

Resume Analysis Using Machine Learning and Natural Language Processing

Sakshi, Sanjana T R, Vanyashree J S, Ganavi H S

Department of CSE, JNN College of Engineering, Shivamogga, Karnataka, India.

Sushma R B

Department of CSE, JNN College of Engineering,, Shivamogga, Karnataka, India.

Abstract - The recent recruitment process has changed from manually to computerized. The majority of resume analysis is carried out utilizing Natural Language Processing and Machine learning. It deals with selecting the best candidates from an enormous pool of applicants who have the necessary skills for a certain job description. By leveraging these NLP capabilities provided by NLTK, resume analysis systems can automate the extraction, classification, and evaluation of information from resumes, enabling recruiters to efficiently screen and evaluate candidates based on their qualifications, skills, experience, and other relevant factors.

Keywords - Cosine Similarity, Machine Learning, TDF-IDF, NLP-Natural Language Processing.

I. INTRODUCTION

In the recent technologically growing world, the corporate recruitment process is evolving to a great extent. Physical copies are no longer used for submitting the resumes of the candidates and the recruitment teams want an e-resume that can be viewed online. In the present system the candidate has to fill in each and every piece of information regarding their resume in a manual form which takes a large amount of time and then also the candidates, are not satisfied with the job which the present system prefers according to their skills. The problem you mentioned about the dissatisfaction of candidates with their job placements is also a common issue in the recruitment process. This can be due to a lack of proper matching between the job requirements and the candidate's skills and interests. In this scenario, an intelligent recruitment system that uses machine learning algorithms can be very useful. Such a system can analyse the candidate's resume, their skills, work experience, and interests and match them with the job requirements and company culture. This can ensure that the candidate is

placed in a job that aligns with their skills and interests, leading to job satisfaction and better retention rates for the company. Machine learning, a pivotal domain within artificial intelligence, represents a paradigm shift in how computers acquire knowledge and perform tasks. Natural language processing (NLP) is a subfield of Artificial Intelligence (AI). This is a widely used technology for personal assistants that are used in various business fields

II. LITERATURE SURVEY

In [1] Alkeshwar Jivtode et al. introduces a comprehensive approach for resume analysis using machine learning and NLP, aiming to automate parsing, extraction, and ranking of resumes. The methodology employs NLP techniques like morphological, syntactic, and semantic analysis to break down unstructured resume data into meaningful components. The Resume Ranking Algorithm, utilizing models like SVM or KNN, categorizes resumes based on their relevance to job profiles. The system emphasizes candidate suggestions and feedback to enhance the recruitment process, achieving high parsing and ML model accuracy, along with an 80% end-to-end resume analysis accuracy.

In [2] Tejaswini K et al. present a machine learning-based resume ranking system using KNN, focusing on matching resumes to job descriptions. The model employs text pre-processing, including stop word removal and stemming, constructing a graph for sentence similarity. Two recommendation approaches are used: cosine similarity matching against job descriptions and KNN for item feature similarity. The system achieves an 85% parsing accuracy and a 92% accuracy in ranking resumes based on similarity.

In [3] Pradeep Kumar Roy et al. propose a machine learning-based resume recommendation system with a two-step approach: preparation and deployment. The dataset, sourced from online portals and Kaggle, undergoes preprocessing, including cleansing, tokenization, and feature extraction using TF-IDF. Two models are built for classification and recommendation based on resume and job category. Accuracy scores for classification models like Random Forest, Multinomial Naive Bayes, Logistic Regression, and Linear SVM range from 38.99% to 78.53%.

In [4] Mungi Naga Venkata Sai et al. present a resume screening model using machine learning, extracting information like abilities and credentials. Data preprocessing involves cleaning resumes by removing special characters and digits. Two models are employed: the first, using KNN, SVM, MLP, and LR, predicts the most suitable job role for a resume. The second, utilizing cosine similarity, recommends roles based on user input. Accuracy scores for KNN, LR, SVM, and MLP are impressive, ranging from 97.40% to 99.48%.

In [5] Chirag Daryani et al. present an Automated Resume Screening System with NLP-based information extraction and a content-based recommendation system using techniques like TF-IDF and cosine similarity. The system summarizes resumes in JSON format and recommends the most suitable candidates for a given job description.

In [6] Nimish Patil et al. present a resume parser and analyzer using NLP, leveraging Python libraries such as NLTK and Spacy for text extraction and processing. The system performs comprehensive parsing steps, including entity extraction, skills identification, and experience analysis, storing the results in a database. The responsive web application offers recruiters visualized insights, resume scores, and skill matching suggestions, contributing to an efficient Automatic Tracking System (ATS) for applicant selection.

In [7] Y. Sowjanya et al. propose a smart resume analyzer using Mong, Cosine Similarity, and Overlapping Coefficient for ranking resumes based on job descriptions. The system employs datacleansing, TF-IDF for term frequency analysis, and

KNN to find closely matching CVs. It utilizes cosine similarity for further comparison, recommends courses based on skills, and utilizes a Kaggle dataset for Java Developer and Project Manager roles. Resumes are converted to PDF format for efficient data handling.

In [8] Riya Pal et al. introduce a resume classification model using machine learning algorithms, gathering datasets from websites like Kaggle, Glassdoor, and Indeed. The methodology involves data cleanup by removing blank spaces, changing text to lowercase, and eliminating stop words. Tokenization, lemmatization, and TF-IDF vectorization are employed, and classification algorithms such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest achieve varying accuracies of 45%, 60%, and 70%, respectively.

In [9] Irfan Ali et al. present a comprehensive resume classification system using NLP and machine learning in five stages: data collection, preprocessing, feature engineering, model construction, and evaluation. They collected a dataset with job categories, performed data cleansing, and employed feature representation. Multiple models like LSVC, SGD, LR, NuSVC, KNN, GNB, MNB, and BNB achieved varying accuracy scores, with LSVC and SGD reaching 99.6%, LR at 99.3%, and others ranging from 97.2% to 79.2%

In [10] Pratham Rasal et al. utilize Support Vector Machine (SVM) and Random Forest Algorithm for resume parser analysis. Their methodology involves data exploration, pre-processing, SVM training, prediction, and evaluation using metrics. Additionally, for Random Forest, they select random data points, build decision trees, and choose the number of trees for an accurate ensemble, considering the trade-off with computational cost.

In [11] V. G. Jagtap's model uses NLP for resume screening, employing tokenization, stemming, and POS tagging for information extraction. The system formats the extracted details into JSON, utilizing techniques like TF-IDF for term significance assessment and cosine similarity for document matching in a multidimensional space. Affinity propagation aids in forming clusters based on

similarities between data points.

In [12] Vansh Nawander's Resume Analyzer integrates Streamlit, Resumeparser, Pandas, pdfminer3, Matplotlib, and PySQL for efficient resume processing. Utilizing Streamlit for web app development, Resumeparser extracts data from resumes, Pandas manages data, Matplotlib aids visual assessment, and PySQL ensures smooth database access. The user-friendly application caters to both regular users and administrators, providing efficient resume analysis with secure data storage in a local MySQL database.

In [13] B.Kelkar's Resume Analyzer focuses on e-recruiting and data mining, categorizing prior research into keyword recovery, DOM tree structure, and semantic-based object extraction. The proposed model involves resume collection, keyword searching, knowledge base enhancement, and candidate ranking using machine learning. Prioritizing proficiency levels in programming languages, it streamlines the recruitment process by automating resume analysis and shortlisting based on specific criteria.

In [14] Patlolla Sruthi's Smart Resume Analyzer, based on RNN-based Keyword Extraction, facilitates resume submission and editing, employing pyresparser for efficient data extraction using NLP. Extracted details are stored, and the system analyzes skills, predicts job roles, rates resumes, suggests improvements, and offers interview tips. The primary goal is automated resume screening, enhancing accuracy through machine learning, and providing a user-friendly experience with feedback and application suggestions.

In [15] Arvind Kumar Sinha's Resume Screening model employs five machine learning algorithms, including Random Forest, Multinomial Naive Bayes, Logistic Regression, Linear Support Vector Classifier, and XG Boost, achieving a notable accuracy of 93.5%. Leveraging Python libraries like NLTK, SpaCy, Scikit-learn, FuzzyWuzzy, and Django, along with pattern matching techniques like Levenshtein Distance and Cosine Similarity, enhances data analysis, visualization, and job description matching in this comprehensive AI and NLP-based solution.

Table 1: Summarization of various Authors

Authors	Title	Research Focus	Remarks
Alkeshwar Jivtode et al.[1],2023	Resume analysis using machine learning and natural language processing.	NLP techniques ,SVM orKNN.	SVM or KNN for ranking, achieving 80% accuracy, and emphasizing candidate feedback for enhanced recruitment.
Tejaswini K et al.[2], 2022	Design and development of machine learning based resume ranking system.	Text pre-processing ,KNN , cosine similarity.	The system achieving an 85% parsing accuracy and 92% accuracy in ranking based on similarity, with a focus on job description matching.
Pradeep Kumar Roy et al.[3] ,2020	A Machine Learning approach for automation of Resume Recommendation system.	Dataset preparation, Preprocessing, and classification for effective job matching.	Accuracy scores suggest potential for further refinement and improvement in model performance.

Mungi Naga Venkata Sai et al.[4],2022	Resume Screening using Machine learning	Model focused on extracting relevant information and recommending suitable job roles through a combination of predictive modeling and cosine similarity.	Accuracy scores for the employed models, suggesting high potential for practical application in automating the recruitment process.
Chirag Daryani et al. [5],2020	An automated resume screening using natural language processing and machine learning	NLP techniques for information extraction and content-based recommendation for efficient candidate matching.	integrating NLP techniques for information extraction and content-based recommendation, potentially streamlining the resume screening process with minimal human intervention.
Nimish Pati et al.[6],2021	Resume parser and analyzer using nlp	Web-based resume parsing and analysis system using NLP, employing Python, NLTK, Spacy, and regular expressions for efficient data extraction, scoring, and visualization	NLP and web technology, enhancing resume parsing accuracy, providing valuable insights for recruiters, and facilitating an effective Automatic Tracking System for streamlined decision-making in the recruitment process.
Y.Sowjanyaal.[7],2023	Smart resume analyzer.	Cosine Similarity, and Overlapping Coefficient, integrating Natural Language Processing	NLP and ML techniques for efficient resume analysis, ranking, and personalized skill recommendations, demonstrating versatility in handling diverse data formats
Riya Pal et al.[8],2022	Resume Classification using various Machine Learning Algorithms	emphasizing data gathering from diverse sources, data cleanup, and the application of classification algorithms such as Naïve Bayes, Support Vector Machine, and Random Forest for improved job profile suggestions.	A systematic approach to resume classification, employing various machine learning algorithms, with Random Forest yielding a noteworthy 70% accuracy .
Irfan Ali et al.[9],2021	Resume Classification System using Natural Language Processing and Machine Learning Techniques	resume classification system integrating natural language processing and machine learning techniques, employing various classifiers such as LSVC, SGD, LR, NuSVC, KNN	Resume classification framework utilizing a diverse set of classifiers, with LSVC and SGD demonstrating remarkable accuracy of 99.6%, showcasing the potential for accurate and efficient job category predictions in real-time environments.

Pratham Rasal et al.[10],2023	Resume Parser Analysis Using Machine Learning and Natural Language Processing	A resume parser analysis model integrating Support Vector Machine and Random Forest algorithms	Resume parsing and analysis, incorporating Support Vector Machine and Random Forest algorithms.
V. G. Jagtapi et al.[11],2023	Resume Screening and Course Recommendation.	NLP techniques, featuring multi-stage information extraction, tokenization, stemming, POS tagging, and TF-IDF analysis for enhanced understanding and processing of unstructured resumes.	NLP techniques and algorithms, showcasing a meticulous approach to resume screening and course recommendation, leveraging tokenization, and cosine similarity for effective information extraction and analysis.
Vansh Nawander et al.[12],2022	Modern Resume Analyser For Students And Organisations	Python modules, emphasizing Streamlit for web application development, Resumeparser for information extraction, and a client-server model with MySQL for secure data storage	Integration of Python modules and web technologies, offering a user-friendly and efficient Resume Analyzer, particularly beneficial for recruiters dealing with a large volume of resumes.
B.Kelkar et al .[13],2020	Resume Analyzer Using Text Processing	A Resume Analyzer utilizing text processing, emphasizing machine learning algorithms for keyword-based resume analysis and categorization to streamline the recruitment process.	Automated resume analysis, emphasizing programming language proficiency to streamline recruitment.
Patlolla Sruthi et al .[14], 2023	Smart Resume Analyser using RNN - based Keyword Extraction	Automated resume screening, data extraction, skill analysis, and personalized recommendations to enhance the candidate experience.	Automated screening, skill analysis, and personalized feedback, emphasizing an improved candidate experience and application enhancement.
Arvind Kumar Sinha et al .[15]	Resume Screening Classification using Artificial Intelligence and Natural Language Processing	Model employing AI and NLP, utilizing a combination of machine learning algorithms, pattern matching techniques, and open-source libraries for enhanced efficiency in the resume screening process.	machine learning algorithms, pattern matching techniques, and open-source libraries to create an efficient Resume Screening Classification model, achieving high accuracy.

III. CONCLUSION

This paper outlined a survey of resume analysis using machine learning and natural language processing that improves the e-recruitment process by reducing the many issues that recruiters experienced when they utilized manual shortlisting of candidates for a specific job post. Our system functions on two levels. First, it employs Natural Language Processing to collect necessary details from the resumes' unstructured formats. It produces a reduced version of each resume

that only contains the information necessary for the selection process. The duty of the screening officials is made simpler by the removal of all the unnecessary content, allowing them to access each resume more effectively. On the other hand, our technology offers the option of evaluating the applicants by comparing the retrieved resume features to the requirements in the job description using cosine similarity

REFERENCES

- [1] Abdul Wahab, Dr. M N. Nachappa. "Resume Analysis using Machine Learning and Natural Language Processing." International Research Journal of Engineering and Technology (IRJET) Volume:05/Issue:05/May-2023
- [2] Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna. "Design and Development of Machine Learning based Resume Ranking System." Global Transitions Proceedings, 2021.
- [3] Pradeep Kumar Roy , Sarabjeet Singh Chowdhary , Rocky Bhatia , "A Machine Learning approach for automation of Resume Recommendation system", Procedia Comp. Sci. 167 (2020)
- [4] Mungi naga Venkata sai Raghavendra."Resume Screening and machine learning", Journal of Engineering Sciences Volume:13/Issue 09,2022.
- [5] Chirag Daryani , Gurneet Singh Chhabra , Harsh Patel , Indrajeet Kaur Chhabra , Ruchi Patel , An automated resume screening system using natural language processing and similarity, Intellig. Comput. Indust. Des. (ICID) (2020)
- [5] Nimish Patil, Shubham Yadav, Vikas Biradar. "Resume Parser and Analyzer using nlp" , International Research Journal of Engineering and Technology (IRJET), Volume:05/Issue:04/April-2023.
- [6] Ms.Y.Sowjanya ,Mareddy Keerthana ,Pulluri Suneeksha ,Dorgipati Sai Sri Harsha." Smart Resume Analyser". International Journal of Research in Engineering and Science (IJRES). Volume:11 /Issue: 3 /March 2023
- [7] Riya Pal, Shahrukh Shaikh, Swaraj Satpute and Sumedha Bhagwat, "Resume Classification using various Machine Learning Algorithm". ITM Web of Conferences. (2022)
- [8] Irfan Ali, Nimra Mughal, Zahid Hussain Khand, Javed Ahmed, Ghulam Mujtaba." Resume Classification using Natural Language Processing and Machine Learning Techniques", Mehran University Research Journal of Engineering and Technology, Vol. 41, January 2022,
- [10] Pratham Rasal, Yashashri Balwaik, Mahesh Rayate, Rohit Shinde, Prof. A. S. Shinde." Resume Parser Analysis Using Machine Learning and Natural Language Processing", International Journal for Research in Applied Science & Engineering Technology (IJRASET) , Volume 11 Issue V May 2023.
- [11] V. G. Jagtap, Payal Bhosale, Sanyogita Londhe,Pooja Meher, Pratiksha Raut." Resume Screening and Course Recommendation System", International Research Journal of Modernization in Engineering Technology and Science. Volume:05/Issue:05/May-2023
- [12] Vansh Nawander, Shrenita Elma, Andhoju Karthikeya, Manuka Koushik Yadav, Sai Karthik Kotala, M. D. N. Akash."Modern Resume Analyser For Students And Organisations".International Research Journal of Engineering and Technology (IRJET).Volume: 09 Issue: 10 |Oct 2022
- [13] B.Kelkar, R.Shedbale, D.Khade, P.Pol, A.Damame."Resume Analyzer Using Text Processing". Journal of engineering Science.Vol 11, Issue 5,May/ 2020
- [14] Patlolla Sruthi P. N. V. K. G. Adithya, MD. Suleman, Pallerla Kunal. " Smart Resume Analyser: A Case Study using RNN-based Keyword Extraction ".(2023)
- [15] Arvind Kumar Sinha, Md. Amir Khusru Akhtar , Mohit Kumar, Shrikant Upadhyay. "Resume Screening Classification using Artificial Intelligence and Natural Language Processing". Research paper,Issue-2023