International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Personality Prediction via CV Analysis

Anjali Kumari
School of Computing Science and Engineering
Galgotias University, India
anjaligupta82526@gmail.com

Utkarsh Raj
School of Computing Science and Engineering
Galgotias University, India
utkarshraj0513@gmail.com

Dr. S. Rakesh Kumar
School of Computing Science and Engineering
Galgotias University, India

ABSTRACT*–The corporate world nowadays doesn't focusjust on the talents a possible worker possesses however additionally their temperament. temperament is what helps one achieve success in skilled still as personal life. Hence, the recruiter should bear in mind of the personalitytraits an individual has. With Associate in Nursing exponential increase in job seekers however a decrease within the variety of jobs, it's troublesome to manually order the simplest match candidate for an acceptable job by gazing the CV. This paper tries to look at totally different machine learning approaches forefficiently predicting temperament through CV analysis victimization language process (NLP) techniques still. Results show that the Random Forest formula achieved higher accuracy compared to alternative algorithms like kNN, supply Regression, SVM and Naive Bayes.*

Keyword: *Personality Prediction, CV, Machine Learning*

I. INTRODUCTION

The word 'Personality' derives from the Latin word persona that refers to a mask worn by actors to act. However, temperament is way quite a mask currently, it might presumably verify whether or not an individual is appropriate for a specific job profile. It tells US if a person's is capable enough to guide, influence and communicate effectively with others. the primary step of enlisting is that the application that consists of non-public details, experience, and most significantly CV. firms generally receive thousands of applications per job gap and have an ardent team ofscreeners to pick out qualified candidates. it's terribly tough for kinsmen to manually undergo the CV of all candidates. several candidates get filtered go into the primary spherical itself on the premise of suitableness, improper CV, not being adept enough. Hiring the rightcandidate could be a terribly tough task as no candidate is ideal, some may not be adept enough or some may not have the correct temperament. Hence, we have a tendency to propose how within which the method of shortlisting gets efficient and quicker by temperament prediction.

CV's will replicate upon the skilled qualifications of a person however don't replicate upon the temperament of someone. temperament is one in every of the important factors that suggests however someone would be ready to add a delegated role, thus temperament analysis and understanding is essential. Our objective doing this project is to create the machine a lot of human, associated analyse the candidate in such how that an actual human reviewer would.

This paper tries to explore and implement varied machine learning algorithms and analyse that one in every of them provides the most effective accuracy with a large array of knowledge provided. we tend to additionally commit to visualise the info and type a affiliation between varied factors.

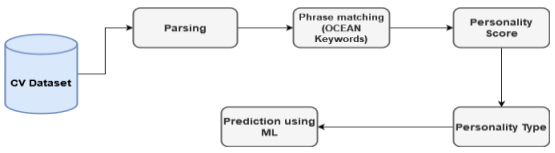
II. PROPOSED SYSTEM

Dataset:

As manual knowledge assortment is long, we have a tendency to collected candidate resumes through tons of internet sites and private interaction with potential job seekers taking the overall count to 708 CVs. The collected CVs were in PDF and DOCx format.

Methodology:

The objective of our paper is to predict the temperament of someone supported their score of openness, extroversion agreeableness, psychological disorder and conscientiousness. For achieving this, we wanted the way to calculate the scores directly from each CV. Our approach as shown in Fig.1 was to analyse the whole resume and rummage around for keywords regarding the 'Big 5 Test'.



III. BACKGROUND

There ar numerous tests that facilitate to work out temperament varieties like the large 5, inkblot test, and MBTI check. during this paper, prediction of personalityis done by considering the large 5 check because it is claimed to be wide used, reliable and conjointly it's direct relation between

performance primarily that specialize in 5 psychological feature domains [1].

BIG FIVE TEST:

Big 5 mental test additionally referred to as OCEAN Model analyses temperament forms of people supported 5 dimensions - Openness(O), Conscientiousness (C), sociability (E), Agreeableness (A), Neuroticism(N). With every of the scale signifying a special temperament sort. It uses keywords to spot traits and analyze during which temperament an individual match.

Openness: As the word suggests, This quality options characteristics like openness and imagination and curiosity.

Conscientiousness: Conscientiousness talks a couple of high quantity of thoughtfulness, a goal- oriented perspective and smart decision-makers.

Extraversion: Extraversion conjointly suggests that sociableness is known by excitement, loquacity and positiveness.

Agreeableness: Agreeableness refers to options like trust, fondness and social behaviour of a personal.

Neuroticism: Disturbance includes attributes like disappointment, moodiness and sharp burst of emotions.

As these 5 dimensions cowl the majority avenuesneeded to grasp somebody it's the correct methodology that forms the premise of a person's overall temperament.

IV. RELATED WORK

Kalghatgi et al. [2] bestowed a Neural Network Approach supported the massive 5 check to predict the temperament of people counting on tweets revealed on Twitter by extracting meta-attributes from tweets. that area unit wont to analyze one's social behaviour. The authors followed a four-step method that is knowledge assortment from tweets, Preprocessing, Transformation and Classification. though neural networks area unit wont to predict temperament there area unit limitations like countering pretend data, automatic analysis of tweets and looking forward to simply Twitter isn't enough to predict someone's temperament however solely user behaviour and trends.

Allan Robey et al [3] planned a system to scale back the load on the Human Resource department of firms by having 2 sides: organization and candidate destined. The authors claim that the planned system are going to be more practical to order CVs from an outsized pool ensuring that the ranking is honest and legal. the most distinction between the present system and therefore the planned system is that rather than simply scanning the CVs, the authors propose to conduct an inherent ability check and a mental test for temperament prediction.

Juneja Afzal Ayub Zubeda et al [4] worked on a projectto rank CVs exploitation tongue process and Machine Learning. The system ranks CVs in any format in keeping with the company's criteria. The authors propose to think about candidate's Github and LinkenIn profile additionally to urge a far better understanding creating it easier for the corporate to search out an acceptable match supported skillsets, ability and most significantly, temperament.

Md. terrorist group Reza and Md. Sakib albizzia analyzed CV of people mistreatment linguistic communication process and Machine Learning by 1st changing CVs to markup language so reverse engineering to markup language code following that, phase closing and qualification feature extraction has been done. The model extracts knowledge from a CV and segments them supported the values. they need classified the CVs mistreatment variable supplying regression. However, the scale of the dataset was terribly less.[5]

V. MODEL TRAINING AND TESTING

Before coaching our model, we tend to label encoded the temperament column of our dataset. Our final dataset had 708 rows and vi rows. mistreatment the sklearn library, we've got used seventieth of our knowledge for coaching functions and half-hour for testing the results. For predicting the temperament of a prospective candidate, we've got used numerous machine learning algorithms like logistical Regression, Naive Thomas Bayes, Random Forest, Support Vector Machine (SVM) and KNN.

Logistic Regression:It is AN algorithmic program analogous to simple regression, except it predicts whether or not one thing is True or False. it's a well-liked algorithmic program for determination classification issues like Binary Classification (Pass/Fail, Rain/No Rain).

Naive Bayes: In likelihood, Baye's theorem is employed to reason the contingent probability. the concept forms the premise of the Naive mathematician classifier, a classification algorithmic rule that assumes robust independence assumptions between the options. consistent with the algorithmic rule, every feature within the downside makes AN equal and freelance contribution to the end result.

KNN: kNN stands for k-nearest neighbours, a supervised machine learning rule capable of determination each regression and classification issues. Intuitively we are able to consider the saw 'Birds of a similar feather flock along as like kNN. The rule assumes that similar information points sometimes occur in closeproximity.

SVM :Support vector machine could be a supervised machine learning algorithmic rule accustomed alter knowledge for classification and multivariate analysis. The goal of SVM is to search out a hyperplane in N-dimensional house (N-variety of features) that may simply classify the info points.

Random Forest :Random forest is another ensemble technique used for classification and regression tasks. It uses multiple call trees to provide the output. sacking or bootstrap aggregation area unit wont to train the random forest algorithm's "forest.

After coaching our model on all of the algorithms, we tend to accomplished that our predictions clad to be rather poor. Even our greatest models may solely notice associate accuracy of concerning thirty p.c

Another issue was that our coaching and testing datasets had terribly completely different distributions. whereas our coaching information was a touch unbalanced, the testing information was even additional unbalanced. however after we place ourselves within the shoes of AN leader, we tend to realise that he would need to rent somebody WHO is 'responsible' and 'lively' quite the rest. so our problemnow turns into a binary
classification drawback (1- accountable or spirited 0-others) [8]

VI. EXPERIMENTAL RESULTS

Now when feeding the info to the models, we have a tendency to managedto spike the accuracy to concerning zero.71. Random Forest formula provides the most effective accuracy followed by the likes of mathematician, kNN, SVM and provision Regression as seen in Table three. evidently, Random Forest additionally has the smallest amount mean square error, that measures the typical of the sq. of the distinction between actual and calculable values.

Table 3:- Accuracy And MSE Values

Model	Accuracy	MSE
Logistic Regression	0.62	0.37
Naive Bayes	0.65	0.37
kNN	0.64	0.35
SVM	0.63	0.36
Random Forest	0.71	0.29

VII. CONCLUSION AND FUTURE SCOPE

In this paper, we've used numerous Machine Learning Algorithms like supplying Regression, Naive mathematician, Random Forest, SVM and KNN for temperament prediction victimization CV Analysis. victimization pyresparser,spaCy and PhraseMatcher we tend to were ready to predict the personalities of varied candidates. The results indicate Random Forest has the utmost accuracy of zero.71 but because of lack of accessible knowledge the accuracy is far lesser than it absolutely was anticipated. The projected system will be used by numerous firms so as to contour the enlisting method by considering the temperament of potential candidates. Future work also can be done to extend the potency and performance of the projected system so as to predict temperament victimization CV analysis additional accurately.

VI. REFERENCES

1.A Demetriou, L. Kyriakides, and C. Avraamidou, "The missing link in the relations between intelligenceand personality" in Journal of Research in Personality,vol. 37 issue 6, December 2003, pp 547-581.

2.M. Kalghatgi, M Ramannavar, and Dr. N. S. Sidnal, "Neural Network approach to personality prediction based on the Big-Five Model" in IJIRAE, vol2 issue 8, August 2015, pp 56-63.

3.A..Robey, K. Shukla, K. Agarwal, K. Joshi, Professor S. Joshi "Personality prediction system through CV Analysis, in IRJET vol 6, issue 02, February2019

4. J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem "Resume Ranking using NLP and Machine Learning", unpublished.

5. Md Tanzim Reza, and Md. Sakib Zaman, "Analyzing CV/Resume using natural language processing and machine learning", unpublished

6.https://towardsdatascience.com/

7. https://spacy.io/

8. https://www.kaggle.com/