iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34335

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

Automated Personality Classification for E-Recruitment on the basis of CV using Machine Learning

Sakshi Kumar[1], Amisha Ghevade[2], Bhushan Patil[3]
[1, 2, 3]*Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, India*

Abstract: ***Personality is an important parameter as it differentiates various individuals from each other. Personality prediction is an evergreen area of research. Predicting personality with the assistance of data through social media may be a promising approach as this method doesn't require any questionnaires to be filled by users thus reducing time and increasing credibility. Due to the dynamic development of job markets, traditional recruitment methods are becoming insufficient. This is because recruiters often receive a huge number of applications (usually unstructured resumes) that are difficult to process and analyze manually. In this work we present a unique approach for evaluating job applicants in online recruitment systems, using machine learning algorithms to solve the candidate ranking problem and performing semantic matching techniques. We propose a system that classifies the personality of the candidate is classified using Myers-Briggs type indicator and the job prediction is done using DNN algorithm.***

Keywords: ***Personality Prediction, Myers–Briggs, e-recruitment***

I. INTRODUCTION

As far as employment is considered, selecting the right candidate for the recruitment process from a huge pool of candidates has been a fundamental issue. Recruitment is considered among the foremost challenging functions for job portals and human resource departments. this is often because employers often receive a vast number of resumes – a number of which are uploaded as unstructured documents in several formats. Conducting different eligibility evaluation tests, interviews, and group discussions have been traditional techniques. These traditional practices are considerably time-consuming, and should end in unfair choices of candidate. As compared to traditional recruitment process, if an internet selection process is conducted, then an honest selection of the candidate is possible. This process usually takes weeks or even months to end, so jobseekers are left hanging until a choice is made a way or another. additionally, if rejected, the candidate is typically not given a reason for the rejection – they have to figure out for themselves what they were lacking. So, to beat the above-mentioned limitations, we propose an automatic online recruitment system that exploits multiple semantic resources in an effort to spotlight and capture the semantic aspects of both job posts and candidate resumes. The proposed system is 2 sided: it'd be candidate oriented and organization oriented. Since we all know that personality is that the most vital factor which reflects a private which keeps on varying. So, our system will simply scan the submitted CVs using Machine Learning Algorithm. The personality of the candidate would be identified by using two metrics, first is that the test series and second is CV analysis. the corporate is responsible to style or update the questions and has the entire control to customize the questions as per organization requirements. the choice is usually made on the thought of the test outcome. The candidate here will register him/herself with all its details and may upload their CV into the system which may be further used. additionally, we'll also extract the info from the candidates Facebook and LinkedIn profile. After we extract the info, we'll be applying Random Forest Algorithm and DNN Algorithm. Machine learning based classification techniques with similarity functions are wont to find most relevant resume. Random Forest classifier performed best for our case compared to a different ML classifier.

II. OBJECTIVES

A. To develop a system to provide a more effective way of short-listing the candidates.
B. To determine the key skill characteristic by defining each expert's preferences and rankingdecisions.
C. To automate the process of requirement specifications and applicant's ranking.
D. To conduct online aptitude, personality and technical test.
E. To produce ranking decisions that would have relatively higher consistency than those ofhuman experts.

III. METHODOLOGY

This project we will develop using python and web technology. For fetch the data from Facebook and LinkedIn we will use selenium which is automation tool. We can classify the Facebook fetched data using Random Forest algorithm. And we will apply DNN algorithm on the LinkedIn Fetched data. For detecting the user's skills, we will take small test and on the basis of result we will show the recommended skills which is useful for predicting the job profile. In this system, we will verify the resume of the applicant's which will have to upload and will compare fetched data from social networking sites with described resume skills and experiences. For this purpose, we take the twitter's dataset and generates the model using machine learning algorithm. And after whatever we get skills from the Facebook and LinkedIn, we will these skills as an input and predict the result. The flowchart of the system is given below:

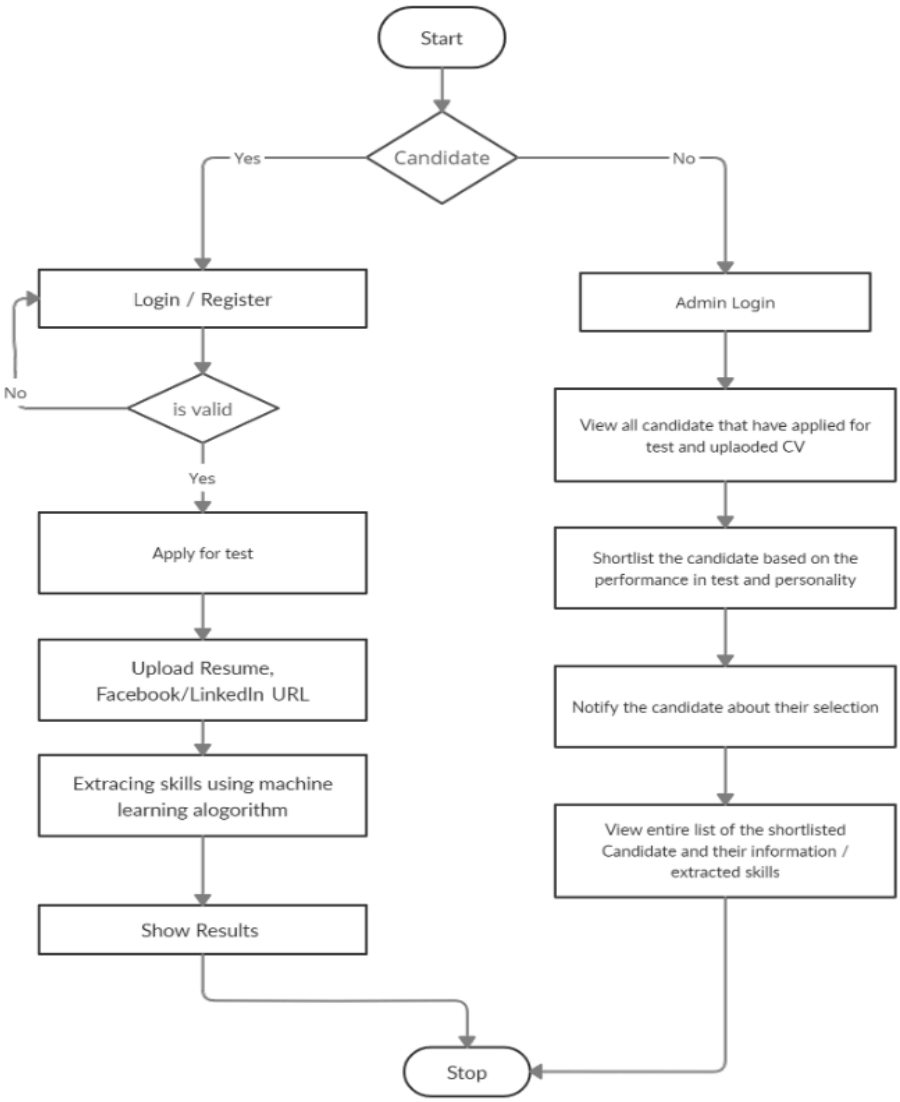


Fig 1. Flow chart of the system

The whole system process can be divided into 7 parts:

A. Admin Module

The admin can login with their username and password. They can conduct the test series. Admin has the right to add questions for the test along with alternatives and can also modify or delete the questions as per the requirement and can view candidate results and shortlist list them as per the requirement and the test score.

B. Candidate Module

First the candidate needs to register/login into the system. Candidate will be directed to the test series page which will consist of personality, technical and aptitude questions. Further the user's Facebook and LinkedIn profile data will be extracted using selenium and Beautiful Soap tool. DNN algorithm is applied to the data fetched from the LinkedIn profile and to extract skills from resume and Random Forest Algorithm is applied to the Facebook data to classify the candidate's personality.

C. Development Tools

For the development process we have used the natural language processing toolkit (NLTK) and Flask framework. NLTK is a leading platform for building Python programs which provides easy-to-use interfaces for text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning. Flask is a micro web framework written in Python. Pandas, numpy, re, keras, seaborn, matplotlib and sklearn are other Python librariesthat were used.

D. Dataset for Training the Model

In this research we have used dataset for Myers–Briggs personality type from Kaggle which contains 8675 rows of data. In this dataset, each row consists of two columns. The first column is for the MBTI personality type of a given person, and the second column includes fifty posts obtained from the individual's social media. The second dataset is the 25 clean job post with their respective job descriptions and the datahas been collected from online forums.

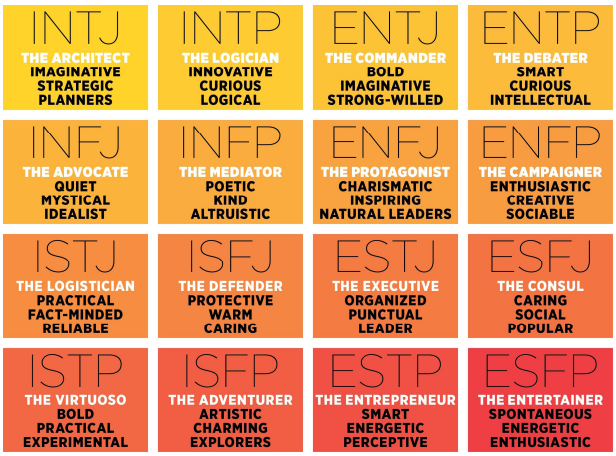


Fig 2. Myers-Briggs personality types

E. Proportionality in Dataset

In this step, seaborn which is a Python data visualization library and matplotlib which is a Python 2D plotting library were used for data preview and to determine the distribution of the MBTIpersonality types in the dataset.

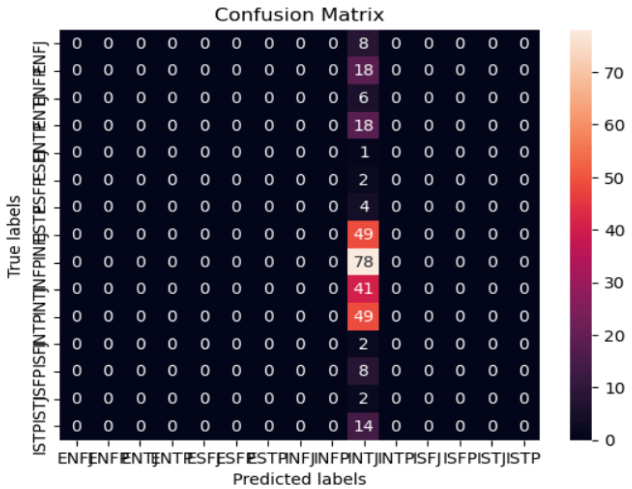


Fig 3. Confusion matrix of MBTI types

F. Classification

The classification was done using four different models and their accuracy score was recorded.

Table 1. Accuracy results

Classifier	Accuracy
Random Forest	0.78
Multinomial Naive Bayes	0.44
Logistic Regression	0.62
Linear Support Vector Classifier	0.68

G. Personality Prediction using Random Forest Algorithm

1) Random Forest Creation Pseudocode

a) Randomly select "k" features from total "m" features where k < m

b) Among the "k" features, calculate the node "d" using the best split point.

c) Split the node into child nodes using the best split.

d) Repeat 1 to 3 steps until "l" number of nodes has been reached.

e) Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

2) Random Forest Prediction Pseudocode

a) Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target).

b) Calculate the votes for each predicted target.

c) Consider the high voted predicted target as the final prediction from the random forest algorithm.

IV. RESULT S AND DISCUSSIONS

The system can be used in many business sectors that may require expert candidate. This system will reduce workload of the human resource department and select right candidate for particular job profile which in turn provide expert work-force for the organization. The test conducted through the Test series is reliable and accurate and will be used by multiple organizations and educational institutes worldwide. Admin of the concern person can easily short-list a candidate based on their online test marks and can select an appropriate candidate for desired job profile.

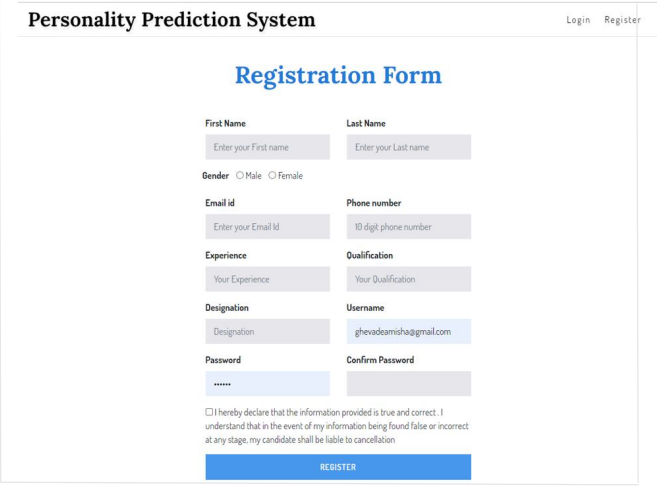


Fig 4. Registration page

Fig4 is the registration page where the candidate first has to register themselves and enter a unique username, password and Email-ID.

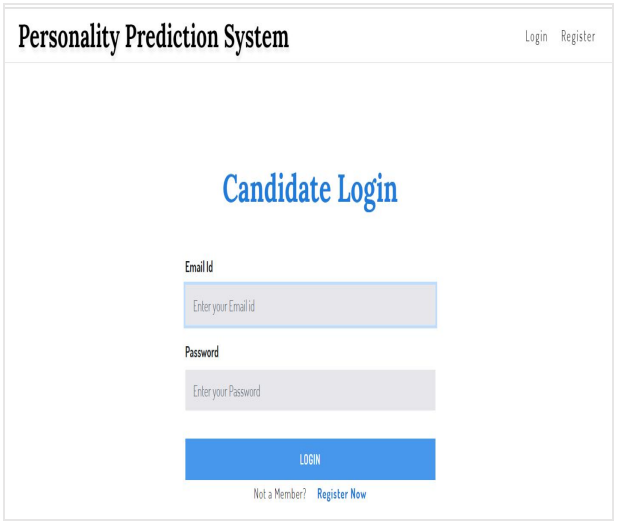


Fig5. Login page

After successful registration the candidates needs to login with their Email-ID and Password as shown in the Fig5.

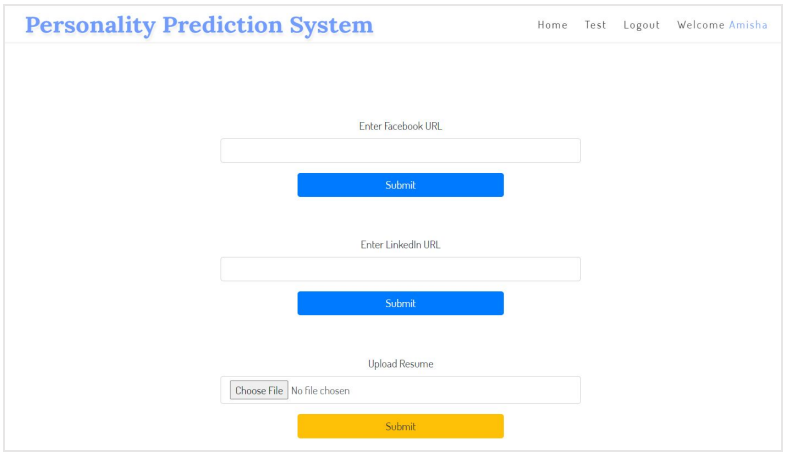


Fig 6. UI page

This UI page basically gains access to the candidate's social media handles and resume. First candidate will enter the Facebook Link followed by LinkedIn link and further upload the resume and using the Machine Learning Algorithms it will extract the data / features.

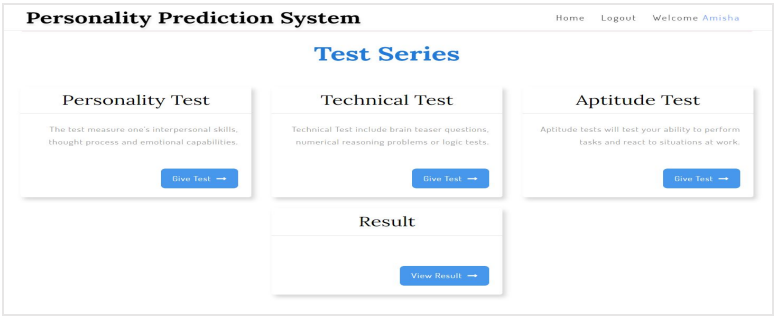


Fig 7. Test Series

This is the test series page where the candidate have to give personality, technical and apptitute test. All three test are mandatory to attend.

Fig 8. Candidate list

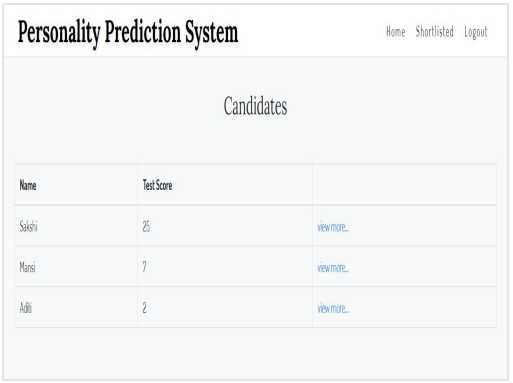


Fig 9. Shortlisted candidate

First the admin needs to login with valid credentials and after successful login they can view the entire list of candidates with their respective marks from test who has given the test and uploaded their CV as shown in Fig 8.
Based on candidate's performance and job requirement, admin can shortlist them and above figure is the list of shortlisted candidates.

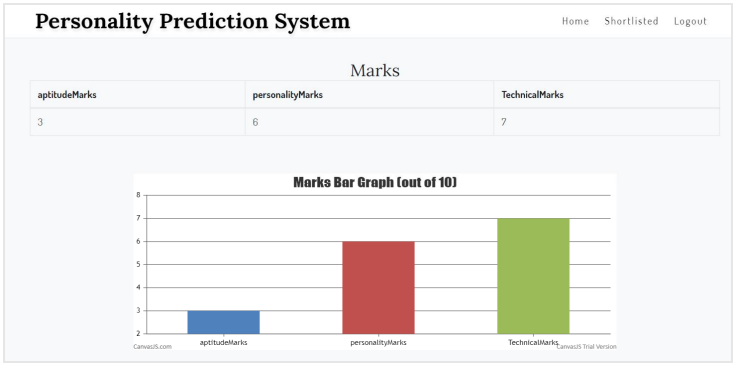


Fig 10. Results

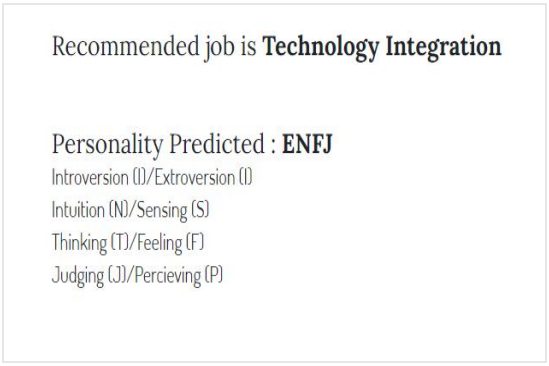


Fig 11. Recommended Job

On clicking on the view more option admin can see the marks in graphical format of that particular candidate along with the job Recommended.

V. CONCLUSION

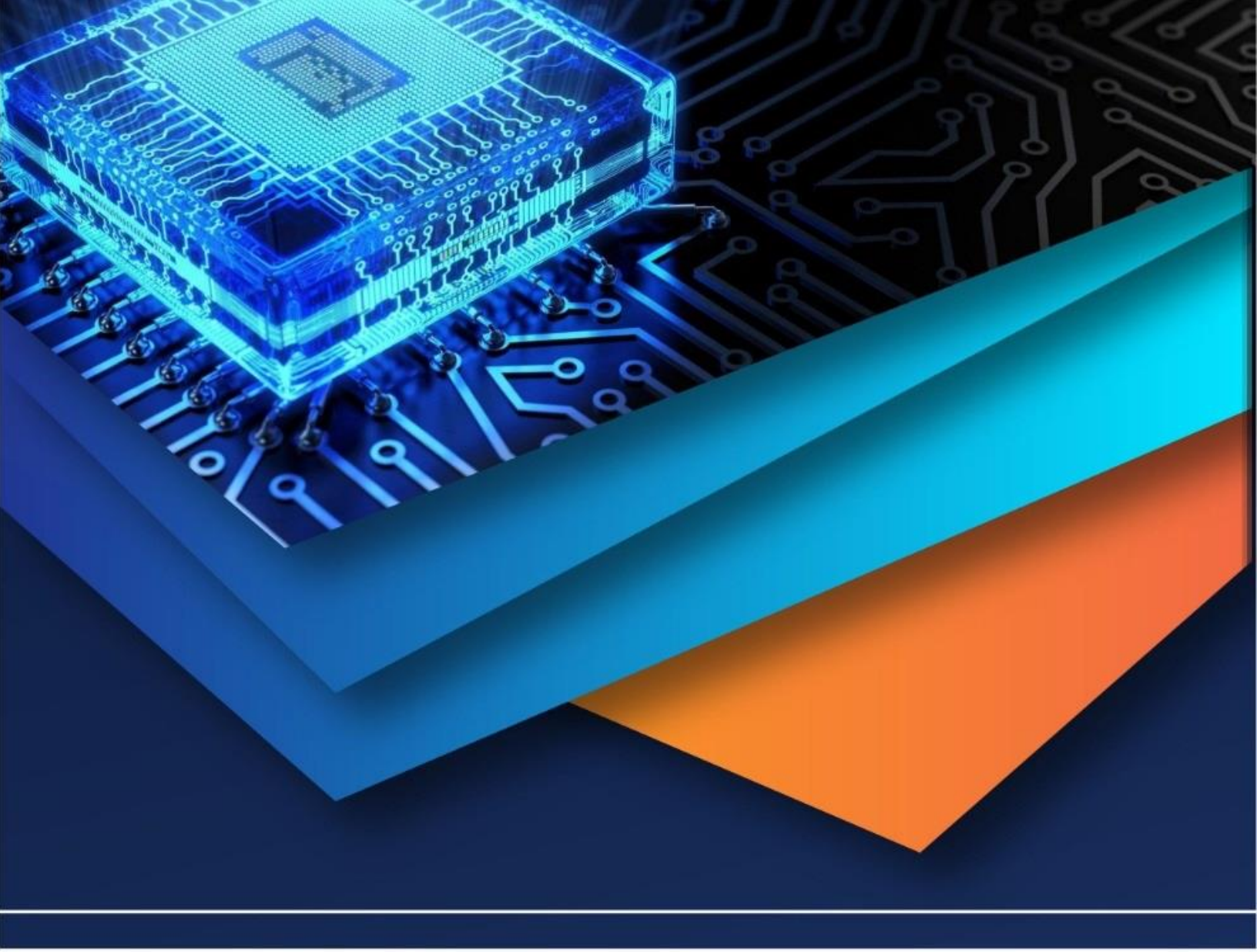
The system suggests the appropriate candidate based on their skills, experience and personality and thereby reducing workload of the human resource department. The Personality Test conducted through the Test series is reliable and accurate, as it will be based on a standardized personality test used by multiple organizations and educational institutes worldwide. The data extracted from the candidates LinkedIn and Facebook profile will add up the accuracy of the system for the classification process. Huge number of applications are received by the organization for every job post. The process of classifying the candidate's resume is manual, time consuming, and waste of resources. To overcome this issue, we have proposed an automated machine learning based model which recommends suitable candidate's resume to the HR based on given job description. The proposed approach effectively captures the resume insights, their semantics and yielded an accuracy of 78.53% with Linear SVM classifier. If an Industry provides a large number of resumes, then Industry specific model can be developed by utilizing the proposed approach. The system would be used in many business sectors that will require expert candidate, thus reducing theworkload on the HR department.

VI. ACKNOWLEDGEMENT

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal and Prof. S. P. Khachane, H.O.D of Department Computer Engineering of Rajiv Gandhi Institute of Technology for providing us an opportunity to do our project work on "Automated Personality classification for e-recruitment on the basis of CV using Machine Learning". This project bears on the imprint of many people. We sincerely thank our project guide Mr. Bhushan Patil for his guidance and encouragement in carrying out this synopsis work. Finally, we would like to thank our colleagues and friends who helped us in completing project work successfully.

REFERENCES

[1] A Kmail, M Belkhatir and S Alhashmi, "An Automatic Online Recruitment System based on Exploiting Multiple Semantic Resources and Concept-relatedness Measures", International Conference on Tools with Artificial Intelligence, 2015 IEEE
[2] A Zaroor, M Maree and M Sabha, "JRC: A Job Post and Resume Classification System for Online Recruitment", International Conference on Tools with Artificial Intelligence, 2017 IEEE
[3] E Faliagka, L Iliadis and I Karydis, "On-line consistent ranking on e-recruitment: seeking thetruth behind a well-formed CV", Issue: 07,2013, Springer
[4] M He, D Shen, Y Zhu and Z Zhang, "Career Trajectory Prediction based on CNN", 2019 IEEE.
[5] M Maree, A Kmail and M Belkhatir, "Analysis and shortcomings of e-recruitment systems: Towards a semantics-based approach addressing knowledge incompleteness and limiteddomain coverage", Journal of Information Science, Issue: 2018, Springer
[6] V. Kumaran and A. Sankar, "Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT)", Int. J. Metadata, Semantics and Ontologies, Vol. 8, No. 1, 2013



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY