

Applicant Personality Prediction System Using Machine Learning

M.Karnakar
CSE Department
Vardhaman College of Engineering,
Hyderabad, India
karnakarmarla933@gmail.com

NageswaraRao Sirisala
CSE Department
Vardhaman College of Engineering,
Hyderabad, India
nagsirisala@gmail.com

Haseeb Ur Rahman
CSE Department
Vardhaman College of Engineering,
Hyderabad, India
haseeb171599@gmail.com

A B Jai Santhosh
CSE Department
Vardhaman College of Engineering,
Hyderabad, India
anumolusanthosh368@gmail.com

Abstract-Recruiting departments in larger organizations are having difficulty in selecting the right candidate because the number of applications for required job profiles are very huge in number. Also, the hiring process takes lot of man force and time management. To overcome this problem we are building a model or tool that will be very helpful to HR department in a particular organization to hire or select a right candidate for desired job profile. The proposed system, Applicant Personality Prediction System Using Machine Learning (APPS) will be predicting the personality of the candidate based on the candidate's resume details and also by using some personality tests. Personality prediction system selects the right and suitable candidate required for desired job profile by grading the resume on certain criteria like experience, skills etc. This system can also be used in other fields where they need to hire candidates by filtering processes in a very less time. Our application uses Spacy module to analyze, summarize and compare to get similarity score between resume and job description. We will be considering some other aspects like conscientiousness, openness, agreeableness etc. through which the personality of user can be predicted. For these characteristics we will give a situation so that user will be able to rate himself on a scale of 1 to 10. We will give this personality test scores to machine learning model which is developed using multinomial logistic regression which will predict the personality of job seeker. These similarity score and personality traits scores will be used for shortlisting candidates.

Keywords-ML: Machine Learning, Natural Language Processing, Spacy, Multinomial Logistic regression, TF-IDF Vectorizer: Term frequency-inverse document frequency, Cosine similarity.

I. INTRODUCTION

This system will be helping to shortlist the candidates in an effective manner from a large number of applications that were submitted. The system will be grading the resumes based on the required skills and other aspects required for desired job profile[3]. The Human Resource department will be saving time by this model since this model provides how much percent of resume is matching to the given job profile. Thus in this way it will be easy for HR department to filter the students based on resume matching percentage with job description. Candidate willing for job shall need to register into this model

with other required details and shall also provide their updated resume to the model and further the provided resume will be used by system or admin to match it with the job description. While registering, the candidate also needs to take test where few question [1][2] on personality and once the test is successfully completed, they can view the results in the form of graphs. According to [4], the results of personality test are further given to ML model to judge or predict the personality of a candidate.

II. RELATED STUDY

In this section we are discussing machine learning technique procedures and Natural Language Techniques to understand APPS. Some of the recent and relevant papers are also discussed.

A. Machine Learning Procedure

According to [11], this chapter explains about the fundamental understanding of machine learning algorithms such as supervised, un-supervised, and reinforcement learning algorithms. The study, then transitions to discussion of supervised machine learning algorithm such as Multinomial Logistic Regression [5]. This technique can be used to classify the Personality of an applicant.

Every algorithm based on machine learning will follow these techniques.

Preprocessing: The database or dataset which we use contains NaN values where NaN indicates 'not a number' values. The program which we use cannot process these non-numerical values, so it is mandatory to covert these values to numerical values[10]. The approach followed is, the NaN values are replaced by the mean of the column.

Splitting: The data in the database is divided into 2 types i.e., one is training and other is testing data sets. Training set is of 80% data and testing set is of 20% data.

Classification: The selected data from the database under training set is trained with algorithm like multinomial logistic regression[7].

B. Natural Language Processing

According to [12] documents similarity can be classified into three categories mainly string based, corpus base and knowledge based. This work utilizes term-based similarity measure nothing but cosine similarity to find document similarity with each other. *Natural language processing*: It is a field of AI in which through various technologies the system is able to analyze, understand and derive the information in our language with ease and smart ways.

By using NLTK and Spacy modules: We can perform text cleaning and preprocessing[8].

Vectorization: TF-IDF vectorization on bag of words[9].

Cosine similarity: For finding similarity between two documents[12].

C. Big five personality characteristics

According to [6], personality is something through which user characteristics are been identified by admin or some another person. These personality traits are mostly recommended for personality theories in scientific and business communities. The Big five characteristics based on [1]:

- Openness
- Conscientiousness
- Neuroticism
- Extraversion
- Agreeableness

According to [16], where the main idea of developing such model was to predict the personality of a social media user based on different aspects. According to the paper, different people having different personalities where related in the formation of unique and different connections on social media and were having different behaviours with each other. This particular project or model was tested on the people of very famous social networking application Facebook.

From[6], personality of a user was predicted form gathering the user's information provided on him on social media platform. It was because on a large scale everyone in world is much active on social platform than in his real life. So taking this as an advantage the system was made to predict the personality by gaining or accessing the user information from social media platforms. This was one of the finest way to judge as the user wouldn't be having any questionnaire which would trouble him. Instead the users will be happy when the personality was predicted without asking any question to user. Here they were using the datasets of active social media platform Twitter where every individual will be sharing his experience and thoughts just like Facebook.

According to [14], artificial intelligence based study which aims to automatically detect personality traits from video by using machine learning and image processing techniques is explained. To determine the character and emotion of people

by analyzing their behaviour is already an important task for especially human resources. If a person's body language is adequately analyzed, much information can be obtained about that person's personality traits. The main task of this study is to predict various dimensions model from video files by using techniques like artificial intelligence and neural networks. In this study, the videos and emotional aspects of user were collected and were utilized and an artificial intelligence based system was been developed to predict the personality of user from the videos automatically.

In [8], its our personality which effects our career from our start to our goals or achievements. It is more easy to design a model which will be much more efficient and can develop efficient strategies by personality traits. In this work, people's personality behaviour was predicted by their social media accounts. Personality prediction is currently an emerging field in research area. Personality can be defined as a psychological model that explains different type of human behaviour with the help of individual characteristics. Automatic personality prediction is something where the personality or behaviour of the user can be predicted using videos, speech, etc. Even though different type of such models were developed in the past few years but it is still difficult to know users current state of art.

As mentioned in [11], defines personality as set of perception, feeling and behavioural patterns that develop from root factors. In general till now personality hasn't got any particular definition but we know that it focuses on provocation and interaction. In olden days personality was predicted manually without the help of any machine but now because of latest mining techniques and advancement in technology we can actually predict personality in the percent values.

III. APPLICANT PERSONALITY PREDICTION SYSTEM

Our aim is to build a model of APPS by using Multinomial Logistic regression Machine Learning algorithm and Natural Language Processing modules like NLTK and Spacy. The ML model uses a CSV file as input for classifying applicant personality whereas NLP model uses resume and Job description for computing matching percentage of resume with Job description. After the successful completion of operation, the applicant's personality predicted and displayed.

The working principle of the APPS is shown in figure-1. The job seeker uploads resume and gives a personality test. The personality test is a test where every user is asked question on each of the characteristics [5] that was described in previous section. Once the users give their personality test then their result will be stored in database after evaluation and it will be used for predicting the personality of user[13]. The personality of the user is finally predicted based on the training and then each individual personality is analyzed using multinomial logistic regression before preprocessing of testing the data.

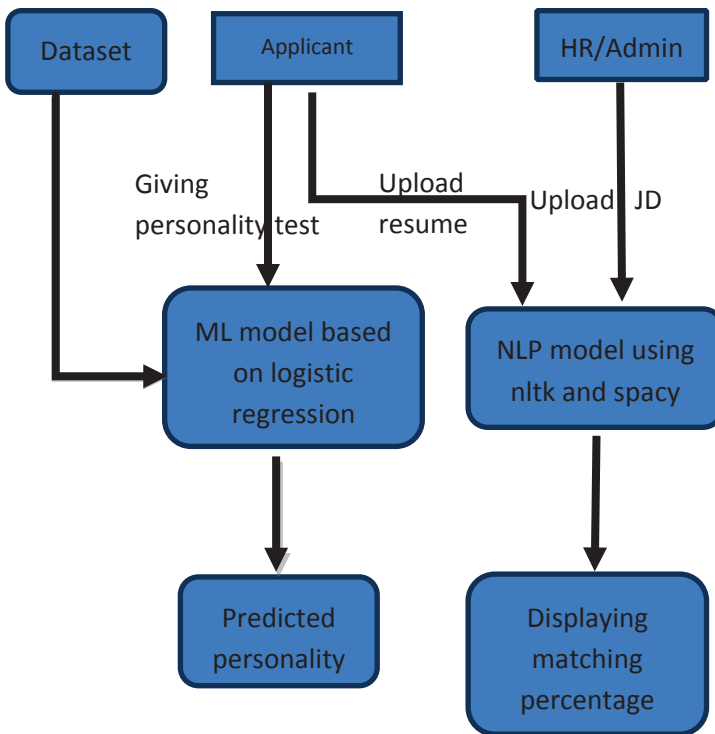


Fig. 1. Architecture diagram.

TABLE I. DATASET DESCRIPTION:

Number of entries	1026
Number of attributes	7
Number of class labels	5

As shown in table-1, for training the ML model using multinomial logistic regression we are using a dataset consisting of 1026 entries with 7 attributes and 5 class labels.

TABLE II. ATTRIBUTE DESCRIPTION:

S. No	ATTRIBUTE	TYPE	RANGE
1	Gender	nominal	Male/female
2	Age	numeric	17-60
3	Openness	numeric	1-10
4	Neuroticism	numeric	1-10
5	Conscientiousness	numeric	1-10
6	Agreeableness	numeric	1-10
7	Extraversion	numeric	1-10

As mentioned in table-2, we have seven attributes in which one is nominal and all other are numeric types.

TABLE III. CLASS LABELS DESCRIPTION:

S.NO	TYPE	CLASS LABEL
1	Nominal	Extraverted
2	Nominal	Serious
3	Nominal	Dependable
4	Nominal	Responsible
5	Nominal	Lively

In table-3, we have the type of outputs that are going to be displayed.

Steps for ML model training:

1. Data cleaning and preprocessing[11].
2. Splitting the dataset into two parts, one is testing set and other is training set.
3. Train the model using logistic regression according to [7][12] by using training dataset.
4. Testing the model accuracy.
5. Improve model performance.

In [15], for NLP model, the resume and Job Description will be given as input. Steps followed here are

1. Cleaning and preprocessing of documents using nltk module.
2. Tokenization of both documents.
3. Removing stopwords from our data.
4. Stemming and lemmatization.
5. Parts of speech tagging.
6. Named entity recognition.
7. Apply TF-IDF vectorizer [9] on both documents.
8. Compute similarity between documents using cosine similarity.

IV. RESULT ANALYSIS

The main aim of APPS is to know whether a person is suitable for a job profile or not and gives personality type of a candidate using logistic regression algorithm and matching percentage of resume with job description. The data set which we have used for ML model training is as follows (sample):

TABLE IV. SAMPLE DATA WITH RESULTS

S. No	1	2	3	4	5
Gender	M	M	F	M	F
Age	17	19	18	18	19
Openness	7	4	7	5	5
neuroticis m	4	5	6	7	5
Conscien tiousness	7	4	4	7	4
Agreeabl eness	3	6	5	6	5
extraversi on	2	6	5	4	4
Class label	extrav erted	seriou s	depen dable	livel y	Resp onsib le

In table-4 we are displaying some of the sample test cases that we used to find out how APPS is working and Personality in the above table is output what we get.

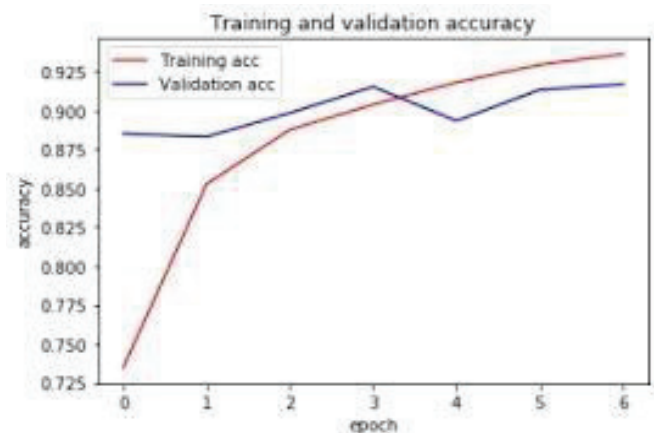


Fig. 2. model accuracy

The above graph figure-2 shows the training and validation accuracies of ML model where x-axis representing an epoch and y-axis representing model accuracy.

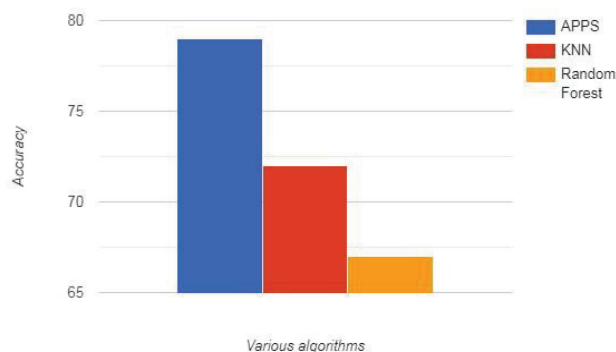


Fig. 3. Accuracy comparison

The figure-3 represents about accuracies of APPS, KNN and Random Forest Algorithm where the accuracy reached by APPS is 79.47 percent.

V. CONCLUSION AND FUTURE SCOPE

APPS is a system which matches the job description and resume of a candidates and displays the matching percentage of resume with job description so that depending on percentage it can be decided whether the candidate should be hired or not. APPS also judges the personality of candidates based on personality test that is conducted at the time of registration process. APPS is developed using logistic regression algorithm and natural language processing where multinomial logistic regression is used for judging applicant personality and natural

language processing used for comparison of job description and resume. At last we can tell that it can reduce the workload and time of hiring department and this APPS has high performance and accuracy where it can help hiring department to hire/select the best and skilled candidate according to their requirement.

REFERENCES:

- [1] Barrick, Murray R, The Personality Dimensions and job performance.
- [2] Gayatri Vaidya, Pratima Yadav, Reena Yadav, Prof.Chandana Nighut, Personality Prediction By Discrete Methodology, IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 22788719 Volume 14, PP 10-13 International Conference on Innovative and Advanced Technologies in Engineering.
- [3] Barrick, M. R., Mount, M. K. (1991). Big Five personality dimensions and job performance: A meta analysis Personnel Psychology, Vol. 44, pp. 1-26.
- [4] F. Harrell, Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis Springer, New York (2001)
- [5] DisethA, (2003). The Personality and approaches to learning as predictors of academic achievement. European Journal of Personality, 17, 143-155.
- [6] Michael M Tedasse; Lin Boxchu; Liang Yang, Personality Predictions Based on User Behavior on the Facebook Social Media Platform, 2018.
- [7] R. Berk. Statistical Learning from a Regression Perspective. Springer, 1st edition, 2008.
- [8] GSaltan. The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice-Hall, Inc.
- [9] J. Ramos, J. Eden and R. Edu, Using TF-IDF to Determine Word Relevance in Document Queries.
- [10] Nageswara Rao Sirisala and C.Shoba Bindu. A Novel QoS Trust Computation in MANETs Using Fuzzy Petri Nets, International Journal of Intelligent Engineering and Systems, Vol.10, No.2, (2017), pp 116-125.
- [11] Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kaufman, 2nd ed.
- [12] Peng, C. Y., & So, T. S. H. (2002). Logistic regression analysis and reporting: A primer. Understanding Statistics, 1(1), 31-70.
- [13] Nageswara Rao Sirisala and C.Shoba Bindu. Uncertain Rule Based Fuzzy Logic QoS Trust Model in MANETs, International Conference on Advanced Computing and Communications -ADCOM, (IITM PhD forum), (2015), pp.55-60.
- [14] Ertan Taş; Mustafa Ersel Kamaşak, Prediction Of Personality Traits From Videos By Using Machine Learning Algorithms, 2019 4th International Conference on Computer Science and Engineering (UBMK), 2019.
- [15] Rajman, M., Besançon, R.: Text mining: natural language techniques and text mining applications. In: Data Mining and Reverse Engineering, pp. 50-64. Springer, US (1998).
- [16] W.H.Gomaa and A.A.Fahmy, A survey of text similarity approaches, International journal of computer applications, vol.63. 2018.
- [17] M. S. Raja, M. Anurag, C. P. Reddy and Nageswara Rao. Sirisala, "Machine Learning Based Heart Disease Prediction System," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5.