

## A Review Paper on Resume scanning using python

Mr. Panil jain<sup>1</sup>, Mr. Nandkishor kamble<sup>2</sup>, Ms. Risha Nadar<sup>3</sup>, Mr. Shaikh Luqman<sup>4</sup>,

<sup>1234</sup>student, Department of Electronics and Tele-comm Engineering, Xavier Institute of Engineering, India

\*\*\*

### Abstract -

Traditional hiring practises are becoming ineffective as internet recruitment continues to rise in popularity. This is because employment portals receive a large number of unstructured resumes – in a variety of styles and forms – from applicants with varying levels of knowledge and specialisation. As a result, structured data extraction from applicant resumes is required not just to assist automated candidate screening, but also to efficiently route individuals to their appropriate occupational groups. This helps businesses manage and organise resumes and sift out irrelevant candidates with less effort.

Finding acceptable candidates for an open position can be difficult, especially when there are a lot of them. It can stymie team progress in terms of getting the right person in the right place at the right time. An automated “Resume Classification and Matching” system could greatly simplify the timeconsuming procedure of fair screening and shortlisting, as well as speed up candidate selection and decision-making. This system could handle a large number of resumes by first classifying them into the appropriate categories using various classifiers. Once that is done, top candidates could be ranked based on the job description using Content-based Recommendation, cosine similarity, and k-NN to find the CVs that are closest to the job description.

**Key Words:** Docopt, Hiring Pattern, Human Resources, PyPDF2, Python, Resume

**1.INTRODUCTION** Within Human Resources, talent acquisition is a critical, difficult, and time-consuming activity. The sheer size of the Indian market is mind-boggling. Not only are one million people entering the labour market every month, but there is also a lot of churn. According to LinkedIn, India has the largest percentage of workers who are “actively looking for a new job.” Clearly, this is a very liquid and large market, but it also contains a lot of irritating inefficiencies. The lack of a consistent organisation and format is the most difficult aspect. For a resume, which makes shortlisting potential profiles for required tasks extremely time-consuming and laborious. To determine the relevance and applicability of a profile for a specific job, effective resume screening necessitates domain expertise. Short-listing presents a difficulty for the human resource department because there are so many various job roles available nowadays, as well as the customary big number of applications submitted. This is

exacerbated by the HR department’s lack of broad skill sets and subject knowledge, both of which are necessary for efficient HR management. India’s market is massive. Not only do one million people enter the labour market every month, but there is also a lot of turnover. The sector is currently confronted with three significant challenges:

- Separating the right individuals from the pack – With millions of people looking for work in India, screening CVs and finding the appropriate fit is nearly difficult. This makes the entire hiring process slow and inefficient, wasting time and money for businesses.
- Making sense of candidate CVs – The fact that CVs on the market are not standard poses a second issue. Practically every resume on the market has a different structure and format. HR must manually review the CVs in order to determine the best fit for the job description. This is time consuming and prone to inaccuracy, as a suitable candidate for the job may be overlooked in the process.
- Confirming that candidates are capable of performing the job before hiring them – The third and most difficult problem is matching the CV to the job description in order to determine whether the candidate is qualified for the position for which she is being hired.

In this research, we provide an automated Machine Learning-based methodology to address the aforementioned concerns in the resume short-listing process. The model takes the features taken from the candidate’s CV as input and categorises them, then maps the categorised resume to the required job description and recommends the best candidate’s profile to HR.

The following are our major contributions:

1. We created a resume recommendation system that is automatic.
2. Classification techniques based on machine learning and similarity functions are applied to find most relevant resume.

### 2. Literature Review

Over 50000 internet recruitment sites exist, all of which need job applicants to submit their resumes through their websites. Classification techniques for filtering resumes are not even used on some of these websites. The firm recruiter’s task is to manually go over all of the candidate resumes. Selecting the most capable candidates for the succeeding phases of the hiring process is a challenging assignment for recruiters. Meanwhile, some job boards have applied the clever idea of automatically scoring or classifying resumes submitted by candidates for a certain

job opening. Indeed, Monster.com, Adecco.com, Top Resume, Ideal, and others are some of these websites or web applications. The advantages and disadvantages of some of these websites are described. 2.1 Resume Ranker, Nileema Pathak IEEE March 2018 Human Resource (HR) agencies use various head hunting tools and online search methods. These search methods connected with the database of millions of resumes. These are the simple search engines who parses the resumes against the given keywords and offers the best match results. The list of the searching keywords is usually prepared by the HR after reading the job description several times. The HR downloads these searched resumes and does the manual work by opening and reading the resumes. By this way, HR person tries to find the resumes which are best match to the Job Description. This is a cumbersome process and requires reasonable time and multiple discussion with the candidate before offering the resume to the client. Usually, due to the complexity of the database, many efficient resumes missed out from the search results or not considered due to stringent time lines of closure.

2.2 : A Job Post and Resume Classification System for Online Recruitment ,Abeer Zaroor, and IEEE 2021 To Overcome the abovementioned limitations, we present a hybrid approach to classify resumes and their corresponding job post by utilizing an integrated occupational categories knowledge base. The exploited knowledge base assists in i) classifying resumes and job offers under their corresponding occupational categories and ii) automatically ranking applicants that best match the announced offers.

2.3 AN INTEGRATED E-RECRUITMENT SYSTEM FOR CV Eleanna Kafeza, IEEE2015

On-line recruitment systems, candidates typically upload their CVs in the form of a document with a loose structure, which must be considered by an expert recruiter. This incorporates a great asymmetry of resources required from candidates and recruiters, resulting in candidates uploading the same CV in numerous HR agencies that become overwhelmed with thousands of CVs. In this work, we follow a different approach in the CV submission process, which is detailed in this section, along with the CV modelling in HRXML format.

2.4 Contour shape and color detection using open cv python Raghav puri, IEEE-2018

OpenCV is Open Computer Vision Library [4]. It was initially launched in 1999 by Intel. With more updates, it has been modified since then to aim for the real-time computer vision. This library has been written under programming languages like C and C+. It can be easily run on operating systems Windows and Linux. This library can be easily interface with programming languages like Python, MATLAB, Ruby and others as well. Along with Numpy [6] and Python image processing (shape & color detection) can be performed at ease.

2.5 Resume Parsing with Named Entity Clustering Algorithm IEEE-2014

This article introduces the notion of the Internet of Things (IoT) ecosystem. There is a need to connect networks of

sensors and smart objects in a way that allows them to interact intelligently. Within the IoT ecosystem, architecture for an Autonomous Vehicles prototype is proposed. Because the GPS-enabled device is indoors at the time a request is made, the GPS signal may be weak or nonexistent, and the users may suffer as a result. As a more precise outdoors technique and as an alternative to low GPS signaling indoors, passengers and cars' positions can be estimated using a mix of GPS data and the Wi-Fi receiver's location estimation.

The retrieved tf-idf features set are put into the RF classifier to predict the resume category, and we've started our trial with it. On 10-fold cross-validation, the RF classifier has an accuracy of 38.99 percent. We utilized another popular classifier called "NB" for this assignment because the results were not adequate. Classification NB Predicted resume categories with a 44.39 percent accuracy, which was higher than the previous classifier's accuracy (RF). However, the NB classifier's accuracy of 44.39 percent suggested that more than half of the resumes were misclassified. On the same data, we tried another classifier called "Linear SVM" and got 78.53 percent accuracy. The LR classifier was employed to increase the model's accuracy, and it yielded a result of 62.40 percent, which was lower than expected. of the transmitter and its interaction with its receiver. Ultrasonic sensor for presence detection. The MaxBotix ultrasonic sensor addresses the issue of human presence detection. Our ultrasonic sensors have high reading speeds and excellent read-to-read stability, so they can detect people over a wide range of distances. Ultrasonic transducers are commonly used in applications that include level or distance measurements. Ultrasonic sensors are often used in level measurements because they are quick and easy to implement.

### 3. Proposed System

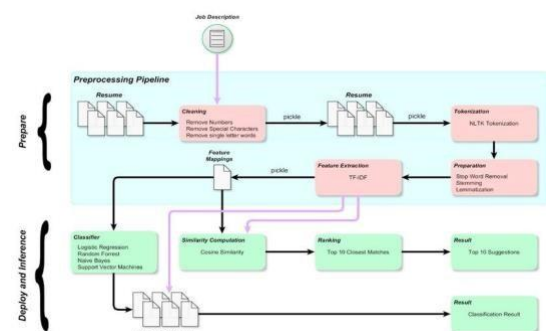


Fig -1: Block diagram of the proposed system

The retrieved tf-idf features set are put into the RF classifier to predict the resume category, and we've started our trial with it. On 10-fold cross-validation, the RF classifier has an accuracy of 38.99 percent. We utilized another popular classifier called "NB" for this assignment because the results were not adequate. The accuracy of the

NB classifier in predicting resume categories was 44.39 percent, which was higher than the accuracy of the previous classifier (RF). The NB classifier's accuracy of 44.39 percent, on the other hand, suggested that over half of the resumes were misclassified. On the same data, we applied a different classifier called "Linear SVM" and got an accuracy of 78.53 percent. The LR classifier was employed to increase model accuracy, and it achieved a score of 62.40 percent, which was lower than the "Linear SVM" classifier's score of 63.40 percent. All of the models' accuracy was determined using a 10-fold cross-validation method, with the average model being found to be the most accurate. The classifiers' accuracy was measured.

#### 4. CONCLUSIONS

For each job opening, the company receives a large number of applicants. Finding the most appropriate candidate's application from a sea of resumes is a time-consuming chore for every company these days. The candidate's resume classification is a laborious, time-consuming, and resource-intensive process. To address this problem, we created an automated machine learning-based methodology that recommends acceptable applicant resumes to HR based on job descriptions. The proposed model consisted of two steps: first, categorise the resumes into different groups. Second, depending on the similarity index with the specified job description, it proposes resumes. If a particular industry generates a high number of resumes, the presented approach can be used to create an Industry-specific model. Involving domain specialists such as HR professionals will aid in the development of a more accurate model, and HR professionals' feedback will aid in the iterative improvement of the model.

#### REFERENCES

- [1] Swapnil Sonar, Resume Parsing with Named Entity Clustering Algorithm, IEEE Research, may2012, <http://www.slideshare.net/swapnilmsonar/resumeparsing-with-named-entity-clusteringalgorithm>
- [2] Machine Learning Methods for Solving Complex Ranking and Sorting Issues in HumanResourcing", Institute of Electrical and Electronics Engineers (IEEE) 2017.
- [3] IntelligentHiringwithResumeParserandRankingusingNaturalLanguageProcessingandMachine Learning", International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE)
- [4] Wei, K., Huang, J., Fu, S., 2007. A survey of ecommerce recommender systems, in: 2007 international conference on service systems andservice management, IEEE.
- [5] Paparrizos, I., Cambazoglu, B.B., Gionis, A., 2011. Machine learned job recommendation, in: Proceedings of the fifth ACM Conference on
- [6] Ramos, J., et al., 2003. Using tf-idf to determine word relevance in document queries, in: Proceedings of the first instructional conference on
- [7] Resnick, P., Varian, H.R., 1997. Recommender systems. Communications of the ACM 40, 56-59.
- [8] Rish, I., et al., 2001. An empirical study of the naïve bayes classifier, in: IJCAI 2001 workshop on empirical methods in artificial intelligence
- [9] Roy, P.K., Singh, J.P., 2018. A tag2vec approach for questions tag suggestion on community question answering sites, in: International
- [10] Roy, P.K., Singh, J.P., Baabdullah, A.M., Kizgin, H., Rana, N.P., 2018a. Identifying reputation collectors in community question answering