International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228240

Resume Screening Using Machine Learning and NLP : A Proposed System

Bhushan Kinge*1, Shrinivas Mandhare2, Pranali Chavan3, S. M. Chaware4
1-3UG Student, Information Technology, SPPU, Pune, Maharashtra, India
4Professor, Information Technology, SPPU, Pune, Maharashtra, India

ABSTRACT

Article Info
Volume 8, Issue 2
Page Number : 253-258

Publication Issue :
March-April-2022

Article History
Accepted: 15 April 2022
Published: 25 April 2022

The Indian Recruitment market has grown substantially over the last half-decade as the need for cheap labor grows and the number of job openings is increasing. And as the job market increases so does the recruitment industry which is a new way of hiring people by outsourcing the hiring process itself to other companies whose sole purpose is to give the correct talent required for the company. This is done because these companies are hiring in bulk and doing such a thing in-house will require a lot of company resources which will hamper productivity. As such companies emerge even for them manually going through all of the Resume of candidates is very time-consuming and tedious so these Talent Acquisition Companies use various Machine Learning models to filter out top resumes according to the job roles, which reduces the efforts for the Human Resource team

Keywords: NLP, Resume, CV, KNN, SVM, NER

I. INTRODUCTION

Machine learning is a field where we train a model with a dataset to predict the desired output when given new data. Screening the resumes is mostly done using Natural Language Processing (NLP), Natural language refers to the way we humans communicate with each other. NLP is concerned with giving computers the ability to understand the text and spoken words in much the same way human beings can. NLP combines computational linguistics- rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, combining these technologies helps computers process the way human language works in the form of texts or voice data and to 'understand' its full meaning. As the job market is growing in India, millions of new job seekers are joining the workforce every year, as per LinkedIn [7]. Around 1.3 million new jobs were created as per2021 Employees Provident Fund Organization (EPFO) [8]. As of this year, the unemployment rate of India is around 7.74% [6] where the urban area has an unemployment rate of 9.06% and the rural area is 7.13%. The number of job seats available is not enough to cover the staggering amount of applications the companies will receive.

Hence, if the companies hire in bulk there are many applications to find the talent that they need which will require a considerable amount of resources and time, this problem Talent acquisition Companies arise as solutions for this problem who fill in the spot and get the job done with less amount of resources costing to the company with an acceptable timeline. Even here the applications are in millions which is a tedious task to go through them hence these companies use various Machine learning models which will rank out the top resumes which are the best fit for the job role.

II. LITERATURE

A. Machine Learning approach for automation Resume Recommendation System

Pradeep Kumar Roy in their research [1], created a system where they can minimize the cost of hiring new candidates for the job positions in the company. They focused on 3 major problems in this process

- Picking the right candidates from the applicants
- Making sense of their CV's
- Finding out if the candidate is fit for the job role

They performed NER, NLP, and text classification using n-grams and used Machine Learning to perform the classification using the algorithms of Random Forest with 38.9% accuracy, Multinomial Naïve Bayes with 44.39%, Logistic Regression with 62.4%, and the highest accuracy was obtained by Linear Support Vector Machine Classifier with an accuracy of 78.53% .

B. Skill Finder: Automated Job-Resume Matching System

In research conducted by Thimma Reddy Kalva [2], they have developed a custom dataset of 3000 jobs and 80 resumes from the website indeed using the web service API [9]. This data is then used to rank the student's resumes comparing their skills required in the job, this is done using the Named Entity Recognition(NER) like Apache OpenNLP [10] and Stanford Name Entity Recognizer [11]. The Skill finder efficiently matches the Resumes according to the job role posted and successfully sends emails to the desiredcandidates

C. ResumeNET : A Learning-based Framework for Automatic Resume Quality Assessment

In research conducted by Yong Luo [3], they have developed a custom dataset of 10,343 resumes which was acquired by a private resume management company. 98.82% (i.e 10,221 resumes) data is unlabeled and the remaining 1.18% (i.e. 122 resumes) data is labeled in 2 categories positive and negative, 33 and 89 of them are labeled as positive and negative.

D. Web Application for Screening Resume

The goal for Sujit Amin [4] was to develop a web application for resume screening, with the help of 220 resumes out of which 200 were used for training and 20 used for testing purposes, further, the web application is divided into 3 divisions

A) Job Applicant side

B) Server-Side

C) Recruiter Side

The applicant side is where the applicant will provide his/her resume, the server-side will process the resume and then be trained using the NLP Pipeline which used SpaCy which is an NLP framework. On the recruiter's side, the rank list of the resumes will be shown which was decided from a score calculator [13] so the recruiter can select the best fit candidate for the job.

E. Design and Development of Machine Learning based Resume Ranking System

The proposed system hereby Tejaswini K [5] is where the resume is submitted by the candidate after an

MCQ test which has a face detection system to detect malpractice. Once the resume is submitted it's run through NLP techniques to get the relevant skills from the resume and use TF- IDF vectorization [14] to convert the words into vectors so the machine can understand it.

The classifier used here is the KNN algorithm to identify the resume that closely matches the JD provided by the recruiter. The system has an average parsing accuracy of 85%.

F. Resume Classification and Ranking using KNN and Cosine Similarity

Riza Tanaz Fareed, Rajath V, Sharadadevi Kaganumath came up with a method to implement the Resume classification with the addition of cosine similarity [15]. The process is the candidate provides his/her resume to the system.

The resume is then passed through an NLP pipeline where the words are extracted out of the resume. Techniques like stop words, lemmatization are used to get the correct set of words. TF-IDF vectorizer [14] is used to vectorize the words for the KNN model to classify the resume into various categories. Now to evaluate the resume on the given JD document similarity detection is necessary so the Cosine Similarity Algorithm is used in which the JD content is matched with the candidate's resume. The accuracy for this trained model is 98.96% .

G. Automated Tool for Resume Classification using Semantic analysis

This study conducted by Suhas Tangadle Gopalakrishna and Vijayaraghavan Varadharajan [16] provides a descriptive view of how they use semantic analysis for resume classification. The received resumes by the HR team are parsed through the Natural Language Processing Pipeline (NLPP) where the Stop deletion is used to delete the words like "and, or, the, etc.". Other techniques like Parts of Speech tagging and NER are also used.

Total 6 types of classification models are used: Naïve bayes, Multinomial Naïve Bayes, Linear SVM, Bernoulli Naïve Bayes, Logistic Regression and KNN. These classification models are run on a dataset of 30,000 employees' resumes which was split in a 9:1 ratio where 27,000 were used for training and 3000 used for testing across various domains including AI, Computer Architecture, CG, Databases, Distributed Computing, CN, Web Technologies and Cloud Computing. Out of all the 6 Classifiers Multinomial Naïve Bayes came out with a top accuracy of 91%.

H. Differential Hiring using a Combination of NER and Word Embedding

The objectives of this examination conducted by Suhas H E and Manjunath A E [17], were to create a model which uses NLP, NER Word embedding, and Cosine Similarity to suggest Resume for Job roles. Resumes and the JD is taken as the input by the system.

Data dump of technical skills [18] used to tag technical skills in each resume document. Tab-separated value (TSV0 file is generated and that file is provided to the Stanford NER model [10] to train the NER model. The output of the NER model (i.e. skills) becomes the input for the word2vec model which uses a shallow neural network.

The last step of the process is Cosine Similarity which determines how much the given resume matches the given JD. The accuracy obtained was 79.8%.

III. LIMITATIONS

- The Above systems has models that don't have any way to improve themselves over the time, the models will be trained only once.
- The above models used Machine learning algorithms which have a tend to plateau in performance when runned over a large dataset.

IV. PROPOSED SYSTEM

A. Problem Statement

Resume Screening is a process that is majorly used in Big Tech companies where they receive a massive amount of resumes, and rank them according to resume strength or how much the resume is relevant to the job description and filtering them according to that. but the student who applies for the job role doesn't know why his resume got rejected and how he can improvise so his resume can become relevant and strong. Currently, there is no such technology available that would benefit the students which can help them strengthen their resume.

B. Solution

The solution for the given problem statement will consist of the machine learning model which takes the student resume as input and extracts the details like skills, certifications from it. for the extra details about the student, it also takes GitHub and LinkedIn profile links where it can extract the student contribution in various fields. The student also has to provide which job role he/she is applying for. The model is trained using a job description and skill set dataset. So ,When the resume is inputted by the student it can tell which job role is suitable for you or how your resume is relevant to the given job description.

C. System Architecture

Below given is the System architecture of our proposed design Fig:1, this shows the entire working of our model and the parts included in it represents the flow of the system.

Figure 1. System Architecture.

The system consists of a database; it will be an SQL database as our data is properly arranged in columns and rows rather than being unstructured. The model will be trained on the existing data which we collected from the open platform of Kaggle. There are 2 Models used in this method the first one being either K-Nearest Neighbour or Support Vector Machine which will help us to get the prediction of what kind of job role our resume is best fit for and the second model will give recommendations of how we can improve our resume to increase its strength by using cosine similarity which will check the user's input of what job role they want what the model predicted on that basis the Recommendation system will give its suggestions for the improvement.

The control flow will be in the following manner, the candidate submits the resume at the front-end the resume is then passed to the resume parser which is a pipeline of NLP techniques that will extract useful information from the resume, and then the system will visit the person's LinkedIn and GitHub profile to scrape useful information from the website which adds more value to the overall extracted data to from vectors and provide it to the Machine learning Model for the prediction.

D. Accuracy Table

Table1: Accuracy in % for different methods

Title	Accuracy
A Machine Learning approach for automation of Resume Recommendation System	78.53%
Skill Finder: Automated Job-Resume Matching System	87%
ResumeNet: A Learning-based Framework for Automatic Resume Quality Assessment	85%
Web Application for Screening Resume	98.96%
Design and Development of Machine Learning based Resume Ranking System	79%
Resume Classification and Ranking using KNN and Cosine Similarity	79.8%
Automated Tool for Resume Classification using Semantic analysis	80%
Differential Hiring using a Combination of NER and Word Embedding	82.67%

Table1 shows accuracy in % when different methods are used for resume recommendation.

V. CONCLUSION

This Paper deals with multiple methods to detect, identify and classify various resumes using multiple machine learning and Neural Network models like SVM, KNN, Word2Vec, Cosine similarity, etc. The accuracy 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 the right algorithm we can get good accuracy and desired output for a large variety of purpose.

VI. REFERENCES

[1]. 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.

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

[3]. Suhjit Amin, Fr.Conceicao Rodrigues Institute of Technology, 2019. Web Application for Screening resume, IEEE DOI: 10.1109/ICNTE44896.2019.8945869.

[4]. 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.

[5]. Riza tana Fareed, rajah V, and Sharadadevi kaganumath, "Resume Classification and Ranking using KNN and Cosine Similarity" In 2021 International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol.10.

[6]. Suhas Tangadle Gopalakrishna, Vijayaraghavan Varadharajan, "Automated Tool for Resume Classification Using Semantic Analysis", International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 10, No.1, January 2019

[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

[8]. Centre for Monitoring Indian Economy Pvt Ltd. (CMIE),2022. The unemployment rate in India.

[9]. Howard, J.L., Ferris, G.R., 1996. The employment interview context: Social and situational influences on interviewer decisions 1. Journal of applied social psychology 26, 112-136.

[10]. Mudit Kapoor, Business Today, 2021. India's formal job creation numbers beat pandemic blues.

[11]. M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," Journal of Machine Learning Research, vol. 7, pp. 2399–2434, 2006.

[12]. A. Zaroor, M. Maree, and M. Sabha, "A Hybrid Approach to Conceptual Classification and Ranking of Resumes" In Czarnowski I., Howlett R. (eds) Intelligent Decision Technologies 2017. IDT 2017. Smart Innovation, Systems and Technologies vol 72. Springer.

[13]. Jabri, Siham, Azzeddine Dahbi, Taoufiq Gadi, and Abdelhak Bassir. "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. .

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

[15]. Jagan Mohan Reddy D, Sirisha Regella., “Recruitment Prediction using Machine Learning”, IEEE Xplore, 2020.

[16]. Resnick, P., Varian, H.R.,1997.Recommender Systems.Communications of the ACM40, 56–59.

[17]. 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.

[18]. 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.

[19]. 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.

[20]. 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

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

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

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

[24]. https://nlp.stanford.edu/software/CRF-NER.shtml

Cite this article as :

Bhushan Kinge, Shrinivas Mandhare, Pranali Chavan, S. M. Chaware, "Resume Screening using Machine Learning and NLP: A proposed system", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 2, pp. 253-258, March-April 2022. Available at doi : https://doi.org/10.32628/CSEIT228240
Journal URL : https://ijsrcseit.com/CSEIT228240