

RESUME SCREENING USING NATURAL LANGUAGE PROCESSING

Anakha K M

Computer Science and Engineering
Adi Shankara Institute of Engineering and Technology,
Kalady
Ernakulam, India
anakhakm2002@gmail.com

Haritha Rajan

Computer Science and Engineering
Adi Shankara Institute of Engineering and Technology,
Kalady
Ernakulam, India
haritharajan981@gmail.com

Honey M

Computer Science and Engineering
Adi Shankara Institute of Engineering and Technology,
Kalady
Ernakulam, India
0127honeym@gmail.com

Sumesh C Raman

Computer Science and Engineering
Adi Shankara Institute of Engineering and Technology,
Kalady
Ernakulam, India
sumeshr.cs@adishankara.ac.in

Abstract---The demand for low-cost labour and the growth in available jobs have both contributed to a significant increase in the Indian recruitment industry during the past five years. The hiring process is outsourced to organisations whose main goal is to provide the specific talents that the organisation requires in the recruitment industry, which is a new method of employing people. As the need for jobs rises, so does this industry. This is undertaken since big businesses hire in large quantities, and carrying out such a task internally would use up a lot of corporate resources and reduce production. Since these businesses are growing and even they find it tedious and time-consuming to manually review every candidate's resume, these talent management firms use various ML models to screen out the best resumes in accordance with the job roles, which lessens the workload on the human resources department.

Keywords—*Natural Language Processing, Resume Analyzer, e-recruitment, Job seeker*

I INTRODUCTION

Screening the resumes of all job applicants is the first task for any recruiter in the recruiting process. Every day, thousands of emails from hopeful process applicants flood the inboxes of each company that has a position advertised for a certain role. Additionally, this procedure includes the involvement of numerous human resources and necessitates intense efforts and resources to select the best candidate for future hiring. At earlier stages of the hiring process, recruiters may be able to identify non-relevant profiles, which could result in considerable time and cost savings. In the realm of machine learning, we train a programme with a dataset to forecast the desired result while using fresh data. The majority of resume screening is done through the use of natural language processing (NLP), which relates to the way that people speak to one another. NLP works to provide computer systems with the ability to distinguish spoken and written words in a manner similar to how individuals

can. NLP integrates computational linguistics-totally rule-based modelling of languages with a combination of statistical and systemic language learning and in-depth fashion mastery. Bringing those techniques together makes it possible for computers to process the human language functions in the context of texts, in order to fully "apprehend" them. Natural Language Processing divides the content material referred to in the resume into small grammatical parts this is referred to as parsing. This lets the system extract the relevant information from the resume and create a summary of it irrespective of the order and format of the resume. After growing the summary, the device lists out the candidates by mapping the similarity between the content material referred to in their resume and inside the process description. So, this is how the NLP is utilised in resume screening. There's no trendy format for resumes so inconsistencies and beside the point statistics are inevitable. The objective of statistics extraction is to extract relevant information or keywords from the unstructured records inside the resume with none human intervention. via information Extraction strategies we have obtained the summary of each carried out candidate. Then the gadget needs to list out the candidates to offer pointers of the right applicants to the recruiters.

II TECHNOLOGY USED

A. Text matching

Text matching deploys and seeks out in raw data, there are specific text matches. Text Matching plays a significant role in applications in the actual world and a significant process in some cases. Furthermore, in this certain dynamic approaches are introduced to aid with the creation of patterns by words. The methods include text file matching, text mining, text clustering, association rule extraction, word cloud, natural language processing, and text similarity measures. String-based methods are the most common type of text mining, utilized in a variety of situations.

B. Vectorization

A geometric structure made up of a number of "vectors" is known as a vector space. They are referred to as scalars because they can be multiplied and added by particular integers, or "scaled." It is an algebraic model that is used to describe text data for text mining, information retrieval, and natural language processing. The term "vectorization" refers to the representation of texts in the vector space paradigm. This is the way a document is transformed into a numeric vector. The fact that most machine learning models require numeric vectors rather than strings as inputs is one of the primary reasons for vectorization. Each potential word is mapped to a particular integer as a standard method of vectorizing text. Each word will be put in a distinct slot if your array is big enough. This index's value represents how frequently the word appears. The matrix size is usually smaller than the corpus vocabulary. Therefore, to account for this, we need a vectorization approach.

C. Natural language processing

Natural language processing (NLP), which relates to the way humans interact with one another, is primarily used to screen the resumes. The goal of NLP is to enable computers to comprehend spoken and written language in a manner similar to that of humans.

The below Figure shows the architecture of NLP (figure 1). The system consists of a database; as our data is appropriately organised in columns and rows rather than being unstructured, the database will be a SQL database. The model will be trained using the available information we obtained from Kaggle's open platform. This approach employs two models, the first of which is either K-Nearest Neighbour or Support Vector Machine. The system will then filter the resumes depending on the specified domain using machine learning algorithms and natural language processing techniques that will extract useful information from the resume.

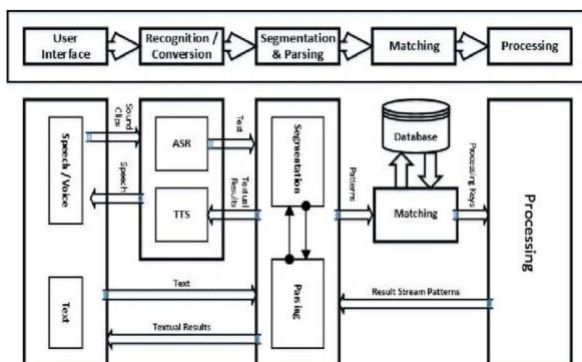


Fig. 1. Architecture of NLP

III RELATED WORKS

Pradeep Kumar Roy et al. [1] proposed a machine learning strategy for automating a resume recommendation system. Finding the best applicants from the list of resumes is the aim of this work. They have created a machine learning-based approach to accomplish this. The two key steps of the suggested model's operation were preparation and deployment and inference.

The main drawbacks of the model is it accepts CVs in CSV format; however, in the actual world, a resume must either be PDF, doc, etc. Because of dataset restrictions, There was no way to enhance the model to take as input, doc or pdf files.

Rajath et al. [2] suggested a Resume Classification and Ranking using KNN and Cosine Similarity. The system receives the candidate's resume during this phase. The words are then eliminated after processing the resume using an NLP pipeline. To receive the right set of words, strategies like stop words and lemmatization are applied. The resume is classified into various categories using the TF-IDF vectorizer, which vectorizes the words for the KNN model. Since it is crucial to compare the CV to the provided JD document, a cosine similarity technique is employed to relate the JD's content to the candidate's CV.

Gopalakrishna and Vijayaraghavan V [3] suggested an Automated tool for Resume classification using Semantic analysis. This study offers a detailed account of how they classify resumes using semantic analysis. The HR team reviews accepted resumes using the NLP Pipeline (NLPP), which employs Stop deletion to get rid of terms like "and, or, etc." Additionally, methods like NER and parts-of-speech tagging are employed. The development and implementation of a resume classifier programme that uses a voting classifier based on ensemble learning to assign a candidate's profile to a suitable domain based on the interests, career history, and competence they have listed in their profile.

Harsh Patel et. al [4] implemented an automated resume screening system using natural language processing and similarity. In this study, they introduced an automated resume screening method that streamlines the e-recruitment process by getting rid of the inconveniences of recruiters who relied on hand-selecting candidates for the available job position experienced. The two phases of this system are information extraction and content-based candidate recommendation. Without requiring human browsing, information extraction extracts pertinent keywords from unstructured text data in a resume. utilising methods such as Stemming, POS Tagging, Named Entity Recognition, and Tokenization, etc. This mechanism uses concepts like Vectorization, weighting techniques like TF-IDF and similarity measures like cosine distance to calculate the similarity between document contents.

Haibo Tan et. al [5] proposed an Enhanced text matching based on semantic transformation. The advanced text matching technique is developed on a Transformer network coupled with CycleGAN. The semantic distance between pairs of texts is decreased as the semantics of text in the source domain are transferred to similar or different target domains. It demonstrates the process of identifying paraphrases and matching answers to questions.

Tejaswini K , Umadevi V , Shashank M Kadiwal , Sanjay Revanna [6] suggested a Design and Development of Machine Learning based Resume Ranking System. Resumes are submitted by candidates after an MCQ test with a facial

recognition system that detects cheating. NLP techniques are used to extract the pertinent skills from a resume after it has been submitted, and TF-IDF vectorization is employed to convert the words into vectors that computers can understand. The recruiter's JD is precisely matched by the resume classifier, which in this case uses the KNN algorithm to detect resumes.

Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M.Kiruthika, Ambarish Gurjar [7] implemented a Web application for screening resumes. This system develops a web application for resume screening. On the applicant's end, the applicant submits a resume; the server side processes the resume; and finally, using the NLP framework SpaCy, the pipeline is trained. A point calculator is used to rank resumes on the Recruiter's page, enabling recruiters to choose the best candidate for the job.

Jagan Mohan Reddy D, Sirisha Regella, Srinivasa Reddy Seelam [8] proposed a system for Recruitment prediction using machine learning. The system uses predictive employee analytics to support human resource management based on a variety of metrics that optimise business outcomes. Also constrained decision rules for forecasting to improve industry needs. We propose a machine learning approach, building a support model to categorise the staff recruitment process, including decision trees, naive Bayes, support vector machines, and random forests.

Irfan Ali, Nimra Mughal, Zahid Hussain Khand, Javed Ahmed, Ghulam Mujtaba [9] implemented a resume classification system using natural language processing and machine learning techniques. In this paper a resume categorization system employs machine learning (ML) and natural language processing (NLP) methods. It categorises resumes by job category to ensure performance. The overall methodology is divided into five phases of data acquisition and visualisation, preprocessing, feature engineering, model building, model validation, real-time graphical user interface testing and evaluation (GUI) environment.

The Resume Screening System Using LSTM was proposed by Navale Sakshi, Doke Samiksha, Mule Divya, Prof. Said S. K [10]. LSTM-based screening of resumes receives input from a PDF file. Data is taken out within the file and saving of the extracted data is done in the CSV format. The data in the csv file is sanitised and forecasts are created using the model LSTM. LSTMs are trained and saved for use in web applications. The dataset is cleaned and vectorized. A vectorized dataset is used for training. Weights are retained for future use and a bidirectional LSTM model is trained.

IV CONCLUSION

The automated resume screening system that we describe in this paper optimises the e-recruitment process by removing the many issues that recruiters encountered when they relied on manually shortlisting applicants for a certain position. Our system works on two different levels. The process begins by using NLP to extract relevant information from resumes' disorganised and different formats. It generates a condensed version of each resume that only contains the details required for the hiring decision. It produces a streamlined version of each resume that only includes the information necessary for the selection process. The removal of any superfluous material has made the verifier's job easier and increased the effectiveness of the resume analysis process. Our system, on the other hand, uses similarities to match extracted resume features to job description requirements.

V REFERENCES

- [1] Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, 167, 2318-2327.
- [2] Rajath, V., Fareed, R. T., & Kaganurm, S. (2021). Resume Classification and Ranking using KNN and Cosine Similarity. *IJERTV10IS080057*, 10(08).
- [3] Gopalakrishna, S. T., & Vijayaraghavan, V. (2019). Automated tool for Resume classification using Semantic analysis. *International Journal of Artificial Intelligence and Applications (IJAIA)*, 10(1).
- [4] Daryani, C., Chhabra, G. S., Patel, H., Chhabra, I. K., & Patel, R. (2020). An automated resume screening system using natural language processing and similarity. *ETHICS AND INFORMATION TECHNOLOGY [Internet]*. VOLKSON PRESS, 99-103.
- [5] Zhang, S., Tan, H., Chen, L., & Lv, B. (2020). Enhanced text matching based on semantic transformation. *IEEE Access*, 8, 30897-30904.
- [6] Tejaswini, K., Umadevi, V., Kadiwal, S. M., & Revanna, S. (2021). Design and Development of Machine Learning based Resume Ranking System. *Global Transitions Proceedings*.
- [7] Amin, S., Jayakar, N., Sunny, S., Babu, P., Kiruthika, M., & Gurjar, A. (2019, January). Web application for screening resume. In *2019 International Conference on Nascent Technologies in Engineering (ICNTE)* (pp. 1-7). IEEE.
- [8] Reddy, D. J. M., Regella, S., & Seelam, S. R. (2020, October). Recruitment prediction using machine learning. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-4). IEEE.
- [9] Ali, I., Mughal, N., Khand, Z. H., Ahmed, J., & Mujtaba, G. (2022). Resume classification system using natural language processing and machine learning techniques. *Mehran University Research Journal Of Engineering & Technology*, 41(1), 65-79.