e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

RESUME RECOMMENDATION SYSTEM USING COSINE SIMILARITY

Anushka Agarwal*1, Dr. Senthilkumar*2

*1,2Computer Science Engineering, SRM Institute Of Science And Technology, Chennai, India.

ABSTRACT

A company's progress gets slowed down if a wrong person gets recruited for the job position. It is a tedious task to find a suitable candidate for an open position when there are many candidates. This research paper showcases about the different recommendation techniques which can make the recruitment process easier and faster. Classification is done as per the job description, which then shows the percentage match of the resume with respect to the job description. Only if there is a 50 % match between the resume and the job description can the recruitment process move any further. Domain knowledge is required for screening of resumes. India is big job market where millions of people seek jobs and it is a tough task to separate the right candidate from this huge market. And so, hiring costs many resources to the company. Resumes do not have a standard format; every resume has a different format and a different structure. The recruitment team then has to manually match each resume with respect to the job description. Manual process has a high chance of missing the right candidate for the job within the process. The classification techniques here come to play and make it easier for the company as well as the candidate it will add value to the company's recruitment drives, these techniques are being researched to bring a more hassle-free experience to the recruitment process.

Keywords: Resume, Machine Learning, NLTK, Cosine Similarity, Decision Tree, Logistic Regression, Resume Matching, Similarity Measure, Content Based Filtering.

I. INTRODUCTION

In today's technologically growing world, recruitment process for the corporate world is getting evolved to a great extent. Hard copies are no longer used for submitting the resumes the candidates and also the recruitment teams want an e-resume which can be viewed online. Validating resumes online is not much flexible and is vulnerable to manual errors. Man power would be required to check the resumes of the candidates. Using the classification techniques for recruitment has a greater benefit over the resumes which are submitted as a hard copy where monitoring the suitable candidates resume is difficult. Even after e- documents took over the hard copies, companies still found it difficult to manage the huge amount of data correctly. According to Jobvite's report 2014, 68% of online jobseekers are either undergraduates or postgraduates. The aim of our paper is to help the recruiting companies to find the most appropriate resume that caters to all the requirements for the actual job description and help them find the foremost suitable candidate in less amount of time and in less resources. This paper portrays the use of collaborative filtering, similarity techniques indicating the percentage match of the resume with the job requirements only when the minimum criteria is met only then the candidate is further recommended for the further rounds of recruitment.

II. LITERATURE SURVEY

[1] Aleksandra Pawlick a, Marek Pawlicki, Rafał Kozik and Ryszard S. Chora "A SYSTEMATIC REVIEW OF RECOMMENDER SYSTEMS AND THEIR APPLICATIONS IN CYBERSECURITY".

This research paper, shows the role of recommendation system in the field of cybersecurity. The paper presents types of recommender systems along with their advantages and disadvantages, their applications and the security concerns. The paper then presents the utilization of recommender systems in the field of cybersecurity; presents solutions as well as future ideas are presented. There are two major contributions of this paper. It shows a comprehensive survey of the type of recommenders.

[2] Yi, X., Allan, J., Croft, W.B. " MATCHING RESUMES AND JOBS BASED ON RELEVANCE MODE AND MODELS".

This article has demonstrated that router misconfigurations are prevalent and can have serious effects for a network's functioning. If misconfigurations are discovered, the network's security will be jeopardized, and it may even create interruptions. This study focuses on identifying IP address or other state misconfigurations that might lead to an anomaly. The disadvantage of this system is that it is only capable of detecting IP address misconfiguration and does not identify additional threats.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

[3] Mahesh wary, S., Misra, H. "MATCHING RESUMES TO JOBS VIA DEEP SIAMESE NETWORKS".
This research paper predicted the problem of matching job descriptions with semi structured resumes where there is a largescale data collection. In this paper the project compared structured relevance models which is an extension of language-based model to the standard approaches of the recommendation system. Structured Relevance model is an extension for retrieving semi structured e-documents. These models after experiments have performed better than unstructured datasets.

[4] Sarabjeet Singh Chowdhary, Pradeep Kumar Roy, Rocky Bhatia, "A MACHINE LEARNING APPROACH FOR AUTOMATION OF RESUME RECOMMENDATION SYSTEM".
In the research paper proposed an automatic machine learning model in which most suitable resume was submitted to the HR with the description that was provided. The models worked in two particular phases. In the first phase the resume is classified into categories. In the second phase, it recommends resume with the similarity index with respect to the description. The model captured the keywords from the resume and semantics and resulted in providing an accuracy of about 78% using SVM classifier.

[5] S.T., Ykhlef, Al-Otaibi," "A SURVEY OF JOB RECOMMENDER SYSTEM".
In this paper literature analysis of many journals related to job recommendation searches were studied. The recommendation system technologies achieved success using range of applications and recommendation techniques.
Survey shows that many approaches for job recommendation systems were proposed combined provided the fit between recruiters and candidates. The paper presented the art of recommendation, it reviewed the typical filtering techniques and also related processes. It concluded that the job recommendation sector still needs further improvements.

Problem Statement

"Creation of a Job based recommendation system for job providers and job seekers using the candidate's skill set and recruiter's requirements by using Machine Learning ". Extracting data from the unstructured resume is the most difficult task. Resume is differently structured depending from candidate to candidate and extracting key information from them is challenging, the e-documents have different data fonts, formats, layout and the writing style is different for each candidate. Classifying such resumes manually is not possible. In a huge market like India companies receive thousands of resumes. The categorization of resume using filtering, similarity techniques will make the process a work of minutes. manual intervention would not be required. To solve this problem, screening of resume can be done using Machine Learning, Natural Language Processing using Python.

Proposed Work

Natural Language Processing is used to extract information from the unstructured and wide formats of the resumes. It creates a summarized version of every resume which has only the entities that are pertinent to the choice process the proposed model works in phases of categories: first the classification of resumes is done into different categories. Second it displays the similarity index with the given verbal description. The technique used to calculate similarity index is cosine similarity. In the final phase, the document which cross the bare minimum similarity value are suggested to the recruiter. Similarity index displays weather or not the candidate is suitable for the duty profile.

CLASSIFICATION TECHNIQUES

Similarity Index -Similarity index is used to calculate rank of the documents by calculating the similarity between the document and the query. Commonly used measures are cosine coefficient etc. Multiplication of the document and query vector and then sum the products. . Larger documents possess the query terms. Real is usually used with long documents.
Where document is d and q are query, similarity is represented by
$\text{Sim}(d, q) = \sum di \times qi$
Where di and qi are the corresponding document and query vectors.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

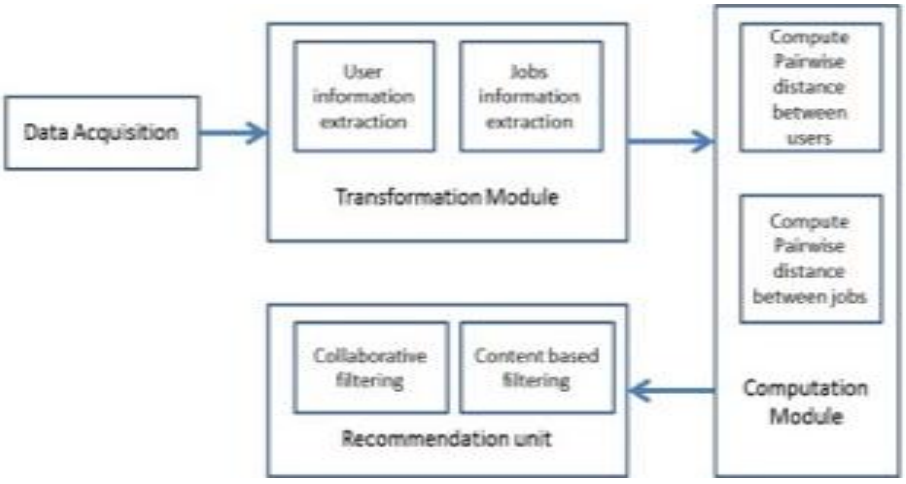

FIG 1.

1. **Cosine similarity -** measuring the similarity between two non-zero vectors in an inner product space that measures the cosine of the angles between them.

Similarity = (C.D) / (||C||. ||D||)

where C and D are vectors.

Numerator presents the inner product of the vectors and denominator indicates the lengths of the vector which lie in the range 0 to 1.

APPROACH

PREPROCESSING OF THE DATA: During pre-processing, the resumes are cleansed so that all special or junk characters that are present in the resumes can be filtered. In the pre-processing process numbers, special characters and all single letter words are filtered out. We will get the filtered data sets after the below steps are followed without any numbers or special characters.

The dataset is then broken into the tokens using Natural Language Toolkit.

- string fragments are masked for escape sequences for example \a, \b, \t.
- masking of numbers in the corpus
- Only letter words are replaced with an empty string.
- Masking of emails.

REMOVAL OF STOP WORDS FROM THE DATA: Removal of stop words like “and”,” the”,” was”, etc which frequently appear are not helpful for prediction process

- Steps to be followed for removal of stop words.

1. Input words are tokenized into individual tokens and are stored in an array
2. Each word is matched with already formed list of stop words present in the library.
3. The words that are present in the Stop Words list, they are then filtered from the main sentence array.
4. Till the last element of the array is not matched the process is continued.
5. Resultant array should not contain stop words.

- STEMMING OF DATA: Stemming is the process of removing the inflection by removing the unnecessary characters and reducing it to the root word. Root is the part of the word where the inflection is added which changes the word. Affixes like ed, s, de, Ing, mis. example: word likes singing, sings are mapped from sing
- LEMMATIZATION OF DATA: Lemmatization is different from stemming, in lemmatization inflected phases are decreased so that base word belongs correctly to the language.

Lemmatization comprises of the below steps:

•Change the text to list of words.

•Create all the item of word list as that of the sentence.

- DEPLOYMENT: In this process the resume will be matched by the job description and the percentage match will be calculated.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

IMPLEMENTATION

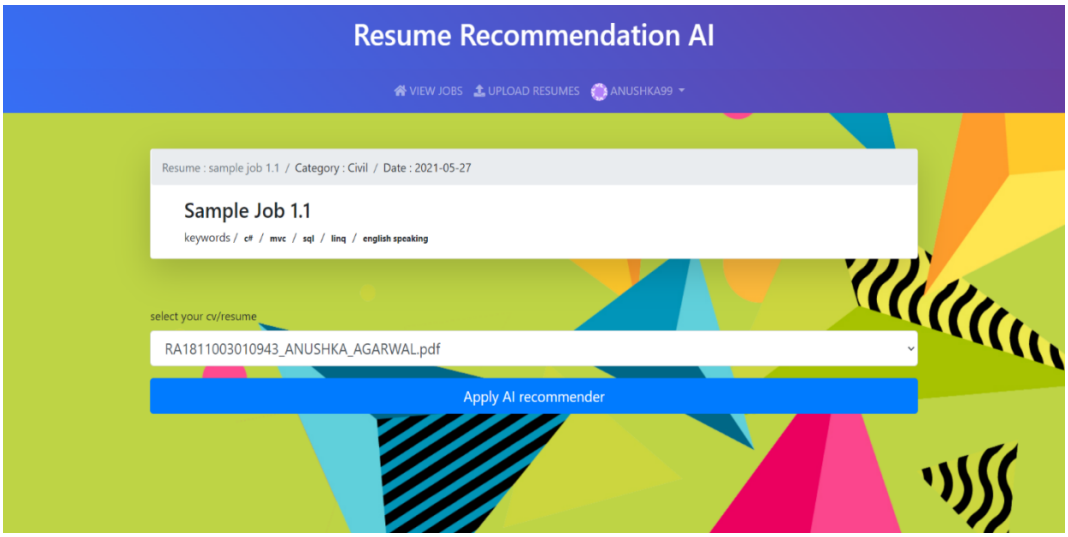

FIG 2.

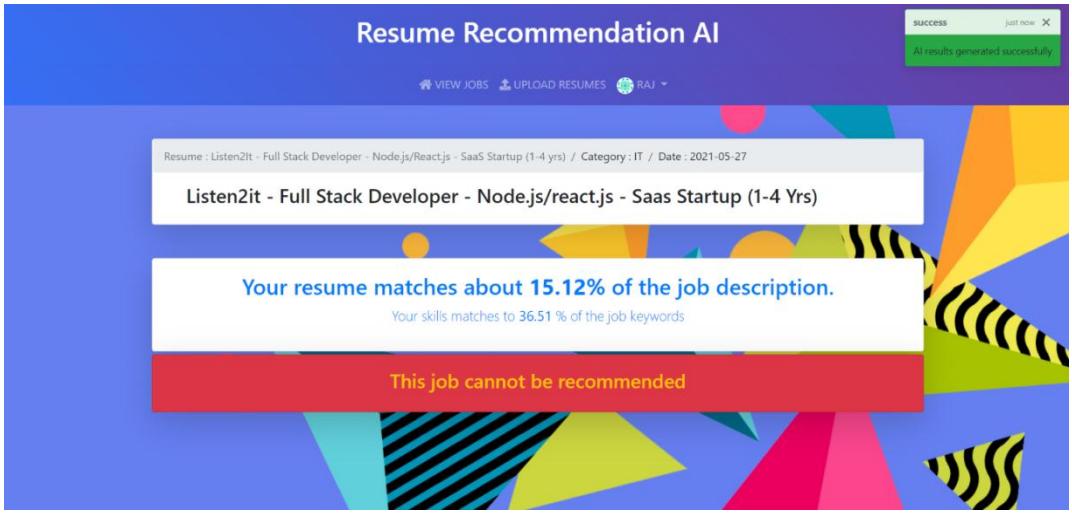

FIG 3.

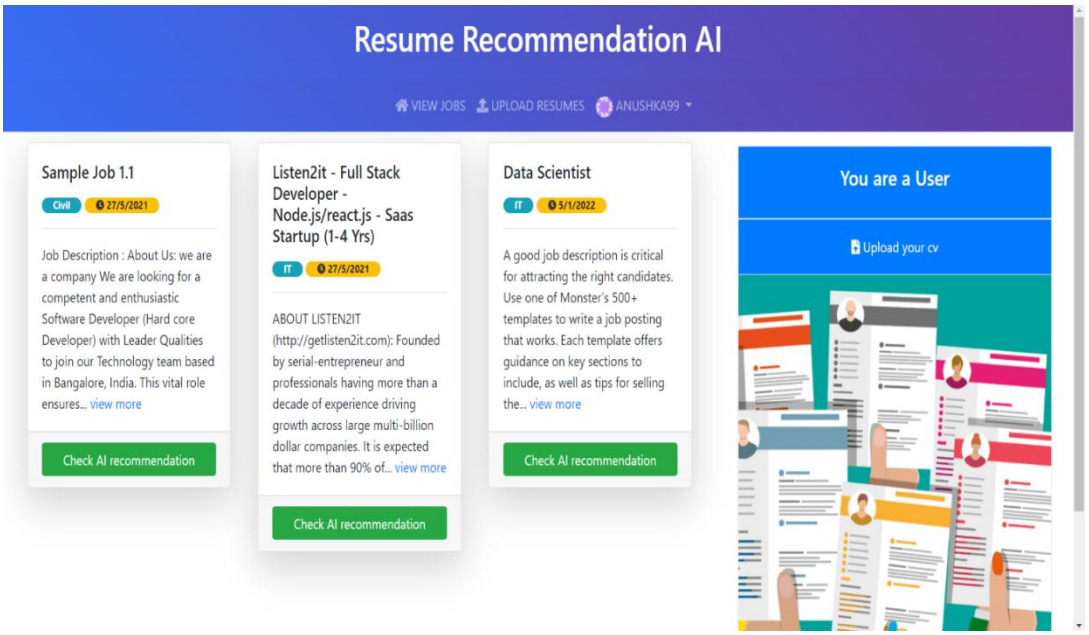

FIG 4.

III. CONCLUSION

The manual process of classification of candidate's resume is a tedious and time-consuming process and also has a scope of manual errors. To overcome this issue, an automatic machine learning model was proposed which helped in recommending the suitable candidates resume to the recruiter according to the verbal job description. Processing of the resume to extract information from the wide-ranging and unstructured formats of the resumes. A summarized version is then created of every resume which only has the keywords. The model works in phases, where in the first phase resumes of the applicants are categorized. In the second phase similarity index is calculated with respect to the job description. Cosine similarity algorithm is used to calculate the similarity between the job description and resume. Finally, the percentage of the similarity index is show to the recruiter. It displays weather or not the candidate is suitable for the work profile.

IV. REFERENCES

[1] Sebastiani F., "Machine Learning in Automated Text Categorization", ACM Computing Surveys, vol. 34 (1), 2002, pp. 1- 47.

[2] Al-Otaibi, S.T., Ykhlef, M.. "A survey of job recommender systems." International Journal of Physical Sciences 7,2012, pp.5127–5142.

[3] Mahesh wary, S., Misra, H., 2018. "Matching resumes to jobs via deep Siamese network," International World Wide Web Conferences Steering Committee. in: Companion Proceedings of the Web Conference 2018 pp. 87–88.

[4] Paparrizos, I., Cambazoglu, B.B., Gionis, A., "Machine learned job recommendation", fifth ACM Conference on Recommender Systems, ACM. 2011pp. 325–328.

[5] Yi, X., Allan, J., Croft, W.B. "Matching resumes and jobs based on relevance models," 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM.2017 pp. 809–810

[6] D. Mladenic, "Text-learning and Related Intelligent Agents: A Survey," IEEE Intelligent Systems, vol. 14, no. 4, 1999, pp. 44–54.

[7] Pradeep Kumar Roya, Sarabjeet Singh Chowdhary, Rocky Bhatia," A Machine Learning approach for automation of Resume Recommendation system", International Conference on Computational Intelligence and Data Science, 2019, pp.2239-2326.

[8] Shaha T. Al-Otaibi., and Mourad Ykhlef., "Job Recommendation System for Enhancing E-recruitment Process", Proceedings of the International Conference ..., 2012, pp.1-6

[9] Beniamin G. BOŞCAI., "THE EVOLUTION OF E-RECRUITMENT: THE INTRODUCTION OF ONLINE RECRUITER ", European Social Fund through Sectorial Operational Program Human Resources development 2007–2013, pp.161-170

[10] LIONEL NGOUPEYOU TONDJI., "Web Recommender System for Job Seeking and Recruiting", African Institute for Mathematical Sciences (AIMS) Senegal, 31 January 2018, pp. 1- 44.