RESEARCH ARTICLE   OPEN ACCESS

Resume Screening and Course Recommendation System

Payal Bhosale**, Sanyogita Londhe**, Pooja Meher**, Pratiksha Raut**

*(Prof.V.G.Jagatp, Computer Engineering, SVPM's College of Engineering Malegaon (Bk), Baramati
Email: hodcomp@engg.svpm.org.in)
** (UG Students, Computer Engineering, SVPM's College of Engineering Malegaon (Bk), Baramati)
Email: payalbhosale570@gmail.com, sanyogitadlondhe@gmail.com, poojameher2222@gmail.com,
rautpratiksha0802@gmail.com)

\-----------------------------------************************----------------------------------

Abstract:

A typical Internet job posting attracts a large number of applications in a short amount of time. It is not practicable to manually filter out the resumes because it would take a lot of time and money, both of which the employing organizations cannot bear. In addition, this technique of filtering resumes is not fair as many suitable profiles don't get enough consideration which they need. This could lead to the hiring of unsuitable candidates. By suggesting the best candidates in accordance with the provided job description, the system will provide a solution that seeks to address these problems. System uses Natural Language Processing technique pyres parser to extract relevant information like skills, education, experience, etc. based on the unstructured resumes, producing a summarized form for each application as a sentence parse tree and storing extracted data in MySQL database. The job description system will perform text processing and the preferred TF-IDF algorithm will extract relevant data from job description. The work of screening is simpler and recruiters may analyze each resume more thoroughly in less time when all the unnecessary data has been excluded. After the completion of this text mining procedure, the suggested solution will compare resume extracted data from database to job description. For model training LSTM algorithm will be used additionally, the algorithm will suggest courses based on market need for those who, unfortunately, do not fit the job profile. [3].

Keywords -- NLP, LSTM, Vectorization, TF-IDF

\-----------------------------------************************---------------------------------

I. INTRODUCTION

Machine learning is a leading technology that enables computers to learn automatically from past data. Machine learning is a technology where we train a model with a dataset to predict the desired output when we give new data. Resume screening is mostly done using Natural Language Processing (NLP), Natural language refers to the way that humans communicate with each other.

Natural language processing (NLP) is the capability of a computer program to understand human language as it is written and spoken. Recruitment is a 200-billion-dollar business. It deals with hiring the best fit candidates who have the relevant skills for a given job profile from an immensely large pool of candidates. If a company has any job openings for a position, scores of candidates mail their resumes to the company to apply for that opening. As the job market is growing in India,

millions of new job seekers are joining the workforce every year, as per LinkedIn [19]. According to 2021 Employees, new jobs were created. As of this year, the unemployment rate of India is around 7.74% where the urban area has an unemployment rate of 9.06 and the rural area is 7.13%. The number of job vacancies available is not enough to cover the staggering amount of applications the companies will receive. Hence, if the companies are hiring in bulk, there are many applications to find the talent that they need which will require a considerable amount of resources and The solution to this problem is to fill in the spot and get the job done with less amount of resources costing to the company with an limited timeline. Even here the applications are in millions which is a tedious task to manually go through each and every resume, hence these companies use various Machine learning models for ranking out the top resumes, which are the best suitable for the job role.

II.LITERATURE SURVEY

A] Machine Learning approach for automation Resume Recommendation System

Through their research, the author created a system [3] for finding new hires for available positions in the company.

Their main focus was on three things:

1. Choosing the top applicants from the pool of submissions;

2. Evaluating resumes;

3. And figuring out whether a candidate was qualified for the job.

In order to perform NER, NLP, and text classification using n-grams, they used machine learning. They used the algorithms of Random Forest, which had an accuracy of 38.9%, Multinomial Nave Bayes, which had an accuracy of 44.39%, Logistic Regression, which had an accuracy of 62.4%, and Linear Support Vector

Machine Classifier, which had an accuracy of 78.53%.

Sr.no	Title	Algorithm	Accuracy
1.	A Machine Learning approach for automation of Resume Recommendation System	SVM, Cosine Similarity	78.53%
2.	Skill Finder: Automated Job-Resume Matching System	NER, Stanford Name Entity Recognizer	87%
3.	Resume Classification and Ranking using KNN and Cosine Similarity	KNN, Cosine Similarity	79.8%
4.	Web Application for Screening Resume	NLP,Spacy, Machine learning algorithm	98.96%

Table.1.Accuracy Table

B] Skill Finder: Automated Job-Resume Matching System

Selecting the right candidate for a job is a tedious process which includes the following tasks:

1] Posting a job

2] Resume screening based on qualifications

3] Selection based on knowledge and skills

For this tedious task Skill Finder is a web application that was developed using J2EE technologies with Apache Tomcat. Server.The author[4], in their research, created a system where, they have created a dataset of 3000 jobs and

also 80 resumes from the website indeed, using the API.In this system, an automatic skill matching algorithmis used by reducing the labour intensity of manual selection of resumes. Then this data is used for ranking the student's resumes comparing their skills required in the job, this process is done using the NamedEntity Recognition (NER) like Apache OpenNLP and Stanford Name Entity Recognizer. Students and recruiters can access theapplication over the internet and perform the tasks which they are authorised to do, like post jobs, apply tojobs and receive email alerts. The skill finder web application matches the resumesaccording to the jobs posted and sends email alerts to the desired candidates.

C] Resume Classification and Ranking using KNN and Cosine Similarity

Resume Classification and Ranking using KNN and Cosine Similarity: Using KNN and cosine similarity, the authors developed a technique to achieve resume classification and Ranking [2]. Using NLP methods like Tf-Idfvectorizer the candidate's résumé is extracted. For document similarity Cosine similarity is used to compare the resumes and the job description. KNN classifies resumes that are most similar to the job description. This trained model has an accuracy of 98.96%.

D] Web Application for Screening Resume

The author [5] mission was to develop a web application for resume screening. The web application is divided into 3 parts:
A) Job Applicant side
B) Server-Side
C) Recruiter Side

In applicant side job seekers upload their resume and resume is extracted by using NLP and stored into a database. At the server side, first, a collection of resumes relevant to the specific type of job post is made. The appropriate text from each resume is then manually uploaded as a JSON file from its ZIP format onto an internet gateway. Passing this JSON file to the NLP module and trained module is created. The job description from the server side passed to pre-trained modules and finding relevant entities. These relevant entities are stored into a database. The rank list of resumes [6] will be displayed on the recruiter's end so that the recruiter may choose the applicant who is most suitable for the position. As per accuracy table accuracy is 98.86%.
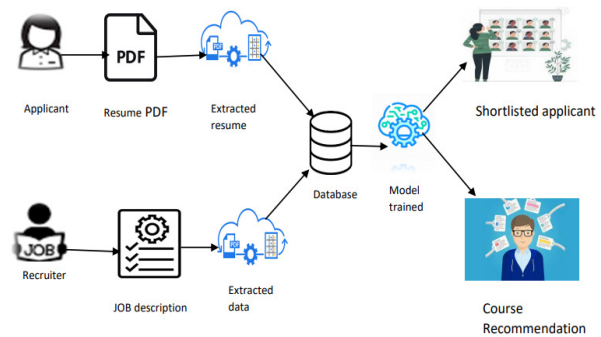
III.TECHNICAL DETAILS

Fig.1.Proposed System

A. Dataset

1. Course dataset: Coursera dataset
2. Job-Dataset: job-listing-dataset-monster-uk

B. DATA PREPROCESSING AND FEATURE EXTRACTION

The following are some libraries that are used for data pre-processing and feature extraction:

1. Pyresparser:

A simple resume parser is used for extracting information from a resume [20]

2. TF-IDF:

TF-IDF stands for Term Frequency Inverse Document Frequency of Records It is the

calculation of how relevant a word in a series to the given text. The meaning increases proportionally as the number of times in the text a word
appears is compensated by the word frequency in the corpus (data-set).

Term Frequency: This measures the frequency of a word in a document.

tf(l,f) = count of l in f / number of words in f

Inverse Document Frequency:

IDF is the inverse of document frequency which measures the in formativeness of term t.

idf(t) = N/df

3. Word cloud:

A word cloud is a visual representation of object for text processing.

C. MODEL BUILDING

Long short term memory (LSTM):

Long-short-term memory (LSTM) is a deep learning architecture that employs a recurrent neural network (RNN) architecture (DL). It can analyze whole data sequences as well as single data points (such as photos or videos).LSTM can be used for problems like unsegmented, linked handwriting recognition. A cell, an input/output gate, and a forget gate are the components of a LSTM unit. The three gates control the flow of data in and out of the cell, and the cell remembers values across arbitrary time intervals. LSTMs were created to solve the task of vanishing gradients that can occur when training traditional RNNs. In many problems, LSTM has an advantage over RNNs, hidden Markov models, and other sequence learning approaches due to its relative insensitivity to gap length.

Natural Language Processing, or NLP:

is a field of artificial intelligence that gives machines the capability to read, understand, and derive meaning from human languages. Some NLP techniques:

Tokenization:

In essence, it's the process of cutting a text into pieces called tokens and at the same time cutting certain characters, such as punctuation.

Stemming:

Refers to the process of cutting the ends or the starting points of words with the intention of removing affixes

Lemmatization:

This reduces words to their lexicon form, or "lemma," for which it calls for comprehensive dictionaries that the computer may search through and connect words to.

D.EXCEPTED RESULT

1. Shortlisted candidate/applicant

2. Course Recommendation

IV. CONCLUSION

The current paper deals with multiple ways to detect, identify and classify various resumes using multiple machine learning and Neural Network models like NLP, KNN, Word2Vec, Cosine similarity, NER etc. The result of the models varies based on the datasets used, the complexity of the learning methods and the size of the dataset, the results range from 78% to 98%. We conclude that with a proper dataset and with the right algorithm, we can achieve the desired accuracy and result.

G. REFERENCES

[1]. The data source for the skills used in the NER train.

[2]. Jagan Mohan Reddy D, SirishaRegella., "Recruitment Prediction using Machine Learning", IEEE Xplore, 2020.

[3]. Pradeep Kumar Roy, Vellore Institute of Technology, 2019. A Machine learning approach for automation of resume recommendation system, ICCIDS 2019. 10.1016/j.procs.2020.03.284.

[4]. Thimma Reddy Kalva, Utah State University, 2013. Skill-Finder: Automated Job-Resume Matching system. 3]Yong Luo, Nanyang Technological University, 2018. A LearningBased Framework for automatic resume quality assessment, arXiv:1810.02832v1 cs.IR].

[5]. Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna, Design and Development of Machine Learning based Resume Ranking System (2021),

DOI: https://doi.org/10.1016/j.gltp.2021.10.002.

[6]. Jabri, Siham, AzzeddineDahbi, TaoufiqGadi, and AbdelhakBassir. "Ranking of text documents using TF-IDF weighting and association rules mining." In 2018 4th international conference on optimization and applications (ICOA), pp. 1-6. IEEE, 2018.

[7]. The data source for the skills used in the NER train.

[8]. Jagan Mohan Reddy D, SirishaRegella., "Recruitment Prediction using Machine Learning", IEEE Xplore, 2020.

[9]. Resnick, P., Varian, H.R.,1997.RecommenderSystems.Communications of the ACM40, 56– 59. [10]. Xavier Schmitt, Sylvain Kubler, Jer my Robert, Mike Papadakis, Yves LeTraon University of Luxembourg, Luxembourg Replicable Comparison Study of NER Software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate.

[11]. Y. Luo, Y. Wen, T. Liu, and D. Tao, "Transferring knowledge fragments for learning distance metric from a heterogeneous domain,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[12]. Mikheev, Andrei; Moens, Marc; Glover, Claire. 1999. "Named Entity Recognition without Gazetteers." Proceedings of EACL '99. HCRC Language Technology Group, University of Edinburgh. http://acl.ldc.upenn.edu/E/E99/E99-1001.pdf.

[13]. Zhou, GuoDong; Su, Jian. 2002. "Named Entity Recognition using an HMM-based Chunk Tagger." Proceedings of the Association for Computational Linguistics (ACL), Philadelphia, July 2002. Laboratories for Information Technology, Singapore

[14]. Zhang, L., Fei, W., Wang, L.,2015.Pjmatchingmodelofknowledgeworkers. Procedi acomputerscience60,1128–1137

[15]. http://www.indeed.com/isp/apiinfo.jsp

[16]. https://opennlp.apache.org/documentation/1.5.3/ manual/opennlp.html#tools.namefind.recognition

[17].https://nlp.stanford.edu/software/CRFNER.shtml

[18].Resume Screening Using Machine Learning and NLP

[19][7]. Suhas H E, Manjunath AE, "Differential Hiring using a Combination of NER and Word Embedding", In 2020 International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Vol.9

[20] https://pypi.org/project/pyresparser/