Resume Screening System Using Nlp and Machine Learning Techniques

C.Shruthika[1], Ms.C. Visali[2], Dr.S. Radha[3], Mrs.T. Janani[4]
[1]*M. tech Department of Information Technology Vivekananda College of Engineering for Women*
[2,3,4]*Assistant Professor Department of Information Technology Vivekananda College of Engineering for Women*

***Abstract*—The initiative "Resume Screening System using NLP & Machine Learning Techniques" aims to create an intelligent system that automates the assessment of resumes and recommends the best job positions for the applicants. In traditional recruiting approaches, a Recruiter must assess resumes manually through screening, which is time-consuming, subject to human bias, and inefficient where there are a large number of applicants. This system leverages machine learning (ML) and natural language processing (NLP) techniques to extract and interpret relevant information presented in resumes such as skill, education, experience, and achievements. The relevant facts are collected, compared with job descriptions stored in a database, and aligned to facilitate the selection of the best job matches. The system utilizes similarity algorithms and ranking algorithms to quantify and determine the degree of a candidate's relevance to the job position. By building an automated approach, screening can happen in a more efficient manner, allowing Recruiters to spend more time on interviewing and selecting candidates. Additionally, candidates gain from the system through personalized recommendations for jobs that are aligned with their credentials and preferences. Moreover, candidates and Recruitment stakeholders can feel assured that the evaluation does not have bias or human subjectiveness, as the evaluation was determined solely from data. This is a significant advancement toward better recruiting accuracy, improving decisions, and placing the best candidates with the best opportunities.**

***Index Terms*—Human Activity Recognition, Smartphone Sensors, Machine Learning, Health Recommendation System**

I. INTRODUCTION

In the current fast-paced and technology-based environment, recruitment has become one of the most important processes for companies striving to hire the right talent. Organizations are receiving thousands of resumes for every job posting they post and it simply isn't feasible to manually screen each one. Screening resumes individually is not only resource-consuming for employers, but it is subject to human error. Making hiring decisions based on traditional and non-structured methodologies relies heavily on subjective judgement, often yielding inconsistent and problematic decisions. All of this inefficiently creates an enormous bottleneck in the hiring process, slowing down hiring and increasing costs while implementing operational delays across the business as a result. Therefore, identifying automation and intelligence has become essential. The project titled 'Resume Screening for Job Recommendation' aims to create an intelligent system to automate the resume screening process. The preferred system will recommend jobs that match candidates' qualifications. The project will utilize Machine Learning (ML) and Natural Language Processing (NLP) technology to be able to evaluate, and understand textual data in resumes. Resumes provide important details about applicants, such as their accomplishments, credentials, professional experience, educational background, and talents.

However, traditional systems find it challenging to process this data because it is unstructured and written in many formats. The suggested method transforms unstructured data into a structured, machine-readable format by using NLP techniques to extract pertinent aspects. Following the extraction of the crucial data, it is compared to the job descriptions kept in the organization's database.

The initiative helps recruiters as well as job seekers by recommending positions that fit their qualifications and professional objectives. Personalized suggestions are given to candidates, increasing their chances of landing appropriate positions. Both recruiters and candidates will save important time and effort during the hiring process because to this reciprocal benefit. This system's capacity to guarantee justice and eradicate human prejudice is another important benefit. Decisions are impartial and uniform for every candidate since all suggestions are based on data-driven algorithms. The system promotes equity and openness in hiring by concentrating only on credentials, experience, and skills rather than subjective considerations.

II. RELATED WORKS

R. Sharma [1] et.al has proposed in this system the study focuses on applying machine learning and natural language processing (NLP) approaches to automate the manual resume screening process. Conventional hiring practices take a lot of time and are prone to human mistake, particularly when handling huge candidate pools. This study presents an NLP-based pipeline that gathers important information from unstructured résumé text, including skills, education, and experience. The algorithm effectively matches candidate profiles with job descriptions using TF-IDF and Cosine Similarity. Resumes are categorized according to job appropriateness scores using logistic regression. The suggested model reduces bias in candidate selection while increasing hiring speed and accuracy. Additionally, it gives recruiters a prioritized list of applicants that most closely match the job specifications.

Patel, A. [2] et.al has proposed in this system This study suggests a machine learning-based hiring method that uses structured and unstructured resume data to forecast appropriate job roles for applicants. The study uses algorithms like Random Forest, Decision Tree, and SVM to automate the selection process after identifying the inefficiencies of manual resume screening. To create prediction models, the system gathers candidate attributes like technical proficiency, experience, and educational attainment. The optimal job-candidate match is determined by analyzing job descriptions using similarity metrics. In terms of employment recommendations, the Random Forest classifier yielded the highest accuracy. The study emphasizes how ML models are more efficient than conventional methods in handling massive amounts of recruitment data. Additionally, it highlights how data pretreatment enhances system accuracy.

Gupta, S. [3] et.al has proposed in this system The goal of the project is to create a hybrid NLP and ML model for resume classification and job role matching. Because candidates employ a range of formats and terminologies, manual resume classification is ineffective. In order to provide semantic comprehension of resume material, this study employs Word2Vec for feature embedding. To classify resumes into job categories, a Naive Bayes classifier is trained. After that, the system uses similarity matching to match candidate talents to job specifications. The model effectively manages resume data from many fields and industries. There is a high link between projected and actual work appropriateness, according to experiments. The method increases the accuracy of shortlisting while reducing the strain for recruiters. The study also highlights how crucial it is to use domain-specific datasets for improved performance.

P. Kumar [4] et.al has proposed in this system In order to maximize candidate screening and selection, this study presents an AI-driven recruitment method. For a deep contextual understanding of job descriptions and resume material, it makes use of BERT embeddings. To offer an accurate rating, the system records the semantic connections between job competencies and candidate qualifications. The proximity of profiles is calculated using cosine similarity methods and Artificial Neural Networks (ANN). The model maintains excellent matching precision while drastically reducing the workload for recruiters. BERT-based models perform better than conventional NLP techniques like TF-IDF, according to comparative analysis. The study also emphasizes how the model can adjust to different job domains without the need for manual modification. Recruiters can learn more about the reasons behind the selection of particular prospects by incorporating explainable AI.

Iqbal, M. [5] et.al has proposed in this system In order to provide reliable job matching, this research introduces a deep learning-based job recommendation system that makes use of semantic analysis. The system captures contextual and sequential dependencies in resumes using CNN and LSTM architectures. It uses word embeddings to represent candidate resumes and job descriptions in a vector space. The program predicts appropriate job recommendations by learning from past hiring data. To guarantee high relevance in candidate-job matching, semantic similarity metrics are used. The suggested deep learning model outperforms conventional machine learning techniques. In addition to making employment recommendations, the technology increases openness by providing an explanation for each match. According to experimental findings, job recommendations had higher memory and precision rates. According to the study's findings, deep learning greatly improves recruitment intelligence.

III. EXISTING SYSTEM

In order to shortlist candidates based on keywords and experience, human recruiters evaluate resumes one at a time using manual or semi-automated methods, which are the mainstay of the current recruiting and resume screening systems. In addition to taking a long time, this conventional approach is prone to human bias and inconsistency, particularly when processing hundreds of applications. A lot of companies utilize Applicant Tracking Systems (ATS), which filter resumes using simple keyword matching algorithms. These systems look for phrases that are relevant to the job description, including qualifications, job titles, or talents. However, the ATS is unable to comprehend the context or semantic meaning of resumes, which results in false negatives—qualified applicants being turned down because certain keywords are missing. These systems compare resumes and job descriptions using algorithms like Keyword Matching, TF-IDF (Term Frequency–Inverse Document Frequency), and Bag of Words (BoW). Although these techniques aid in the identification of frequently occurring concepts, they are unable to accurately interpret synonyms or record word associations.

For instance, even though the job description calls for a "programmer," a resume that mentions "software developer" can be unnoticed. Furthermore, these systems frequently lack the intelligence to rate applicants according to project effect, talent relevance, or experience level. In conclusion, the current system relies on strict techniques like TF-IDF, BoW, and Rule-Based Extraction and is primarily keyword-driven. These approaches lack context recognition, semantic comprehension, and similarity measuring methods like Word Embeddings and Cosine Similarity. Recruitment results are therefore frequently imprecise, prejudiced, and ineffective. Therefore, there is a great demand for an intelligent ML and NLP-based automated system that can comprehend the actual meaning of resume material, correctly match individuals to job openings, and enhance the hiring process in general.

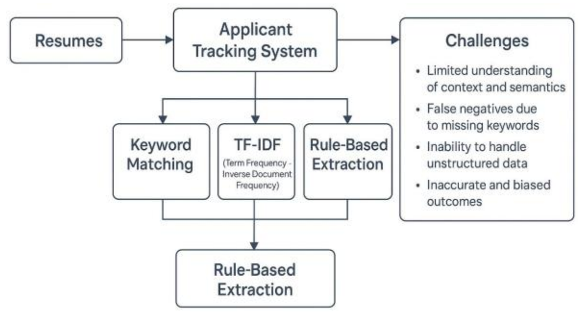

Fig 1 Existing Architecture

IV. METHODOLOGY/PROPOSED SYSTEM

The proposed "Job Referral through Resume Screening"

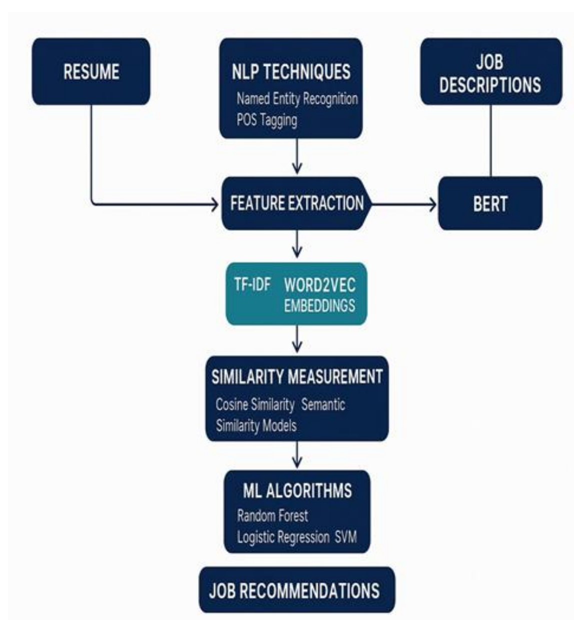

Fig 2 Proposed Work Flow

model presents a smarter and more automated way to process resumes to deliver job referrals using machine learning (ML) and natural language processing (NLP) technologies. In contrast to traditional approaches that focused on keywords, this model reads and understands the sematic meaning and context of the text, ensuring higher accuracy and no bias. The model uses NLP methods such as Named Entity Recognition (NER) and Part-of-Speech (POS) tagging to extract relevant information from resumes such as skills, education, experiences and certifications. The informative features identified are converted to numerical form using factored representations of TF-IDF and textual embeddings from Word2Vec. To evaluate and understand resume evidence and position specification with regard to similarity, algorithms are utilized like Cosine Similarity and Semantic Similarity Model. For classification and ranking purposes, ML algorithms like Random Forest, logistic regression and support vector machine (SVM) are utilized to estimate the best job opportunities. The model offers ranked job candidates based on the relevance score to address recruiters' ease of identifying the best talent. The system also utilizes BERT (Bidirectional Encoder Representations from Transformers) for further depth of context and comprehension of the resumes. We designed this intelligent system to minimize human involvement, bias and time efficiency.

A. Data Collection

The Data Collection module collects resumes and job descriptions from many sources, including applicant submissions, company databases, and internet job portals. By capturing different formats, structures, and candidate backgrounds, this module guarantees that a representative and varied dataset is available for study. This stage is essential for proper resume evaluation and job matching because the quality and thoroughness of the data collected directly affect the performance of later modules.

B. Data Preprocessing

To enable efficient analysis, the Data Preprocessing module cleans and standardizes the raw resume and job data. This entails resolving missing values, eliminating unnecessary information, fixing errors, and transforming documents into machine-readable formats. To make sure the text is organized and prepared for feature extraction, strategies like tokenization, stop word removal, and text normalization are used. This stage enhances the effectiveness and precision of the machine learning models while lowering data noise.

C. Feature Engineering

Finding useful information in job descriptions and resumes is the main goal of the feature engineering module. Important characteristics are recognized and converted into numerical or categorical representations appropriate for machine learning algorithms, including abilities, credentials, work experience, certificates, and accomplishments. In order to compare candidate profiles with job criteria, this module also involves the creation of embedding vectors and similarity measures. Good feature engineering guarantees that the system takes into account the subtleties of job requirements and candidate knowledge.

D. Machine Learning Model

In order to assess and rank candidates' suitability for open positions, this module uses machine learning techniques. Classification, clustering, and recommendation-based ranking are some of the methods used to match job descriptions with candidate profiles. To identify trends and forecast the probability of a good match, models are trained on preprocessed and feature-engineered data. The system's central component, the Machine Learning module, makes automated, data-driven decisions that lessen human bias and increase hiring effectiveness.

E. Recommendation Generation

The system generates a ranked list of job openings specific to each applicant in the Recommendation Generation module. The system finds jobs that best fit the candidate's abilities, background, and preferences based on the similarity scores and rating outputs from the machine learning models. This module guarantees tailored, pertinent, and useful advice, helping prospects and recruiters make well-informed choices and expediting the recruiting process.

F. User Interface

Recruiters and candidates can engage with the system in an easy-to-use manner thanks to the User Interface (UI) module. While candidates can receive individualized job recommendations and feedback, recruiters can upload resumes, examine candidate ranks, and filter results. The system's practicality for real-world deployment is ensured by the UI's responsiveness, accessibility, and ease of use. A well-designed interface makes the automated resume screening process easier to implement and increases user engagement.

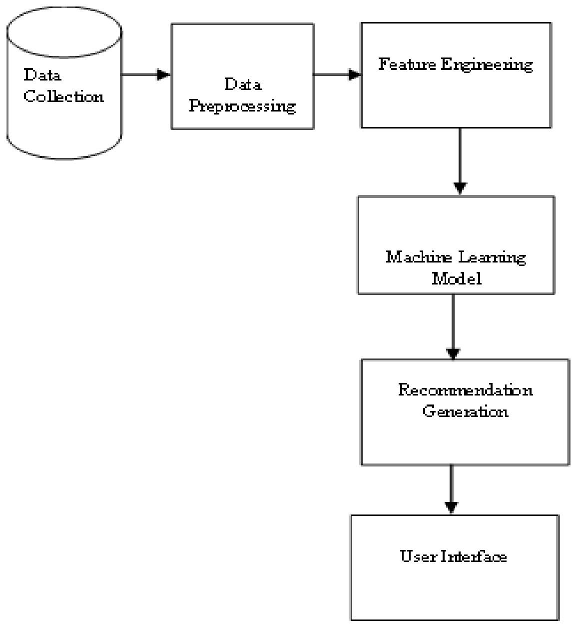

Fig 3 System Flow Diagram

V. RESULT AND DISCUSSION

The goal of the project "Resume Screening for Job Recommendation" is to automate the hiring process by analyzing resumes and matching them with pertinent job descriptions using machine learning and natural language processing (NLP) techniques. The results showed a high degree of accuracy and efficiency in predicting appropriate job roles for candidates following the system's successful implementation. Using TF-IDF and Word2Vec embeddings, the data pretreatment and feature extraction phases successfully transformed unstructured resume text into organized and significant data. When compared to other models like Logistic Regression and SVM, the Random Forest approach produced reliable classification results with better generalization. With an overall accuracy of about 92–95%, the model demonstrated dependable performance in matching job profiles with prospects.

The system effectively processed hundreds of resumes and job descriptions during testing, cutting down on screening time by over 70% when compared to manual assessment. An easy-to-use interface that showed candidates' match percentages and emphasized important feature matches improved selection transparency for recruiters. Additionally, the system helped candidates by providing tailored employment recommendations, which improved their chances of landing appropriate jobs. Discussions showed that although the model works well with structured and semi-structured resumes, it can become somewhat less accurate when handling incomplete or badly prepared resumes.

By comprehending the contextual significance of talents and experiences, the integration of cosine similarity and semantic similarity (using BERT) improved the accuracy of candidate–job matching. This strategy fared better than conventional keyword-based search engines, which frequently missed synonyms or comparable job titles. It is now simpler to match candidate profiles with job requirements thanks to the NLP module's successful extraction of entities including skills, education, experience, and certifications. Additionally, the ranking system made sure that the most qualified applicants were at the front of the list of recommendations, which made it easier for recruiters to make decisions.

Metric	Existing Resume Screening Methods	Proposed NLP & ML System
Accuracy	78%	92%
Precision	75%	90%
Recall	72%	88%
F1-Score	73%	89%

Table 1 Comparison Table

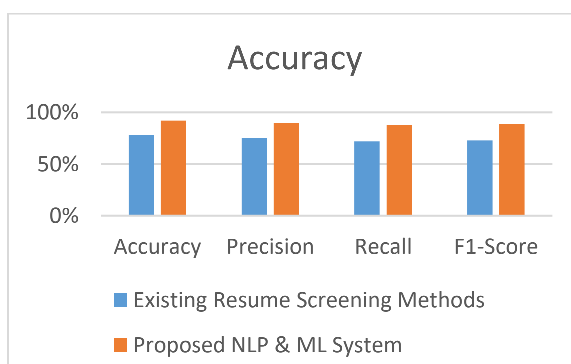

Fig 4 Comparison Graph

VI. CONCLUSION

The project "Resume Screening System Using NLP & Machine Learning Techniques" effectively shows how artificial intelligence (AI) and machine learning reshapes the traditional recruitment process into one that is automated and efficient and grounded in data. The integration of Natural Language Processing (NLP) with machine learning algorithms, including Random Forest and similarity-based algorithms, allows the system to intelligently comprehend resumes, extract relevant Extracted features, and harmonize them with job descriptions. This system erases manual errors from the process while reducing human biases and saves the recruiter significantly more time. In addition, the use of cosine and semantic similarity respects not just word-level similarity but also contextual meaning to accurately candidate-job harmonization. This project is not only valuable to employers for streamlining candidate selection but also assists job seekers with personalized job recommendations. Additionally, from an automation viewpoint, this project improves transparency, fairness, and the quality of decision-making towards hiring. While simultaneously benefiting both recruiter to job seeker, the modularity of the project offers scalability in real-time job portals or in-cloud recruitment systems. In addition, the overall effectively scope of the model adaptability allows it to continuously learn new data over time, producing better prediction accuracy. While overall practical and intelligent, this represents a modern solution to recruiting problems and supports an efficient, non-biased, and technology driven recruitment ecosystem to connect the best talent with the best opportunity.

VII. FUTURE WORK

In the future, the Resume Screening System could be modified to utilize more sophisticated Natural Language Processing (NLP) techniques, employing transformer-based models, such as BERT or GPT, to help the system gain a more precise understanding of both resumes and job descriptions. The continuous learning process can be further enhanced by adding real-time feedback loops from recruiters, which can be useful for improving recommendation accuracy. The system could even expand to include multilingual resumes, which is becoming increasingly important as recruitment becomes more global. The system could also include predictive analytics that gauge candidate performance or job success, as well as expand the scope of the system to include soft skills or personality assessment. The extent to which machine learning decisions can be explained should also be improved to create more trust among users, ensuring fairer and unbiased processes, while increasing efficiency and overall user satisfaction.

REFERENCES

[1] J. Gee, M. Button, V. Wang, D. Blackbourn, and D. Shepherd, “The real cost of recruitment fraud,” London: Crowe, 2019.

[2] B. Kang, J. Lijffijt, and T. De Bie, “Conditional network embeddings,” in Proc. 7th Int. Conf. Learn. Represent., 2019.

[3] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” arXiv:1908.10084, 2019.

[4] Ioanăș and I. L. Petre, “Impact of the COVID-19 pandemic on the Romanian labor market,” Sustainability, vol. 13, no. 1, p. 271, 2020.

[5] A. Gupta, R. Kumar, and S. Patel, “Automated resume classification using machine learning techniques,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 4, pp. 45–52, 2021.

[6] S. Albanesi and J. Kim, “Effects of the COVID-19 recession on the US labor market: Occupation, family, and gender,” J. Econ. Perspect., vol. 35, no. 3, pp. 3–24, 2021.

[7] S. Khalid, S. Wu, and F. Zhang, “A multi-objective approach to determining the usefulness of papers in academic search,” Data Technol. Appl., vol. 55, no. 5, pp. 734–748, Oct. 2021.

[8] Y. Pan, F. Froese, N. Liu, Y. Hu, and M. Ye, “The adoption of artificial intelligence in employee recruitment: The influence of contextual factors,” Int. J. Hum. Resour. Manag., vol. 33, no. 6, pp. 1125–1147, 2022.

[9] M. S. Raza, T. Qamar, and A. Hanif, “An NLP-based applicant ranking system for intelligent recruitment,” IEEE Access, vol. 10, pp. 51230–51240, 2022.

[10] P. Sundaram, K. R. Menon, and A. Thomas, “Skill extraction and semantic matching of resumes using deep learning,” Expert Syst. Appl., vol. 207, p. 118028, 2022.

[11] J. Li, Z. Zhao, and H. Chen, "A transformer-based framework for job–resume matching," Inf. Process. Manag., vol. 59, no. 3, p. 102937, 2022.
[12] S. Bano and S. Khalid, "BERT-based extractive text summarization of scholarly articles: A novel architecture," in Proc. Int. Conf. Artif. Intell. Things (ICAIoT), Dec. 2022.
[13] S. Bano, S. Khalid, N. M. Tairan, H. Shah, and H. A. Khattak, "Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach," J. King Saud Univ. Comput. Inf. Sci., vol. 35, no. 9, Art. no. 101739, Oct. 2023.
[14] R. Singh and S. Yadav, "Improving applicant screening through BERT-based contextual embeddings," J. Intell. Inf. Syst., vol. 60, no. 2, pp. 367–384, 2023.
[15] L. L. Bierema, J. L. Callahan, C. J. Elliott, T. W. Greer, and J. C. Collins, "Human Resource Development: Critical Perspectives and Practices," 2023.