System will automatically generate a threshold value so that candidates can evaluate based on this score.

After checking with personality, the candidate is requested to submit his/her CV. The organization requirements for a particular post are assign as input to machine learning algorithm for evaluation of candidates. Then each CV is analyzed by the algorithm and evaluate based on the requirement for the post. This algorithm also generates two scores, from which we can eliminate the rest of candidates. Thus, this system is more helpful for evaluation of candidate.



**Fig 3.1. Implementation Flow**

The proposed system is developed on the basis of the implementation flow shown in (Fig 1). The candidate has to undergo the Personality test. Followed by this process, a

candidate can submit their CV based on score from the test. The TF-IDF algorithm is used to perform the analysis as a graph in terms of the programming skills on x-axis and the respective scores on y-axis. These scores will help recruiters in selecting the right candidate.

**CHAPTER 4**

**Details of Design/Technology/Analytical and/or Experimental Work**

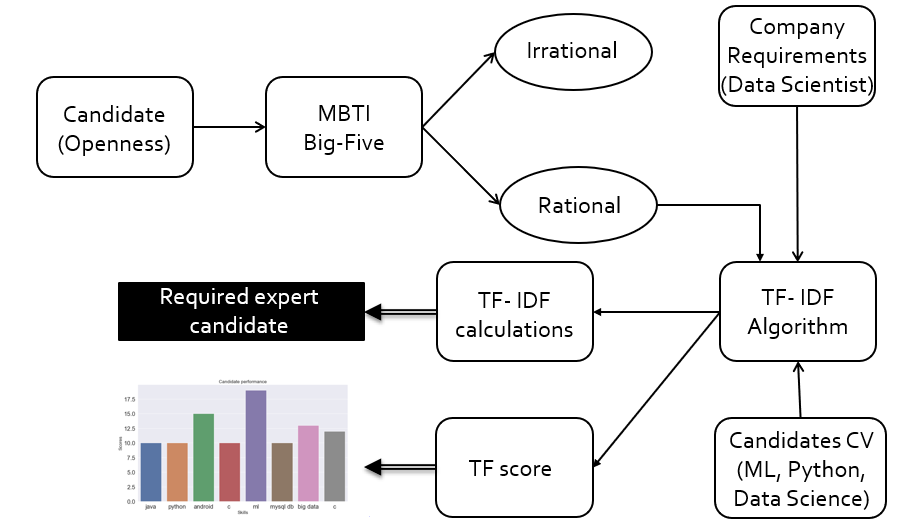
**4.1 Method**

If you have a high IQ, you have a better chance of being successful at school and professional life. For prediction of personality, psychometric questions are used. There are various personality prediction models are available. In this report, we use MBTI test and Big-Five Personality Model to detect the personality of the candidate. This model gives some required score. System will automatically generate a threshold value so that candidates can evaluate based on this score.

As depicted the process start with **checking the personality of the candidate** and his personal interest by using MBTI test and big-five model. We can achieve this by doing the personality test on individual candidate. We can make a certain limiting score for this test and if he passed the test then he we will submit the cv, which is followed by **removing unnecessary symbols**, numbers and stop words from the data or text and chopping the words down to the root words and creating these words in to numerical data using the **TF-IDF vectorizer.** The TFIDF Vectorizer is used to encode the data and create the feature vectors. After that we can pass that feature vector to train the model after doing the hyper parameter training to find out the best parameters and those values that can affect the training process. The last step of the process being generating the **personality traits** which is also that is to be generated from the model using TF score. Each CV is analyzed by the algorithm and evaluate based on the requirement for the post. This algorithm also generates two scores, from which we can eliminate the rest of candidates. Thus, this system is more helpful for evaluation of candidate.

**4.2 Architecture**

The Architecture System (Fig 2) shows sequential flow of the workflow. These will show us the basic idea about the interaction of candidate, organization and our proposed system (CV analysis).

****

**Fig 4.1. Architecture System**

Below, I discussed about these three fields-

1. Candidate

- Add recruiter request to system

- Perform Personality test

- View score

- Attach CV

- View application selection result

1. Organization

- Decide threshold score for personality test

- Add requirements for post

- Decide threshold for TF-IDF

- Decide the candidate result

1. CV analysis

- All CVs stored in database

- CV analyze by TF-IDF algorithm

**4.3 Algorithm One**

The algorithm used for predicting personality.

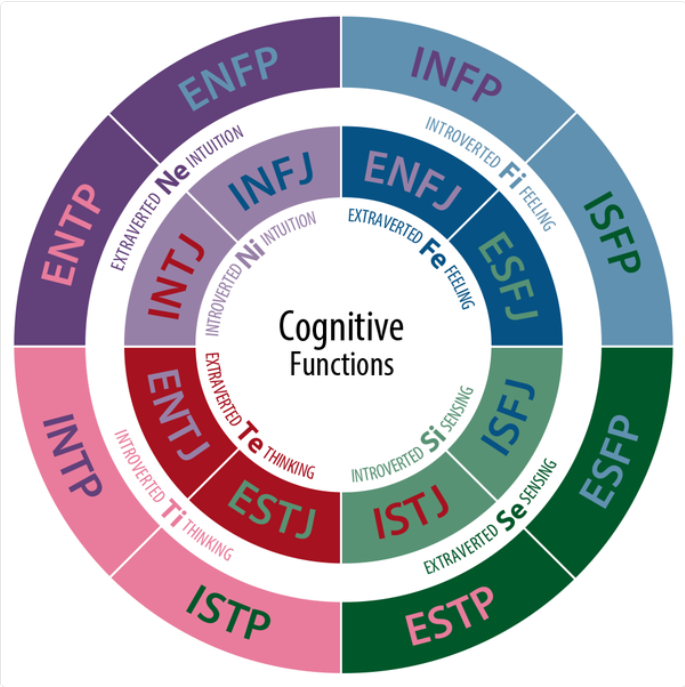
1. **MBTI Test**

Myers-Briggs Type Indicator (MBTI) it is a test carried in physiological study which can determine the person personality in 16 unique types. This theory provides output in two functions, first one is rational which is related to thinking and feeling and other one is irrational which relates sensation and intuition. Thus, for our case we can focus on the relational function score and evaluate the candidate. The goal of the MBTI is to further explore candidate on their personalities including their likes, dislikes, strengths, weaknesses, possible career preferences, and compatibility with other people.

MBTI measures people's preferences for where they get their energy (extraversion vs. introversion), how they handle information (sensing vs. intuition), how they make decisions (thinking vs. feeling), and how they structure the outside world (judging vs. perceiving). Someone's type, a combination of four letters (such as IMFJ or ENTP), is a shorthand way of understanding how that person tends to take in information and make decisions.

Personality psychology boils down to this: we each have certain personality traits, and we try to identify and measure those traits and predict how they affect our behavior and attitudes. If you'd like to find out what your MBTI type is,[16 Personalities](https://www.16personalities.com/) offers a free test and breakdown of each personality type (including workplace habits), using labels like "advocate," "virtuoso," and "logician."

Here is of 16 designations in full -



**Fig 4.2. MBTI personalities**

1. **Big-Five Model**

Big Five model is also a personality prediction model which predict based on personality test and gives the personality in 5 different categories. It Perform psychometric analysis on the selected candidates for further evaluation. This approach detects persons personality and classifies in categories as Openness, Extraversion, Agreeableness, Neuroticism and Conscientiousness.

We can make a certain limiting score for this test and if he passed the test then he we will submit the cv, which is followed by **removing unnecessary symbols**, numbers

and stop words from the data or text and chopping the words down to the root words and creating these words in to numerical data using the **vectorizer.**

The Big Five model is helpful for describing personality differences. People who have high agreeableness, for example, are more likely to cooperate and go along with group decisions, whereas those low in this trait would be more competitive and go against the grain. If you're high in conscientiousness, you're likely organized and have good impulse control, as opposed to those who fly by the seat of their pants. The last step of the process being generating the **personality traits** which is also that is to be generated from the model.

**4.4 Algorithm Two**

The algorithm used for CV analysis.

**TF-IDF Algorithm**

TF-IDF algorithm is used to evaluate submitted CVs of candidates. Before applying this algorithm, we can use some pre-processing techniques like text reorganization, removal of stop words, data cleaning, etc.

The algorithm can take bunch of requirements from the company which will add as input, then it will further search for those particular requirements in submitted CVs and shortlist them for interview purpose. The algorithm can also analyze the individual candidates CV and score them accordingly.

1. TF- Term Frequency

Number of times a keyword appeared in a document is calculated by Term Frequency

TF = No. of times keyword appeared in document

Total no of keywords in document

1. IDF- Inverse Document Frequency

Used to priorities the keywords in the required document.

IDF = log Total no of CV

No of CVs with required keywords

1. TF-IDF Weight

Weight = TF \* IDF

Higher the weight more relevant is CV. This step gives the higher and lower weights of all the CVs, based on this we can further classify them in different groups.

This calculation helps to find the most specific and accurate candidate who is eligible for the post. The high-frequency of some keywords may impact on candidate overall score.

By using TF score we can calculate individual score of the candidate. The scores and Skills are taken into consideration for overall evaluation and analysis of the CV.

**4.5 Experimental Work**

In our system, we proposed a system which can check the personality of the candidate and also the system can evaluate the candidates CV based on organization

requirements. This system provides with an expert right workforce for the organization which will help the HR department to select the right candidate for the particular job profile. There are various personality prediction models are available. In this paper, we use MBTI test and Big-Five Personality Model to detect the personality of the candidate. This model gives some required score. System will automatically generate a threshold value so that candidates can evaluate based on this score.

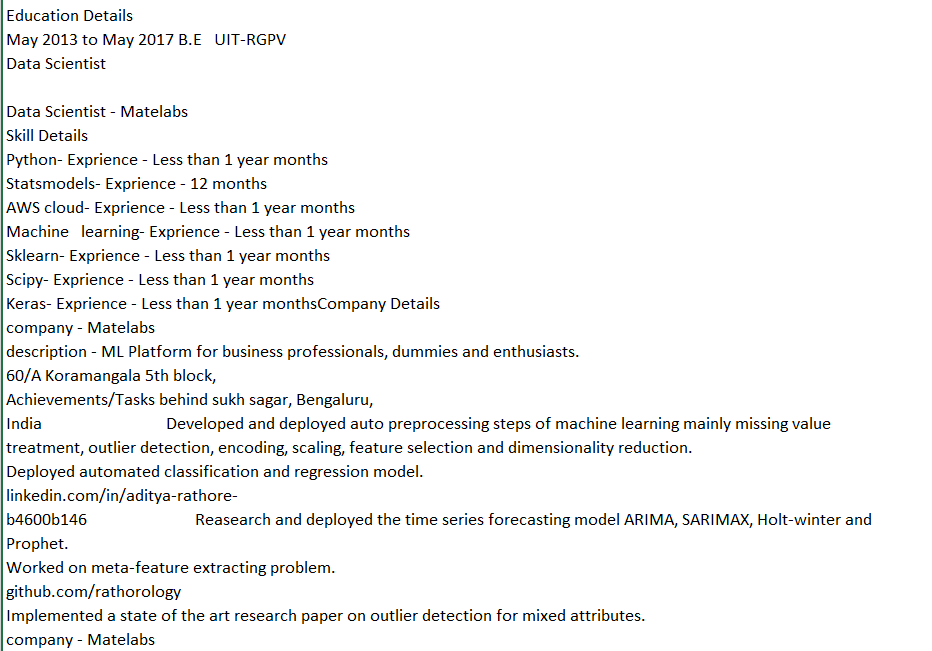
After checking with personality, the candidate is requested to submit his/her CV. The organization requirements for a particular post are assign as input to machine learning algorithm for evaluation of candidates. Then each CV is analyzed by the

algorithm and evaluate based on the requirement for the post. This algorithm also generates two scores, from which we can eliminate the rest of candidates. Thus, this system is more helpful for evaluation of candidate.

**4.6 Implementation**

For the implementation purpose I used the natural language toolkit (nltk) and for TF-IDF I used the sklearn feature extraction libraries. First, I took a sample CV as an input. I use preprocessing techniques for removing stop words from the data, data cleaning, text reorganization, etc.

Sample CV-

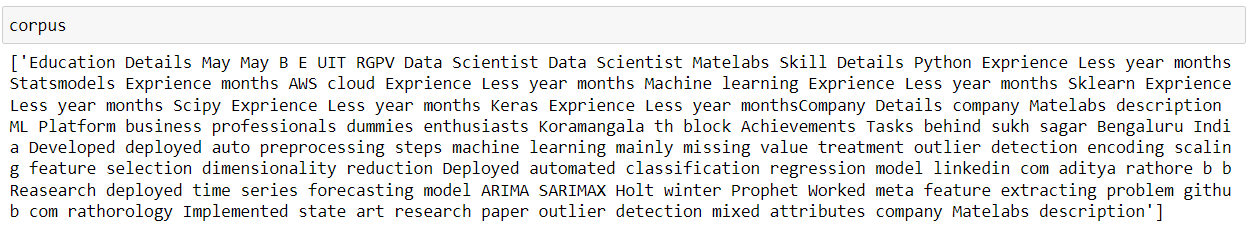


**Fig 4.3. Sample CV**

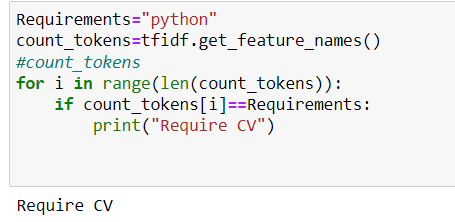
**4.6 Result- TF-IDF algorithm**

This proposed a system that can evaluate the CV. System will reduce the workloads of HR department and also reduces the overall time for evaluation of candidate. This automated the process of evaluation of candidate. The candidate can identify the Personal Interest from the personality test. Time efficiency is the main advantage of this system.

First, I took a sample CV as an input. I use preprocessing techniques for removing stop words from the data, data cleaning, text reorganization, etc.



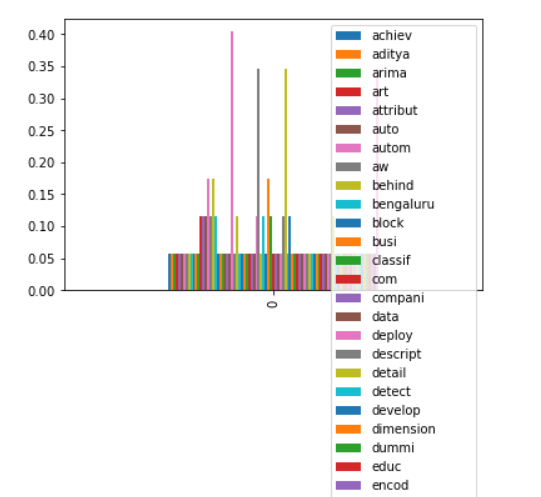
After cleaning of data, I give these data as an input to TF-IDF vectorizer for chopping of words into root words and for finding weather the gives required words (Requirements of company) are satisfying by the particular candidate CV or not. If the word is present in the CV, then it is required one, else we can evaluate the CV.



**Fig 4.4. Output**

As our sample CV contains the python skill so that the result becomes **Require CV.**

The TF-IDF also calculates the TF score from that we can get a particular candidate profile and also, we can find in which field the candidate is Expert.



**Table 4.1. TF Score**

**4.8 Advantages**

1. Admin can easily shortlist the candidate
2. Reduce workloads of HR department and Evaluation time
3. Automate process of Requirement specifications
4. Personal Development

**Disadvantages**

1. Requires vast storage space because of information that is to be stored identified with CV
2. Should have a dynamic Web association
3. May give off base outcomes if information not entered legitimately

**CHAPTER 5**

**Conclusions and Future Work**

**5.1 Applications**

The system is efficient and effective approach to rank and evaluate the candidate. This framework will decrease workload of the human asset office. This proposed system can also be implemented in real time applications like career adviser. As system calculates the personal profile it can also be used for personal development for individuals.

**5.2 Future Scope**

Further, I can modify the existing system to perform analysis of social media data and decide whether the candidate socially active or not, other features. I will add more classification algorithms of machine learning can be integrated to provide much better performance. We can further, increase the efficiency of the system.

**5.3 Conclusion**

We have presented in this report, the prediction of human personality by using MBTI test, Big-Five model standard TF-IDF algorithm that is provided by the HR Department according to the job selection criteria. The system is efficient and effective approach to rank and evaluate the candidate. This framework will decrease workload of the human asset office. This proposed system can also be implemented in real time applications like career adviser.

The prediction of human personality by using MBTI test, Big-Five model standard TF-IDF algorithm that is provided by the HR Department according to the job selection criteria. The system is efficient and effective approach to rank and evaluate the candidate. This framework will decrease workload of the human asset office. This proposed system can also be implemented in real time applications like career adviser.

**References**

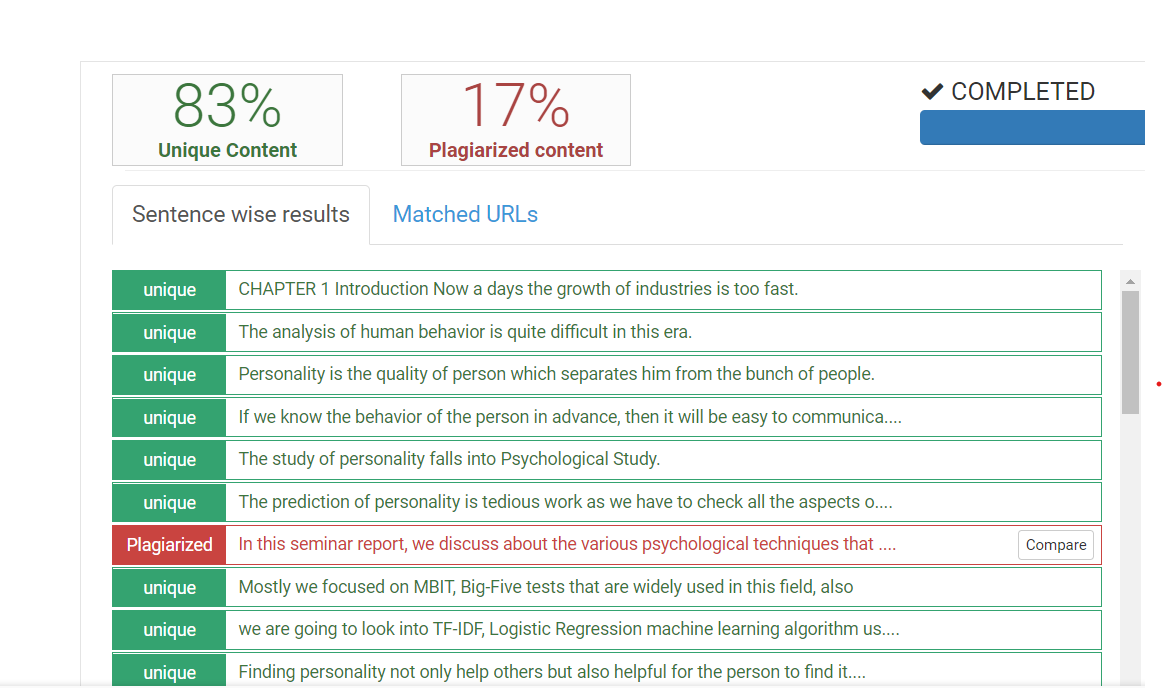
1. Ms. Kavita Agarwal Assistant, Ms. Vimala Manohara Ruth Assistant, Kedarnath Chaturvedula, Vishal Chandra Jongoni “Analysis of Human Traits Using Machine Learnin”- 2017.
2. Sudhir Bagade, Pooja Yede, Nirmiti Patil“Personality Evaluation and CV Analysis using Machine Learning Algorithm”-2018
3. Alakh Arora and N. K. Arora “Personality Prediction System Through CV Analysis”- 2018
4. Gagandeep Kaur, Shruti Maheshwari 1 Assistant Professor, IT Department, Symbiosis Institute of Technology, Pune “Personality Prediction through Curriculum Vitae Analysis Involving Password Encryption and Prediction Analysis”- 2019
5. Vishnu M Menon Computer Science and Engineering Department TKM College of Engineering Kollam, India vishnummenon@computer.org H A Computer Science and Engineering Department TKM College of Engineering Kollam, India “A Novel Approach to Evaluate and Rank Candidates A Recruitment Process Estimating Emotional Intelligence through Social Media Data”- 2016
6. Ashish Kumar Tripathi Department of Applied Computer Science and Society University of Winnipeg “Personality Prediction with Social Behavior by Analyzing Social”- 2017
7. FoDRA – Nikolaos D. Almalis George Tsihrintzis ,Aggeliki , “A New - Content-Based Job Recommendation Algorithm for Job Seeking and Recruiting”, 2016.
8. Data, Vishnu M Menon Computer H A, “A Novel Approach to Evaluate and Rank Candidates in A Recruitment Process by Estimating Emotional Intelligence through Social Media”,May 2016.
9. Manasi Ombhase, Prajakta Gogate, Tejas Patil “Automated Personality Classification Using Data Mining Techniques” 10.13140/ RG.35949.59363- 2018
10. S. Amdouni and W. Abdessalem, "Web-based recruiting", Proc. of

International Conference on Computer Systems and Applications ICCSA), 2010, pp. 1-7.

1. P. De Meo, G. Quattrone, G. Terracina and D. Ursino, “An XML-based multi-agent system for supporting online recruitment services,” Systems and Humans, vol. 37, July.
2. Ramezani, M., Bergman,, , L., Thompson, R., Burke, R. and Mobasher, B., “Selecting and applying recommendation technology”. In Proc. of International Workshop on Recommendation and Collaboration, in Conjunction with 2008 International ACM Conference on Intelligent User Interfaces. Canaria, Canary Islands, Spain, 2008.
3. Shereen Albitar, Sebastien Fournier, Bernald Espinasse, An effective tf/idf-based text-to-text semantic similarity measure for text, spinger, pp. 105-114, 2014.
4. Stephen,"Understanding inverse document frequency: on theoretical arguments for IDF", Journal of, Vol. Stephen Robertson,"Understanding inverse document frequency: on theoretical arguments for IDF", Journal of Documentation, Vol. 60 Issue: 5, pp.503-520, 2004.
5. R.S. Walse, G.D. Kurundkar, P. U. Bhalchandra, "A Review: Design and Development of Novel Techniques for Clustering and Classification of Data", International Journal Scientific Research Computer Science and Engineering, Vol.06, Issue.01, pp.19-22, 2018. Bhalchandra, "A Review: Design and Development of Novel Techniques for Clustering and Classification of Data", International Journal of Scientific Research in Computer Science and Engineering, Vol.06, Issue.01, pp.19-22, 2018.
6. Shahzad Qaiser,Ramsha Ali ,International Journal of Computer Applications (0975 – 8887), “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents” Volume 181182 – No.1, July 2018.

**Annexure 1**

**Plagiarism Check Report**



**Fig. Plagiarism**