iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43978

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: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

A Machine Learning Approach for Automation of Resume Recommendation System

Lakshmi Naga Vishnu Babu Korrapati[1], Venkata Jaswanth Chowdary Malineni[2], Praveen Kumar Rama[3], Sai Raghava Konuru[4], Udith Siva Sai Raju Konduru[5], Aemi Kalaria[6]
Presidency University

Abstract: Finding qualified candidates for a vacant post can be difficult, notably if there are many applications. It can stifle team growth by making it difficult to hire the right person at the right time. "A Resume Recommendation System" has the potential to significantly simplify the time-consuming process of fair screening and shortlisting. It would certainly enhance candidate selection and decision-making. This system can handle a large number of resumes by first classifying them using multiple classifiers and then recommending them based on the job description. We offer research to improve data accuracy and completeness for resource matching by combining unstructured data sources and incorporating text mining algorithms. Our method identifies categories by extracting and learning new patterns from employee resumes.

I. INTRODUCTION

The problem of automatically recommending resumes to appropriate job descriptions has sparked a lot of study in recent years. However, it is widely considered that better user-friendly filtering approaches are required for automated recommendation systems to be more frequently adopted. We discuss here our study on three Machine-Learning strategies for disseminating or that may broadcast job offers to the appropriate person at the right time, which are based on Random Forest, Naïve Bayes, and Support Vector Machines, respectively. Today's job market is extremely volatile. Indeed, this is one of the primary reasons for the growing desire for improved means of publishing or finding intriguing employment opportunities. Furthermore, this interest is bilateral, which means that it emanates not just from Human Resources (HR) departments in companies, middlemen, or makers of recruitment tools, but also from job searchers searching for new professional challenges.

This implies that, as a first step, it is believed that a tentative reduction of the most potential applicants and job postings would result in significant enhancements and benefits for both sides. In this context, job portals and online recruiting platforms have historically been created to assist recruiters and job seekers in readily locating relevant people and job offers. Many job boards and web-based recruitment tools are now available all over the world. However, there is a considerable body of evidence suggesting that the functionality of existing portals should be improved. In general, only online job advertising is monitored, and these are then categorized using a basic textual description or key characteristics.

The scientific community has been working on a sort of information filtering mechanism to help job seekers and recruiters (HR) efficiently locate what they are looking for. More specifically, resume recommender systems will try to automatically suggest job applicants' resumes in this manner that HR could manually check through those resumes, rather than HR looking through every resume that is received.

II. BACKGROUND

For many years, human resource information systems were primarily restricted to tracking applicant data via applicant monitoring systems. However, as the workforce and commercial sectors become more diversified, the process of locating a suitable individual for a vacant position and vice versa becomes more cumbersome. It is obvious that new social media outlets, as well as an astounding number of job portals, necessitate new techniques and technology for both recruiters and job hunters.

Job sites are now employing the same recommendation algorithms and related technology to provide better outcomes. The concept is straightforward. Thousands of companies post job vacancies and job descriptions. Thousands of applicants and job seekers publish their credentials on these portals at the same time. The better a website matches these resumes to the job description, the higher the odds of conversion and the more popular the website will grow through referrals and publicity. Traditional approaches can only produce findings with smaller amounts of data.

©IJRASET: All Rights are Reserved

Previously, job portals relied on decision trees and tried their hardest on a case-by-case basis. performed their best under certain conditions; The situation has altered today. Concrete rules are no longer used to group people into occupations. Instead, dynamic algorithms have been implemented. These algorithms examine resumes, collect data, format it, and then attempt to match it to any previously stored resume in the database. When it discovers the best fit, it retrieves and shows the job suggestion engine that is displayed with the highest similarity.

III. USE CASES

A lot of firms who want to automate their e-recruitment processes are now interested in services for the automatic recommendation of resumes or job applications. Human resources, market makers, internet job platforms and gateways, and application developers are among the most crucial. We take a deeper look at each of them here.

A. HR Department

Human Resources (HR) departments in businesses must deal with challenges like these on a regular basis. Currently, huge corporations' HR departments get a tremendous number of e-mail applications. All application materials must be manually processed in order for the essential information collected to be put into the internal recruitment systems. This procedure takes a long time and consumes a lot of resources (time, money, effort). As a result, the HR department needs a system that automatically summarizes and suggests applicants who best match the job description. so that they can easily screen the system's resumes.

B. Market Entrepreneur

HR Recruiters and recruitment agencies are frequently given the task of identifying the best applicant for a certain job description. The work is so complicated that many corporations are prepared to spend large sums to complete it effectively. Solutions for resume suggestions can assist to ease this issue, making it much more functional and effective.

C. Electronic Job Platforms and Agencies

It is also vital to mention recruitment software producers since this industry is continually seeking to grow its software solution with extra and unique modules in order to boost client satisfaction and create additional earnings. As a consequence, software developers of recruitment solutions may benefit from outcomes that lead to a suitable applicant suggestion.

Job offers with descriptive features such as job title, location, firm, and necessary abilities were stored in databases As a result, only positions matching the specific search parameters may be discovered. When employing IR techniques, the full-text search is supplemented with keywords, allowing regular search engines to be incorporated. However, IR-based approaches enable the use of semantic relations in keywords, which is only partially offered by ordinary search engines. On the other hand, individuals have been proven to readjust the top-ranked results when presented with sorted lists of URLs using these strategies. For all these considerations, and independently of how job offers are processed and digested, suggesting the correct offer to the right user has always been a critical duty. The scientific community is using this strategy in an attempt to create answers that will meet the requirements of all parties participating in the process in an effort to establish a win-win situation for everyone concerned.

IV. EXISTING METHODS

Automatic resume recommendation techniques are especially; developed to solve the network congestion by optimizing content distribution for anonymous consumers based on their acquired inclinations, Nowadays, the most popular way to process this information is to automatically process the papers included in the e-recruitment process. In practice, most techniques attempt to use solutions based on the Vector Space Model to calculate the similarity ratio between the initial job description and the acquired application. It is an easy-to-implement method with extremely minimal computing costs that have traditionally produced very good results in the context of applicant or resume suggestion. However, new trends are betting on the application of machine learning technology to overcome old restrictions such as the inability to move beyond the syntactical representation of documents.

V. PROBLEM STATEMENT

With so many diverse employment opportunities available today, as well Because of the large number of applications that have been received, shortlisting is a difficult task for those in charge. This is exacerbated by the HR department's lack of diversified skill and topic knowledge, which is essential for successful screening. Being able to filter out irrelevant profiles as early in the funnel as feasible leads to cost reductions, both in terms of time and money.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The sector is now facing three key problems.

1) When it comes to sifting through applicants' resumes, the fact that they aren't, all the same, presents another challenge. Almost every resume on the market has a unique structure and format. HR must manually go through resumes to identify the best fit for the job description. This is a time-consuming and error-prone procedure in which a qualified candidate may be overlooked.

2) Identifying the most qualified candidates within a large pool of candidates - Finding the right candidate in India is practically impossible because there are so many individuals searching for jobs. This results in a lengthy and inefficient hiring procedure, which costs companies money.

3) Knowing that candidates are capable of performing the job before hiring them - The third and most difficult problem is matching comparing a resume with a job description to see if an applicant meets the requirements for the job for which they are applying.

In this paper, we provide a machine learning-based automated solution for dealing with the aforementioned difficulties in the resume short-listing phase. The model characterizes the qualities collected from the person's resume as input. Relying on the needed position description, the classified resume is plotted and the best-suited candidate's profile is recommended to HR. Our most significant contributions are mentioned below.

We created a mechanism for automatically recommending resumes. Machine learning-based classification techniques employing algorithms are used to find the most relevant resumes.

VI. METHODS

In order to improve prediction accuracy, significant research efforts have been made in recent years to identify methods of combining a range of fundamental techniques. To characterize fresh data, these approaches generate a set of theories and integrate the ensemble predictions in some way.

The accuracy attained by combining possibilities is typically greater than the precision of each different component. Random forests are one of the most used approaches in this area. The use of algorithms that depend on n-dimensional geometry and are given a collection of instances that have already been solved is also becoming more common.

Techniques based on n-dimensional geometry, which are provided a set of previously solved examples, are also gaining prominence. In this manner, the classes may be labeled and the algorithm trained to produce a geometric model that properly classifies a new sample. The key here is to be able to gauge the proposed method's performance with respect to past problems that have been addressed, each iteration of the procedure introduces new parameters into the algorithm, which then modifies the existing ones.

VII. RANDOM FOREST

The Random Forest is the first method that we consider while carrying out our research endeavors. The idea of RF is to deal with a set number of decision trees at once. Each tree casts a vote for a certain class. All trees iterate on this process. The RF then displays the results with the most votes. One of the benefits of employing RFs is that, in most cases, this method may avoid overfitting of the training set, which is not always achievable when using other ml algorithms. Figure 1 depicts an RF example. Please keep in mind that each decision tree must be designed in accordance with these instructions in order to function properly.

1) N is the number of test instances, and M denotes the number of variables in the classifier.

2) The number of input variables to be considered to generate a node's decision is m more m must always be less than M.

3) Select a training set for this tree and calculate the error using the remaining test instances.

Decide on each node of the decision tree by selecting m random variables at random. Use this variable to find the best way to distribute training sets.

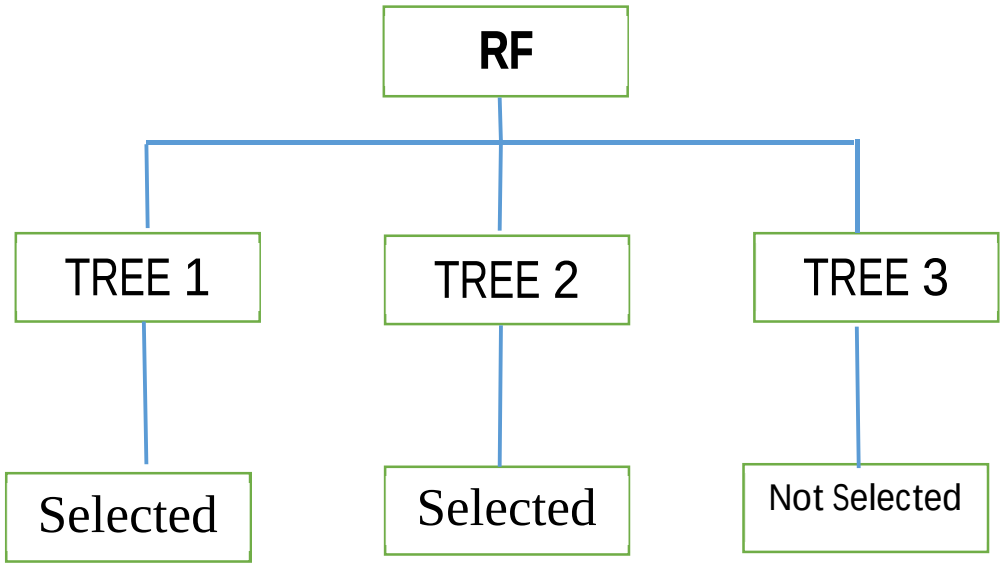
A major advantage of using RF, in this case, maybe summarised as follows:

a) As often, the number of trees in the RF is the only parameter that may be modified.

b) RF is founded on probabilistic concepts.

c) RF in general is easily extensible to handle numerous classes

FIG 1



For classification, the algorithm takes the majority vote, and for regression, the average of all votes, which in this instance is a classification, which means that the method takes the majority vote as the outcome and majority votes were selected, hence the output is selected.

VIII. SUPPORT VECTOR MACHINE

Support Vector Machines (SVM) is a cutting-edge classification technology that divides data samples based on the structural concept of hyper-plane. The SVM principle is extremely simple and straightforward: If we have already classified data samples, SVM may be used to construct multiple separation hyperplanes so that the previously classified data samples can be separated into segments. The concept is that each of these portions contains only one class. The SVM approach is often effective and relatively accurate in circumstances involving some form of classification. The rationale for this is because SVM is intended to decrease classification error while increasing geometric margin.

SVM has already shown excellent performance in a variety of settings involving data sample classification. We also believe that SVM has various advantages in the area of automated Resume suggestion.

These benefits are as follows:

1) SVM features a linearization method that helps it to prevent overfitting.
2) SVM is defined by an optimal solution for which there are a variety of efficient solutions.
3) SVM approximates a constraint on the test error, making it exceedingly resilient.

We only want the classifier that is closest to the hyperplane out of all those that can accurately categorize previous data, depicts the reasoning behind SVM using a two-dimensional space as an example. The goal here is to find the hyperplane that best separates the relevant resumes from the non-related resumes. When a new instance is added, this hyperplane must be updated to allow for future classifications.

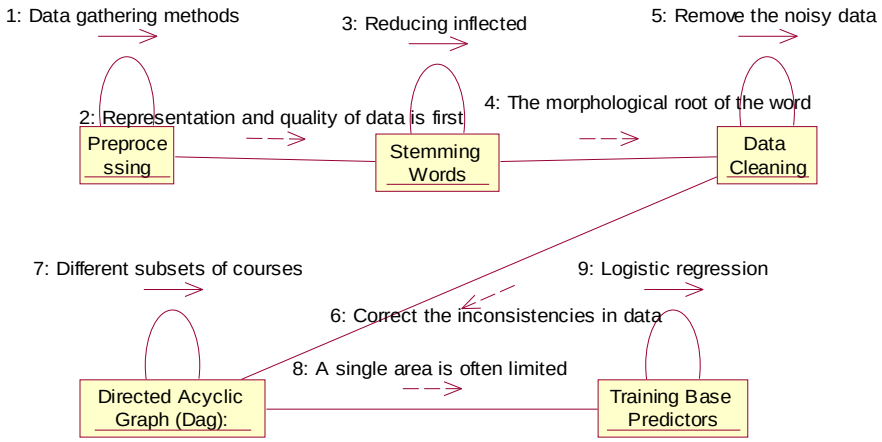
IX. METHODOLOGY

The goal of this task is to choose the most qualified individuals' resumes from a pool of resumes. To accomplish this goal, we created a machine learning-based solution. Figure 3 depicts the whole framework for the proposed concept. The suggested model operated primarily in two stages:

I) Prepare, ii) Deploy, and Inference Description of the underlying data set: Data was gathered from a variety of sources, including the portals mentioned above and Kaggle. The information is in Excel format, with three columns labeled ID, Class, and Resume text. Figure 4 demonstrates the number of occurrences for each domain.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

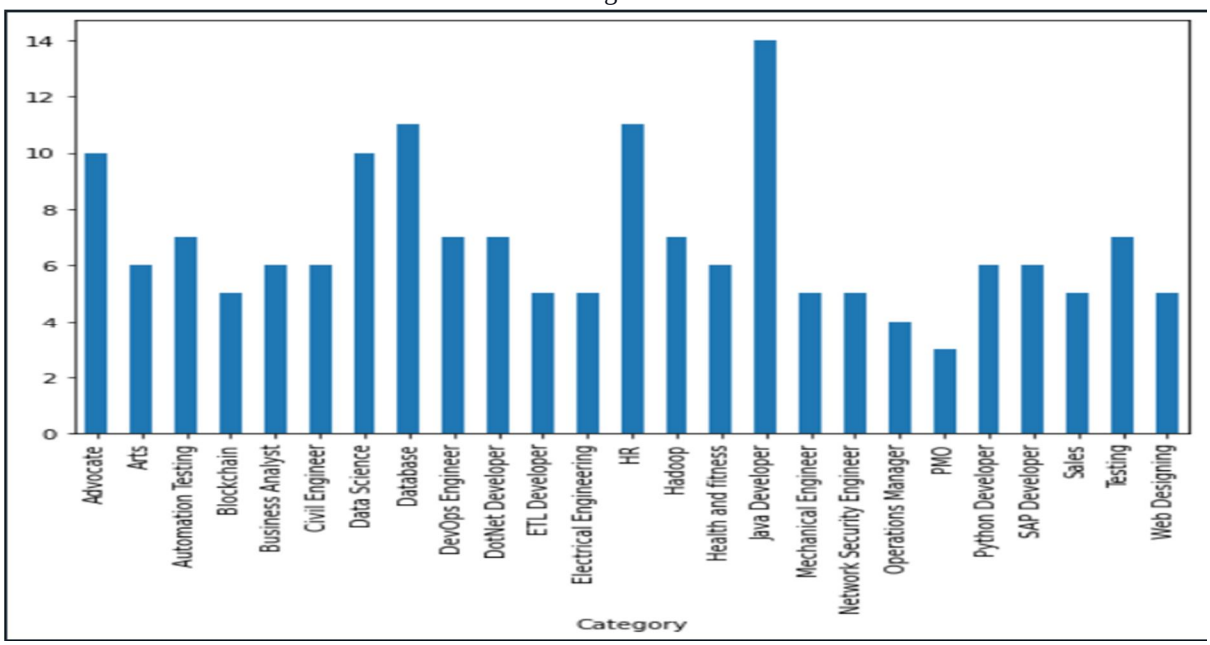
FIG 2



A. Pre-processing

Resumes submitted as input will be cleaned throughout this method to remove any special or rubbish characters. All special characters, digits, and single-letter words are eliminated during cleaning. Our dataset was free of special characters, digits, or single-letter words using these steps. The dataset is tokenized using the NLTK tokenizes. Preprocessing processes such as stop word removal, stemming, and lemmatization is then applied to the tokenized dataset. The raw resume file will be imported, and the resume data were cleaned to eliminate the numbers and additional spaces in the date

Fig3



B. Stop Words Removal

Stop words like and, the, was, and so on appear often in the narrative and are not useful in the prediction phase, therefore they are eliminated. Methods for Filtering Stop Words

1) The input words were tokenized into separate tokens and kept in an array for filtering out stop words.
2) Now, each word corresponds to a Stop Words list in the NLTK library.
3) From nltk.corpus import stopwords
4) From nltk.tokenize import word_tokenize, sent_tokenize

©IJRASET: All Rights are Reserved

5) Stop_words=stopwords.words('english')
6) Stop_words.extend(['from','subject','re','edu','use','email'])

It turns up a total of 185 stop words, which we can verify by using the code or command print(len(stop words)), so that it can be viewed, and to check the stopword, we can use the line print(stop words) function, which allows us to view all of the stopwords that were previously stored in the array we created.

C. Stemming

Despite stemming, lemmatization reduces inflected sentences to guarantee that the base word accurately pertains to the speech. Lemmatization often progresses as follows:

1) Create a glossary of words from the text corpus.
2) Generate a corpus concordance, which includes all of the word list entries as they appear in the corpus.
3) The word should be assigned to its lemmas based on the correlation.

The following phase is feature extraction. The data was preprocessed and the features were extracted using the TFidf transformer. Data that has been cleaned was imported. To process, the machine learning-based categorization model or learning algorithms require a fixed-size numerical vector as input. ML-based classifiers did not parse input text with varying lengths. As a consequence, the texts are converted to the requisite equally spaced vector form throughout the preparation operations. Using the scikit learn library function, sKlearn.feature_ectraction.text.countvectorizer is used to calculate count vector.

Deployment and Inference:

In this phase, the tokenised resume data and the job descriptions (JD) will be compared, and the model would deliver resumes relevant to the job description as an output, as well as feedback to the job applicant on which profession his resume fits more closely.

X. RESULTS

Classifier	accuracy
Random Forest	0.589
Naïve Bayes	0.224
Logistic Regression	0.385
Linear Support Vector Machine	0.787

We determined this model to be dependable and best suitable for our purpose since its accuracy score was better than that of other models. Continuing with our best model, we will examine the confusion.

We began our study with the Naive Bayes classifier, and the retrieved tf-idf feature set is input into the RF classifier, which recommends the resume. The NB classifier has an average accuracy of 22.4 percent. Because the acquired results were unsatisfactory, we employed another prominent classifier known as "Logistic Regression" for this purpose. The LR classifier predicted the resume with an accuracy of 38.5 percent, which was higher than the performance of the previous classifier (LR).

However, the LR classifier's 38.5 percent accuracy suggested that more than half of the resumes were incorrectly forecasted. On the very same data, we tried another classifier, "Linear Support Vector Machine," and got 78.7 percent accuracy. The Random Forest classifier was employed to increase model accuracy and produced 58.9 percent accuracy, which was worse than the accuracy of SVM.

XI. CONCLUSION

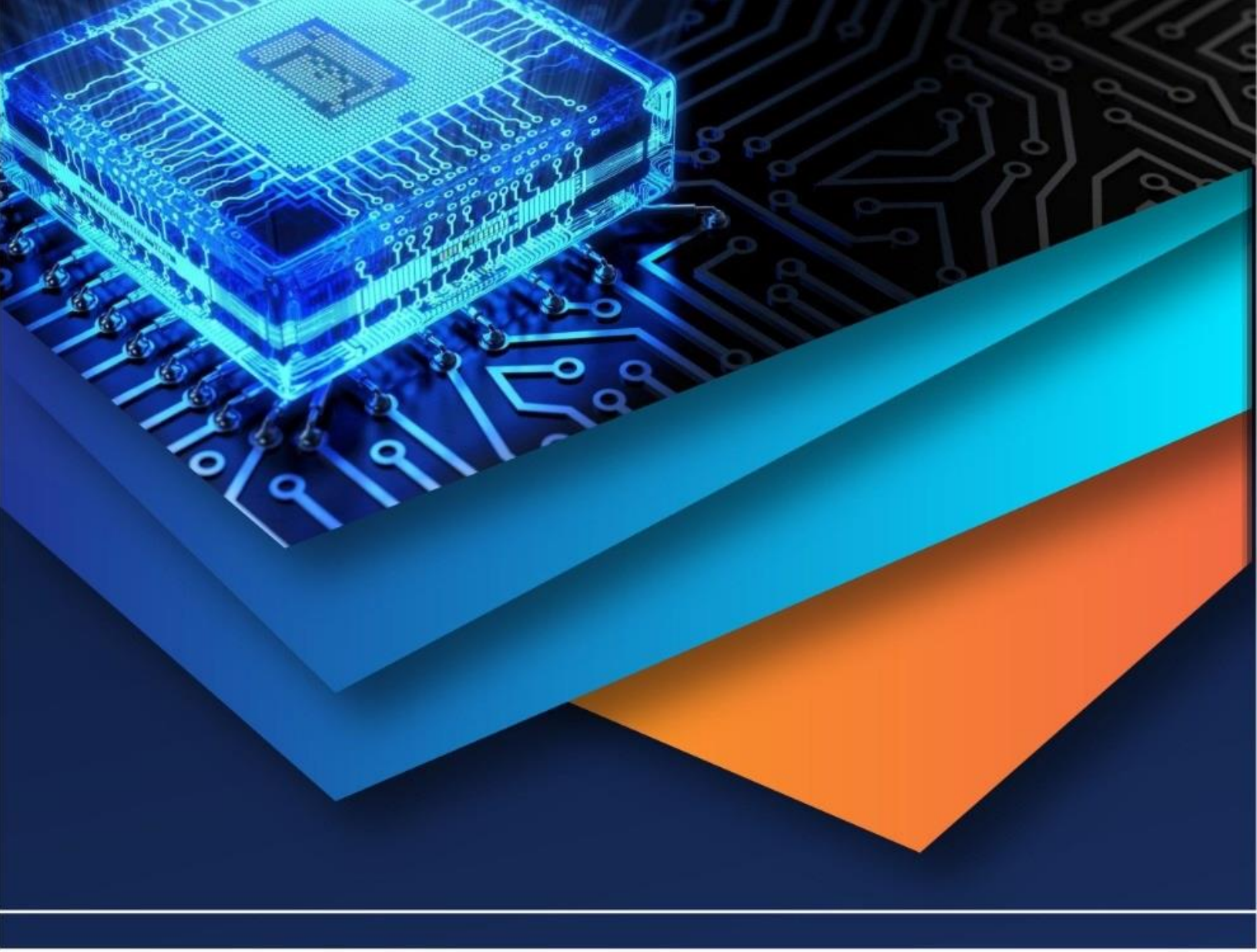
We provided our concept for automated resume recommendation in this paper. Our objective here is to be able to construct ways and give acceptable job applicant profiles to the needed job descriptions that may be interested in them. Today, we have focused our research efforts on two well-known classification methods: random forests and support vector machines. The company gets a large number of applications for each job posting.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Finding the appropriate candidate's application from a collection of applications is a time-consuming chore for every firm these days. The process of categorizing the prospective employee's résumé is one that is arduous, time-consuming, and resource-intensive. To address this issue, we suggested an integrated machine learning-based algorithm that suggests acceptable applicant resumes to the HR job specification provided. A two-step process was devised: First, the resumes were categorized into several categories. Second, it suggests resumes based on similarities with the job description. The suggested method effectively captures resume insights and semantics, yielding an accuracy of 78.7 percent using the Linear SVM classifier. If an industry has a high number of resumes, the suggested technique may be used to construct an Industry-specific model. A more realistic model may be built by incorporating specialists in the field, such as department professionals, who can provide input on the model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)