Online CV Analysis And Ranking System

Shrikant Padhy
Department of Information Technology
Ramrao Adik Institute of Technology
Mumbai University
shrikantpadhy18@gmail.com

Ritik Patil
Department of Information Technology
Ramrao Adik Institute of Technology
Mumbai University
patil.ritik.17it2044@gmail.com

Akshay Pawar
Department of Information Technology
Ramrao Adik Institute of Technology
Mumbai University
akshay3708@gmail.com

Anita Senathi
Department of Information Technology
Ramrao Adik Institute of Technology
Mumbai University
anita.senathi@rait.ac.in

Abstract— **This system can used in many business sectors that may require expert candidate. This system will reduce workload of the human resources. This system will help the human resource to select right candidate for desired job profile, which in turn provide expert workforce for the organization. Admin can easily shortlist a candidate based on their online test marks and select the appropriate candidate for particular job profile. This will enable a more effective way to short list submitted candidate CVs from a large number of applicants providing a consistent and fair CV ranking policy, which can be legally justified. Using Natural Language Processing (NLP) can be defined as a process that enables a machine to become more like a human, hence slashing the distance between machines and humans. In this field challenges like understanding the NLP i.e. permitting the machines to understand the natural language of humans are faced. This system will help the HR dept. to easily shortlist the candidate based on their CV ranking policy. This system will focus not only in qualification and in experience but also focuses on other important aspects, which are required for particular job position**

Keywords— scan cv, NLP system;

I. INTRODUCTION

This system will enable a more effective way to short list submitted candidate CV's from a large number of applicants providing a consistent and fair CV ranking policy, which can be legally justified. This system will rank the experience and key skills which is required for particular job position.

The system will rank the CV's based on the experience and other key skills which are required for particular job profile. System will help the HR department to easily shortlist the candidate based on their CV ranking policy. This system will not focus only in qualification and experience but also focuses on other important aspects which are required for particular job position. This system will help the human resource dept. to select right candidate for particular job position which in turn provide expert workforce for the organization. Candidate here will register him/herself with all its details and will upload their own CV into the system which will be further used by the system to shortlist their CV [1]. Candidate can also give the online test which will be conducted on personality questions as well as aptitude questions. After completing the online test, candidate can view their test results in graphical representation with their marks [2].

II. LITERATURE SURVEY

Gayatri Vaidya[3] This system will be constructed by first constructing a dataset consisting of photographs, quality factors, personality measurement and intelligence measurement for building an end to end network for personality prediction where selfreported personality traits can be predicted reliably from an image by Discrete Methodology. The main goals of the proposed system is to distinguish internal properties of the person

from overt behaviors, and to show the results using chart or in percentage.

Mayuri Pundlik Kalghatgi[4] This work analyses social media data to predict significant personality traits, i.e. qualities or characteristics are specific to an individuals, using the Big Five Model. A parallelism between an individual's personality traits his/her linguistic information is explored for analytics. The Big Five model enables the identification of personality traits through linguistic information. This reveals the personality traits which can be applied to various domains like business intelligence, marketing and psychology. Thus this work provides a personality prediction system based on Big-Five model, which can predict the traits of an individualusing the group of tweets posted by him.

In 2014 an Integrated E-Recruitment System for Automated Personality Mining and Applicant Ranking was proposed by Faliagka et al. an automated candidate ranking was implemented by this system. It was based on objective criteria that the candidate's details would be extracted from the candidate's LinkedInprofile.
The candidates' personality traits were automatically extracted from their social presence using linguistic analysis. The candidate's rank was derived from individual selection criteria using Analytical Hierarchy Process (AHP), while their weight was controlled by the recruiter (admin).
The limitations of the system were that senior positions that required expertise and certain qualifications were screened inconsistently[5].

Liden et al. published The General Factor of Personality: The interrelations among the Big Five personality factors (Openness, Conscientiousness, Extraversion,
Agreeableness, and Neuroticism) were analyzed in this paper to test for the existence of a GFP. The meta-analysis provides evidence for a GFP at the highest hierarchal level and that the GFP had a substantive component as it is related to supervisor-rated job performance were concluded by this paper . However, it is also realized that it is important to note that the existence of a GFP did not mean that other personality factors that were lower in the hierarchy lost their relevance
[6].

III. METHODOLOGY

In this step are as follows [9]:

(NLP) Natural Language Processing is the area of application and research that learn how to computers can be used to understand and manipulate natural language text or speech to do useful things. NLP is a branch of computer science and artificial intelligence (AI) which is concerned with interaction between computers and human languages. Natural Language Components-

1) NLU (Natural Language Understanding) : Natural language understanding (NLU) is a unique category of natural language processing that involves modeling human reading comprehension or in other words, parses and translates input according to natural language principles.

2) NLG (Natural Language Generation) : Natural language generation (NLG) is a particular Artificial Intelligence complete task that involves generating language from non-language inputs. Some experts might be refer to a natural language generation application as a "translator" of text or other informational formats into the spoken language.

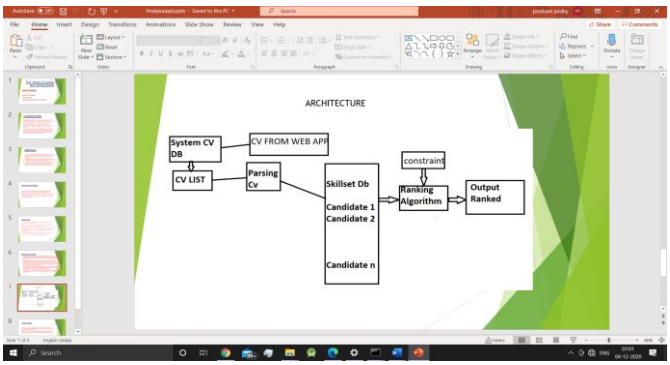

Fig. 1 : System Architecture .

The System Architecture consists of two modules:
1. Outer World System
2. Resume Ranking System

1. Outer World System Consist Of:
    1. Client Company.
    2. System C.V's Data base.
    3. Social Profile.
2. Resume Ranking System Consist Of:
    1. Parser System.
    2. Candidate Skillset Database.
    3. Resume Ranking algorithm.

Client Company : This is the client company who will provide us the bulk of the resume or C.V's with the specific requirement and constraints, according to which it should be ranked.
System C.V's Database : This is the large database which is used to store the bulk of resumes provided by the client company in a distributed environment.
Social Profiles : Social Profiles include LinkedIn Profile of the candidate, Github Profile of the Candidate. This social profile module can be extended to different community too.
Parser System : Parsing system includes the parsing of the following candidate resume and their social profiles using NLP. That is without any manual interaction.Here, using Natural Language Processing the this is how we are going to parse the resume one at a time. NLP (Natural Language Processing) requires following constraint for parsing : Morphological Analysis Syntactic Analysis Semantic Analysis

IV. Design Algorithm

In this section we state the machine learning algorithm (*TF-IDF*) for CV analysis.

The TF-IDF [10] Algorithm is used to find out the important keywords in a document/CV. Below, we give the working of TF-IDF in detail.

Step 1: Calculate TF (Term Frequency)
Term Frequency (TF) - Number of times a keyword appeared in a document is calculated by Term Frequency.

TF ('keyword') =number of times 'keyword' appears in document /Total number of keywords in the document.

Here, the term 'keyword' signifies any job specific skill which the algorithm is searching for.

Step 2: Calculate IDF (Inverse Document Frequency) value.
The problem of rare and frequent words is resolved.
This helps our system to give more priority to the required word or skills.
IDF sets the *log* value=1 for the required CV as per skill sets and *log* value=0 for the unwanted CV.

IDF ('keyword') =log (total number of CV/Number of document with term 'keyword')

Step 3: Calculate TF-IDF weight

*Weight= TF ('keyword')*IDF ('keyword')*

Higher the weight, more relevant is the CV and lower the weight, less or no relevance of the CV for the selection process [11]. This step returns the CV with highest and lowest weight values which is further useful for classification [12].

The system determines the candidate on the score obtained. The high-frequency of some keywords may impact on candidate overall score. TF-IDF is widely used in text mining techniques [13]. The algorithm takes into account the effect of high-frequency keywords and negates the low-frequency keywords.

V FUTURE SCOPE

Further, we can modify the existing system to perform sentiment analysis of social media data [7]. Many more classification algorithms of machine learning can be integrated to provide much better functionalities [8]. Further, the efficiency and performance of the application can be tested and analyzed. The application can be extended further to other domains like Telecom, Healthcare, Ecommerce and public sector jobs.

VI. CONCLUSION

This system will help the human resources department to select right candidate for particular job position, which in turn provide expert workforce for the organization. This system will help to get shortlisted CV's according to their ranking. Ranking is based on their test result and experience, qualification etc. This system will reduce work of the human resource department.

VII. REFERENCES

[1] http://ieeexplore.ieee.org/document/5566454/

[2] The "Personality dimensions and job performance" includes what is the important of the personality dimensions by Barrick Murray. R.

[3] 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): 2278-8719 Volume 14, PP 10-13 International Conference on Innovative and Advanced Technologies in Engineering

[4] Mayuri Pundlik Kalghatgi, Manjula Ramannavar , Dr. Nandini S. Sidnal, "A Neural Network Approach to Personality Prediction based on the Big-Five Model", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 8, Volume 2 (August 2015).

[5] F Evanthia, T Athanasios, et al. An integrated e-recruitment system for automated personality mining and applicant ranking. Internet Research 2012; 22: 551568

[6] L D. van der, J te Nijenhuis, et al. The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. Journal of Research in Personality 2010; 44: 315-327

[7] Amit Palve, Rohini D.Sonawane, Amol D. Potgantwar, "*Sentiment Analysis of Twitter Streaming Data for Recommendation using, Apache Spark*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.3, pp.99-103, 2017.

[8] Huan Liu, Lei Yu, "Towards integrating feature selection algorithms for classification and clustering", *IEEE Transaction on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491-502, April 2005.

[9] Mayuri Pundlik Kalghatgi, Manjula Ramannavar , Dr. Nandini S. Sidnal, "A Neural Network Approach to Personality Prediction based on the Big-Five Model", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 8, Volume 2 (August 2015).

[10] Shereen Albitar, Sebastien Fournier, Bernald Espinasse, An effective tf/idf-based text-to-text semantic similarity measure for text, spinger, pp. 105-114, 2014.

[11]Stephen Robertson,"Understanding inverse document frequency: on theoretical arguments for IDF", Journal of Documentation, Vol. 60 Issue: 5, pp.503-520, 2004.

[12] 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 of Scientific Research in Computer Science and Engineering, Vol.06, Issue.01, pp.19-22, 2018.

[13]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 181 – No.1, July 2018.