International Conference on Innovative Data Communication Technologies and Application (ICIDCA-2023)
IEEE Xplore Part Number: CFP23CR5-ART; ISBN: 979-8-3503-9720-8

Applicant Screening System using NLP

B. Lalitha[1]
Department of CSE
GokarajuRangaraju Institute of Engineering & Technology, Telangana, India.
lalli.c22@gmail.com

Sirisha Kadiyam[2]
Department of CSE
GokarajuRangaraju Institute of Engineering & Technology, Telangana, India.
sirishakadiyam2@gmail.com

Ritisha Varma Kalidindi[3]
Department of CSE
GokarajuRangaraju Institute of Engineering & Technology, Telangana, India.
ritishakalidindi@gmail.com

Sri Maukthika Vemparala[4]
Department of CSE
GokarajuRangaraju Institute of Engineering & Technology, Telangana, India.
maukthi.vemparala@gmail.com

Kshiraja Yarlagadda[5]
Department of CSE
Gokaraju Rangaraju Institute of Engineering & Technology, Telangana, India.
yarlakshiraja15@gmail.com

Sri Vinutna Chekuri[6]
Department of CSE
Gokaraju Rangaraju Institute of Engineering & Technology, Telangana, India.
vinutnachekuri@gmail.com

2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) | 979-8-3503-9720-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIDCA56705.2023.10099953

ABSTRACT- This research work is mainly focused on selecting or short-listing the eligible candidates from the pool of applicants in a relatively short span of time. As technology is rapidly progressing, manpower requirement also increases exponentially. Thus, to calibrate the cream of the crop, this online application screens the applicants resume for a specific recruitment ad. This is designed in such a way that the applicant as well as the hiring agency can both be benefited, i.e., the applicant can use it to avail the job opportunities, apply for it and improve their abilities if they don't meet the criteria. Hiring agencies can mention the details of the job openings available. This bilateral website allows applicant to upload their resume, the resume uploaded will be compared with the job occupation requirement posted by the hiring agencies by using Natural Language Processing [NLP]. Results are generated using cosine similarity, then the similarity of both the uploaded documents in percentage is displayed. The eligibility of the candidate is decided based on the displayed result. Recent techniques include CNN KNN algorithms which are complex and time consuming, this project uses NLP tools, which simplifies the process, reduces the time consumption and also gives accurate answers.

KEYWORDS - Natural Language Processing [NLP], Natural Language Tool Kit [NLTK], Cosine Similarity, Section-based Segmentation, spaCy.

I. INTRODUCTION

Around the world recruiting and hiring the best talent is a tedious task for any firm. The challenges a company faces to take in the suitable candidate for their work is a toilsome process, as the applicants applying for jobs are numerous.

When a company announces a job opening many aspirants send in their resumes to apply for the position. In this process the companies receive many resumes that need to be screened to find the best fit that meets all the necessary qualification for the job role. The main goal of any recruiter is to the select right applicant for their company, due to many incoming resumes it is difficult to go through them manually and consumes time to look into each of them separately to see if they are meeting the criteria of the company.

Over the years there have been many recruiting sites that have been developed to make the hiring process easier. In all these projects the job seeker can scan the resume and upload on the site. The approaches that have been used in these to develop these sites are mainly using machine learning algorithms such as the KNN (K-Nearest Neighbour) method to rank the most suitable resume, the LSTMS (Long Short-Term Memory) and the Linear SVM (Support Vector Machine), some of the methods also include different kinds of classification algorithms [1,2,3,7]. All these methods that have been mentioned have given accurate results the only disadvantage is that it can only be used for a limited data and not for large amount of data. Later, the researchers have adopted the NLP approach to address the issue of storing enormous amount of data [4,5,9,10]. From the literature, it is understood that there are many methods that have been introduced to skim through the resumes. In this view, the paper proposes a different approach to reduce the time consumption and match resumes to the job descriptions to get the accurate percentage.

II. METHODOLOGY

This section provides, architecture, algorithms used and modules used to build the application.

A. Existing method

The eligibility of the candidate was first done manually, reading everything in the resume and then selecting and shortlisting them. But over time there is an increase in the growth of industries and reading thousands of applications, emails, resumes manually is not possible, it is but time consumption is more, so to overcome this situation there are many proposed methodologies that help the agencies to take the candidates who are fit for a specific role which they are searching for.

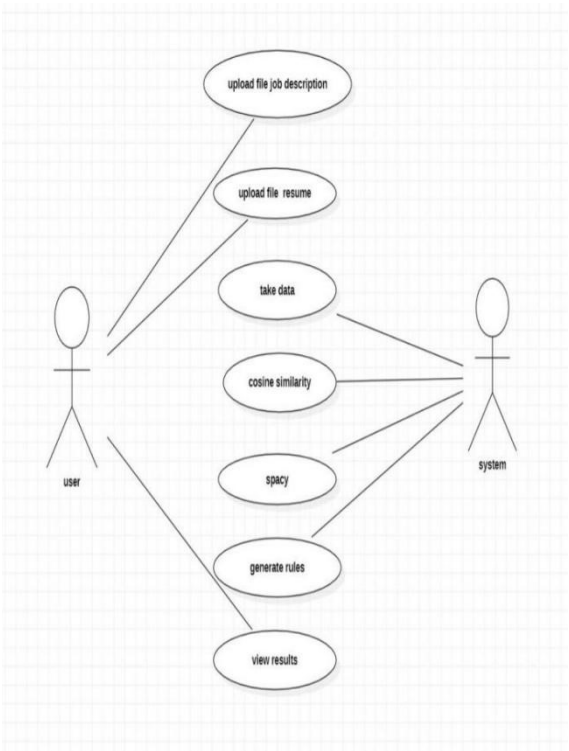
Day by day there are many things invented which are enhanced form of the traditional idea, they are improved to reduce time consumption. The proposed system relies on this fact.

B. Proposed method

In order to determine whether the skill set listed on the resume corresponds to the job description or not, this article addresses the development of a web application where the job description and resume are compared. The user interface will show the skills that have been matched. Natural language processing techniques like cosine similarity and other built-in Python libraries like NLTK and Spacy are used in the development of this application. To overcome the existing methodologies drawbacks, this approach uses NLP tools to enhance the process and increase the performance.

C. Overall working model

The user first uploads a file including their resume as well as a file containing the job description. Then, the Spacy procedure is carried out, treating words and sentences as objects. After deleting the stop words, the provided description and resume are normalised. The percentage of similarity is then calculated by performing a cosine similarity on the normalised job description and résumé. The acceptance or rejection of a given resume is determined using a percentage system.



D. Natural Language Processing Tools

1 .NLTK

A strong Python package for working with human language data is called NLTK (Natural Language Toolkit). For simpler tasks like tokenization, stemming, and tagging as well as more complex ones like parsing and semantic reasoning, it offers a wide variety of tools. Developers frequently use NLTK for tasks like text classification, sentiment analysis, and machine translation because it makes it simple to perform natural language processing (NLP) operations on enormous amounts of text data.The vast array of pre-trained models and data sets available through NLTK is one of its distinguishing qualities. This includes well-known data sources like the Penn Treebank and the Brown Corpus, as well as models that have already been pre-trained for tasks like part-of-speech tagging, named entity recognition, and other similar ones. Developers may use tools like Stanford CoreNLP and the Spacy library with ease because NLTK also offers interfaces to other well-known NLP libraries. NLTK is a useful tool for any developer dealing with human language data because to its robust set of tools and wide range of resources.

2. Tokenization

Tokenization is the process of separating individual tokens, or words, phrases, symbols, or other meaningful pieces, from a string of text. Natural language processing (NLP) applications including text categorization, sentiment analysis,

and machine translation all involve tokenization, which is a fundamental step in NLP. Tokenization can be done in a variety of ways, such as by separating words with white space and punctuation, looking for patterns in the text with regular expressions, or identifying the boundaries between tokens with machine learning algorithms. Tokenization produces a list of tokens that can be used as input for additional NLP activities.

3. Stemming

The act of stemming involves stripping a word down to its fundamental or root form. Stemming aims to reduce words to their most basic forms so that they can be quickly studied and contrasted. Natural language processing (NLP) tasks including text categorization, sentiment analysis, and information retrieval can benefit from this. The Porter stemmer, which employs a set of heuristic guidelines to eliminate frequent morphological affixes from a word, is the most often used stemming method. There are numerous distinct stemming algorithms that each have their own advantages and disadvantages. Snowball stemmer and Lancaster stemmer are some further examples.

4. Lemmatization

A term is lemmatized when it is stripped down to its dictionary definition. It is a more complex kind of stemming that reduces words to their most basic forms so that they can be quickly compared and examined. Lemmatization determines a word's base form using a dictionary or a morphological analyser, as opposed to stemming, which utilises heuristic criteria to eliminate frequent morphological affixes. To get at the base form, it considers the word's context and part of speech.

5. Cosine Similarity

A measure of similarity between two non-zero vectors in an inner product space is called cosine similarity. It is frequently used in information retrieval and natural language processing (NLP) to gauge how similar two texts or other text-based data are. The cosine of the angle between two vectors is used to calculate the cosine similarity. Cosine values vary from -1 to 1. The vectors are said to be identical if their cosine similarity is 1, completely opposite if it is -1, and orthogonal if it is 0. (not similar).Cosine similarity, as its name suggests, measures the cosine of the angle formed by two vectors in a high-dimensional space. To do this, divide the dot product of the vectors by the product of the magnitudes of the vectors. The resultant value is then applied to the two vectors to determine how similar they are. Because it is comparatively simple to compute and is not sensitive to the relative lengths of the text documents, cosine similarity is particularly well-liked for text data. Cosine similarity can be used to find related documents, assess how well a search query matches the documents, and for other similar purposes. Cosine similarity in this application can be achieved by text Similarity to determine how the two text documents are close to each other in terms of their context or meaning, here it is between the applicant's resume and the job description uploaded. In this approach Cosine similarity is one of the metrics to measure the text-similarity between two documents irrespective of their size in Natural language Processing.

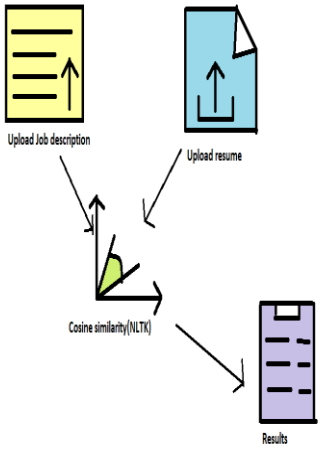
6. Spacy

Popular Python open-source module spaCy is used for sophisticated natural language processing (NLP). It is quick and efficient, with a focus on usability and readability, and it is especially built for production use. SpaCy is based on the most recent research and is frequently updated to enhance functionality and include new features.One of spaCy's important characteristics is its capacity to carry out numerous NLP operations, including dependency parsing, named entity recognition, part-of-speech tagging, tokenization, lemmatization, and named entity recognition, all inside a single library. Additionally, it offers pre-trained models for a variety of tongues, including English, German, Spanish, French, Portuguese, Italian, Dutch, Greek, and many more.

III .IMPLEMENTATION

The application can be explained in two divisions: front end and back end. Front end deals with the user interface which is developed using HTML (Hyper Text Markup Language) and CSS (Cascading Style Sheets) which is connected through flask library and in the backend the data is stored in SQL YOG enterprise. Once the user opens the application, the displayed page is the about page which consists the login button and the registration button. The new user should register to use the application to login . After signing up, the user goes to a new window and uploads two files, one of which is a resume and the other is a job description, these documents could be of any format like PDF(Portable Document Format), word document and text files.The application shows the degree of similarity between the user's resume and the job description, when the documents are uploaded. After the user signs in, on the back end the database is updated with their information. The documents are saved in the database once the user uploads them. The documents are translated to text format, and the text is then sent into a programme called Spacy, which breaks up the huge text into smaller units known as objects, such as sentences and words. The stop words are removed to normalise the phrases and words. Then the similarity between the assorted data is checked using cosine similarity,

which is an important tool to produce the eligibility criteria through percentage.



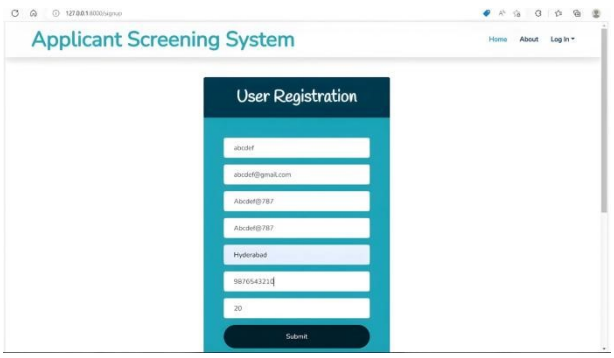
7. Data Acquisition and storage

In this application a database is created , this database contains all the information about the user. Once the user gets register onto the application the details of the user are stored into a database termed as sql yog enterprise . Each time the user logins into the site with their credentials the data fetches the required details and lets the user login.

IV.EXPERIMENTAL RESULTS

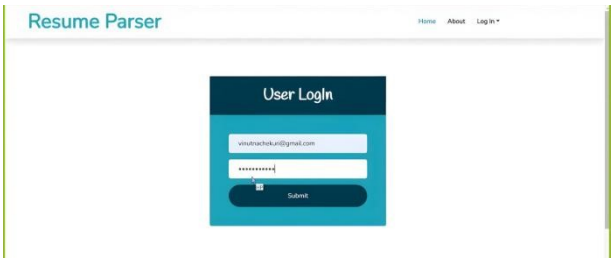
The new applicants should register in the application, and then login with the same credentials. Next step is to upload the job description as well as the resume of applicant. Using cosine similarity, the similarity percentage between the two files is given.

Home page - It contains the synopsis of our application.

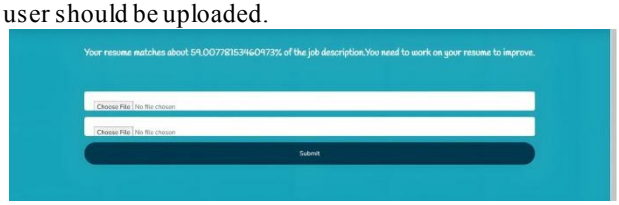


Registration page - It is a page in which the required details like name, mail, password, phone number should be filled.

Login page - The registered user should login to in the login page to upload the documents.



Upload page - Here the job description and the resume of the user should be uploaded.



Result page - The similarity percentage between the uploaded documents is displayed.

V. CONCLUSION AND FUTURE SCOPE

This application helps the job applicants by simplifying the process of job application. It gives the similarity percentage between the job description and the resume uploaded by the applicant. Cosine similarity is used to give the similarity percentage. Any percentage above 70 conveys that the applicant is eligible for job.

Future scope :

In future , this application can be upgraded such that the resumes and the job description can be matched exactly to a specific category like years of experience and the skills the candidate holds , the application can navigate the job seeker directly to the category they required.

VI. REFERENCES

1. Rajath V, Riza Tanaz Fareed, Sharadadevi Kaganurmath, 2021, Resume Classification and Ranking using KNN and Cosine Similarity, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 08 (August 2021),

2.J Chen, Z Niu, H Fu, "A Novel Knowledge Extraction Framework for Resumes Based on Text Classifier," Proceedings of the International Conference on Web-Age Information Management. Springer International Publishing, pp. 540-543, 2015.

3.T Schmitt, P Caillou, M Sebag, "Matching Jobs and Resumes: a Deep Collaborative Filtering Task," Proc. of the 2nd Global Conf. on Artificial Intelligence, pp.1-14, 2016.

4. S. Bharadwaj, R. Varun, P. S. Aditya, M. Nikhil and G. C. Babu, "Resume Screening using NLP and LSTM," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 238-241, doi: 10.1109/ICICT54344.2022.9850889.

5. T. M. Harsha, G. S. Moukthika, D. S. Sai, M. N. R. Pravallika, S. Anamalamudi and M. Enduri, "Automated Resume Screener using Natural Language Processing(NLP)," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1772-1777, doi: 10.1109/ICOEI53556.2022.9777194.

6.F Javed, Q Luo, M McNair, F Jacob, M. Zhao, and TS. Kang, "Carotene: A Job Title Classification System for the Online Recruitment Domain," Proceedings of the IEEE First International Conference on Big Data Computing Service and Applications (BigDataService), pp. 286 - 293, 2015.

7. M. F. Mridha, R. Basri, M. M. Monowar and M. A. Hamid, "A Machine Learning Approach for Screening Individual's Job Profile Using Convolutional Neural Network," 2021 International Conference on Science & Contemporary Technologies (ICSCT), Dhaka, Bangladesh, 2021, pp. 1-6, doi: 10.1109/ICSCT53883.2021.9642652.

8.A. Zaroor, M. Maree, and M. Sabha, "A Hybrid Approach to Conceptual Classification and Ranking of Resumes and Their Corresponding Job Posts," In: Czarnowski I., Howlett R., Jain L. (eds) Intelligent Decision Technologies 2017. IDT 2017. Smart Innovation, Systems and Technologies, vol 72. Springer, Cham.

9.S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika and A. Gurjar, "Web Application for Screening Resume," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-7, doi: 10.1109/ICNTE44896.2019.8945869.

10. Sinha, A.K., Amir Khusru Akhtar, M., Kumar, A. (2021). Resume Screening Using Natural Language Processing and Machine Learning: A Systematic Review. In: Swain, D., Pattnaik, P.K., Athawale, T. (eds) Machine Learning and Information Processing. Advances in Intelligent Systems and Computing, vol 1311. Springer, Singapore. https://doi.org/10.1007/978-981-33-4859-2_21.

11. Daryani, Chirag, Gurneet Singh Chhabra, Harsh Patel, Indrajeet Kaur Chhabra, and Ruchi Patel. "An automated resume screening system using natural language processing and similarity." ETHICS AND INFORMATION TECHNOLOGY [Internet]. VOLKSON PRESS (2020): 99-103.

12. E Faliagka, L Iliadis, I Karydis, M Rigou, S Sioutas, A Tsakalidis, and G Tzimas, " On-line consistent ranking on e-recruitment: seeking the truth behind a well-formed CV," The Artificial Intelligence Review, 42(3), 515, 2014.

13. R Kessler, N Béchet, JM Torres-Moreno, M Roche and M. El-Bèze, "Job Offer Management: How Improve the Ranking of Candidates", in Foundations of Intelligent Systems, J. Rauch, et al., Editors. Springer Berlin Heidelberg, pp. 431-441, 2009.

14. K Yu, G Guan, and M Zhou, "Resume information extraction with cascaded hybrid model." Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, pp. 499–506, 2005.

15. S Clyde, J Zhang, and CC Yao, "An object-oriented implementation of an adaptive classification of job openings," Proceedings of the 11th Conference on Artificial Intelligence for Applications, IEEE, pp. 9-16, 1995.