Job Portal With Ai-Based Resume Matching

Siddhi Kumari[1], Sakshi Priya[2], A. Boobalan[3]
[1,2]*Department of Computer Science & Engineering, Galgotias University Greater Noida, India*
[3]*Galgotias University Greater Noida, India*

***Abstract*—The increase in number of job applications and the subjectivity of recruitment process inherent in manual hiring are tremendous drawbacks to the modern hiring systems which leads to inefficiency, unconscious discrimination, and poor candidate job fit. The research question is what is the effectiveness of an AI based job portal to automatize and optimize the process of screening candidates, parsing resumes and matching jobs, as well increasing the level of fairness and effectiveness in recruiting process. The suggested system, Smart Hire AI, uses the advanced techniques of Machine Learning (ML) and Natural Language Processing (NLP) to calculate the semantic similarity between resume and job description, extract the skills that are relevant, and prioritize the candidates, according to predetermine minimum skill-match threshold of 60%. The architecture to be used in the design of the system is a modular and scalable one, where MongoDB is used to provide a flexible and large volume data storage, Fast API to perform efficient processing and models integration on the backend, and React.js to provide an interactive and friendly user experience. The matching model based on the resume and similarity is designed using the vector-based representation of text and proximity points to guarantee objective and fact-driven recruiting suggestions.**

***Index Terms*—Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Resume Parsing, Automated Candidate Screening, Job Matching System.**

I. INTRODUCTION

The process of hiring in organizations in the recent years has experienced a permanent change and has become more data-driven, time-sensitive and technology-dependent. In the fast-growing digital recruitment environment, it is common to get hundreds or thousands of applications on one job advertisement. Such large numbers of applications can prove to be a significant burden on recruiters because it is time-consuming and difficult to evaluate such large amounts of data manually and without any sort of consistency. Due to this, the recruitment processes tend to be long-lasting, costly in operation and the best qualified candidates might be missed accidentally. Unstructured data on recruitments is a major cause of this inefficiency. Free-text contents prevail in resumes, cover letters, and job descriptions and it is hard to systematically analyze and compare qualifications of the candidates. Older Applicant Tracking Systems (ATS) are commonly used in the industry, but to a big extent these systems are highly reliant on filtering processes and key-word matching methods. Such systems do not understand semantic relationship, contextual relevance, and tacit skill relationship between candidate profile and the job requirements. As a result, shortlisting based on keywords may result in poor ranking of the candidates, the elimination of qualified candidates because of the absence of the right vocabulary, and the promotion of unconscious discrimination in hiring. The insufficiency of the conventional ATS systems are the reasons of being more intelligent and dynamic There are need for recruitment solution. To solve these problems, Artificial intelligence (AI) in particular through the use of Machine Learning techniques (ML) and Natural Language Processing (NLP) is a technology which has become one of the possible solutions. Natural Language Processing (NLP) is the ability to train the systems to draw meaningful information from unstructured text and by an understanding of the meaning of the words in the context and not rely on just matching keywords. Through Machine Learning models, recruitment system can learn from past data and improve accuracy in matching candidates over time and also help in taking decision based on data and reduce the need of human bias. The proposed system Smart Hire AI was able to automate the important tasks in recruitment

process such as parsing of resumes and candidate-job match. With the help of Natural Language Processing, it can extract the structured information like skills, experience, and qualification and analyze the suitability of a candidate based on a similarity score. To ensure that only qualified candidates go to the next stage minimum 60. Compared to traditional ATS Systems Smart Hire AI is less about strict rules of the principles of association of words and more about the context of understanding which means: A more efficient, scalable and fair hiring process.

II. LITERATURE REVIEW

Artificial Intelligence (AI) recruitment and the talent acquisition is the field that has been researched in many scholarly and industrial studies throughout the last twenty years. Early recruitment that was automated systems were mainly based on the use of keyword search engines as well as Boolean logic to sift through list of applicants on pre-established qualifications like job titles, skills and academic qualifications [1].

Although they provided basic automation and simplified the number of people needed to work on these, those systems could find explicit key words only. These rule-based systems did not find semantic relations, contextual relevance and transfer of skills and this tended to rule out good candidates based upon differences in terminology, resumes formatting, which are small or language expression. With the development of Natural Language Processing (NLP), scholars began to work on these deficiencies by applying method of linguistic analysis to unstructured recruitment data. Techniques like Named Entity Recognition (NER), Part-of-Speech (POS) tagging, and dependency parsing allowed extracting structured data in the form of skills and organizations and job roles and experience years, aided by the resumes and job description [2].

These methods made a great contribution towards the understanding of the resume by converting the free-text documents to the machine-readable format. NLP was however not good enough to match candidates and job without including predictive learning models. Machine Learning (ML) algorithms were implemented to recruitment pipelines to boost the categorizing and decision-making processes. Such models as Support Vector Machines (SVM), Naive Bayes classifiers, and Random Forests were popular in the tasks of identification of skills, identification of job roles, ranking of candidates, and prediction of experience. These methods proved to be more accurate than rule-based systems, learning trends on previous hiring data and changing with the emerging recruitment needs. However, most ML-based systems were susceptible to manually designed features and were not resistant to various resume formats and industries. The more recent advances in AI-assisted recruitment brought about sophisticated methods of candidate evaluation that are not based on resume. Hire-Vue and similar commercial platforms were the first to apply AI-powered video interviewing, where facial expressions, accent and word use are analyzed to assess the suitability of candidates [3]. Likewise, recruiting web sites such as the Indeed Hiring Platform, LinkedIn Talent Insights, use massive ML-based recommendation systems which use behavioral data, career history, and professional networks to calculate candidate-job similarity scores and hiring hints [4]. These systems represented a big advancement to comprehensive candidate evaluation and predictive hiring analytics. Irrespective of such developments, current AI-based recruitment solutions have been met with increasing criticism due to their interpretability, openness and impartiality. A large number of commercial solutions are black-box models and do not provide much information on the hiring decisions process and how bias is identified and overcome. The issue of Their usage, or implementation, facilitates algorithmic bias, data privacy and ethical responsibility have also raised the need for interpretable and fair recruiting technologies, more specifically in contexts of high stakes hiring settings. Based on these previous works and the industry, Smart Hire AI attempts to address the shortcomings of the current recruitment systems with the focus on the transparency, scalability, and contextual accuracy. The suggested design incorporates real-time pipelines of NLP and scalability of web-related technologies, such as FastAPI as a backend service and React.js as a frontend interface. Contrary to the opaque deep-learning-based systems, Smart Hire AI uses a transparent cosine similarity-based matching method to determine the fit between resumes and job descriptions in terms of semantic alignment. The system is able to provide the same shortlisting criteria with inclusivity since it has a

minimum 60skill-match requirement. Such a method is a significant improvement over the historical ATS and black-box AI recruiting systems as it strikes a balance between effectiveness in computation, interpretability, and equity. The Smart Hire AI promotes the creation of responsible AI in hiring by designing its product in a modular fashion and by explaining its matching logic, thereby creating a base towards ethical, data-driven hiring.

III. METHODOLOGY

To enhance the system, accuracy, scalability, and reliability, the proposed AI-based resume matching system was developed and deployed in a structural and systematic approach. The approach consists of a combination of Natural Language Processing (NLP), Machine Learning (ML), and up-to-date web technologies to automatize the process of resume screening and job matching. The general process should start with data collection and preprocessing, then continuing with feature extraction and computing similarities and finally, the final stage should be visualization of the result and evaluation of the system graphic

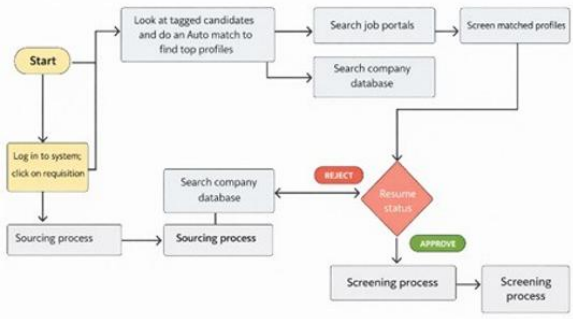

Fig. 1. Workflow of AI-based Resume–Job Matching System

The system takes in unstructured data on resume and job description which is processed to extract the relevant textual and semantic features. These characteristics are then changed into numerical values to allow similarity to be calculated between the job requirements and the candidate profiles. The web-based interface is helpful to recruiters and job seekers to connect with each other, and the underlying A.I model analyses the information and produce a ranking of compatibility scores. Extensive testing is done for testing the accuracy of the system, efficiency of performance and usability by the users.

A. System Architecture

The system architecture takes the paradigm of 3 tier modular hence defining the concerns and enabling scalability. It has three main layers, which are discussed in the sections below.

- Frontend Layer: The frontend is implemented using React.js, providing an interactive and responsive user interface. Recruiters can post job descriptions, specify skill requirements, and view ranked candidate recommendations. Job seekers can upload resumes and monitor the status of their applications through the user-friendly interface.
- Backend API Layer: The backend is developed in FastAPI and it is responsible for implementing the business logic, parsing the resume, NLP and calculating the similarity. This layer exposes the API on the basis of the crashing in a relevant restful manner for the frontend to work with the AI model with a capacity of real-time data processing and the high-performance system operation.
- Database Layer: MongoDB is used as the database management system due to its flexibility in handling semi structured and unstructured data. It is an efficient way to store parsed resumes, job descriptions, extracted skills and similarity score and the user metadata with the capability for a scalable data management.

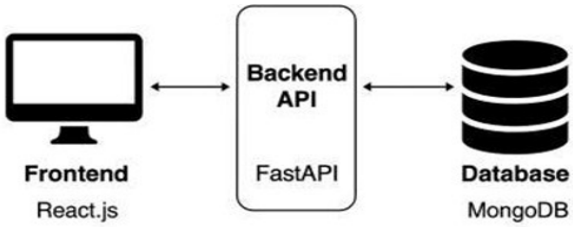

Fig. 2. Smart Hire AI (Placeholder) System Architecture

The modular architecture is leading to the scalable access to each layer individually and sanctions the eventual merging of complex artificial intelligence tools into current AI models, or merging of third-party. The overall architecture of the Smart Hire AI system in its framework is discussed in fig.2.

B. Data Preprocessing and Feature Extraction

Before the actual computations of similarity, there is a rigorous preprocessing program on resumes and job

descriptions that is stipulated in order to improve the quality of the data and abate the undesirable noise. Some of the stages that are available in the preprocessing pipeline include the following steps:

1. The elimination of stop words, punctuation characters and special characters.
2. Lemmatization and tokenization are important process which are involved in the normalization of texts.
3. They use case normalization which is used to provide consistency in the textual data.

Table I: Preprocessing Steps

No.	Step	Purpose
1	Stop-word Removal	Removes common irrelevant words
2	Tokenization	Splits text into tokens
3	Lemmatization	Converts words to base form
4	Case Normalization	Standardizes text case
5	NER	Extracts skills and experience

Then, Named Entity Recognition using spaCy is implemented to recognizes and extracts the essential entities, which include technical skills, education, working experience, and domain specific terms. The step facilitates representation of unstructured resume data that is mentioned in a structured way.

C. Algorithm Design

The system of matching resumes and job belongs to the type of semantic similarity analysis. Textual resumes and job descriptions go pre-processed and relevant entities are extracted. The processed text is converted to numerical vector representations using either TF–IDF or Word2Vec engine which captures the contextual and semantic meaning beyond just matching keywords with.

Cosine Similarity is similarity function is used to calculate the similarity between candidate's resume and job description by calculating the angular distance between feature vectors. Candidates are shortlisted, by predefined threshold value, which is 0.6 (60%).

Resume–Job Matching Process:

The system extracts textual information, firstly on resume, job description. The text extracted is preprocessed on removing stop words, and special characters, this is then followed up by tokenization and lemmatization. Named Entity Recognition or NER is then applied to identify key entities, skills, education and experience being key. Next both documents are converted into the vector version of the data using TF–IDF or Word2Vec. Cosine similarity is calculated in between the two vectors to get a matching score. Candidates having Similarity Score more or equal to 0.6 are made short-list and others are rejected.

Table II: Candidate Selection Based on Similarity Score

Similarity Score Range	Decision
≥ 0.60	Candidate Shortlisted
< 0.60	Candidate Rejected

D. Result Generation and Ranking

After figuring out similarity scores, the shortlisted candidates are ranked in descending order of their match scores. Recruiters are given a prioritized list of applicants that makes it easier and more informed to make decisions. This ranking mechanism greatly decreases the manual screening effort and guarantees both consistency and objectivity of evaluation in candidates.

E. System Evaluation

In an attempt to prove the effectiveness of the proposed system, various evaluation metric is used, which are as follows:

- Accuracy Evaluates fit of resume and job.
- Performance Efficiency: It measures the responsiveness and the speed of the total processing.
- Usability: Consists of the customer satisfaction and the performance of the interface.

Experimental evidences indicate that there is significant reduction in candidate shortlisting time and the matching accuracy is high as compared to conventional Applicant Tracking System (ATS) systems.

IV. DISCUSSION AND RESULTS

The results of the experiment with Smart Hire AI system testify to the fact that the combination of Natural Language Processing (NLP) and Machine Learning (ML) algorithms can help to increasing the efficiency, accuracy and impartiality of modern hiring. The system reduces human interference and creates an objective and consistent assessment of applicants by automating resume parsing and candidate -job matching. Moreover, Smart Hire AI will have a data-driven strategy, as the factor of human subjectivity

will be reduced, and the impact of bias and inefficiencies typical of the traditional hiring process will be minimized.

A. Performance Evaluation

The system was tested with several cases of sample resumes and job descriptions in different areas of technical expertise, such as web development, data analysis, and machine learning. Individual resumes have been processed by the NLP pipeline to obtain the appropriate information, which includes skills, experience, and education background. Afterward, similarity computation was carried out methodically using vectors to match applied information of the candidates against job demand which resulted in a quantitative score of the skills match of an individual candidate. This resulted in shortlisting the candidates who scored a similarity of 60 percent and above on the filter automatically to be considered by the recruiter. This is a threshold method of being inclusive where applicants are given an opportunity to advance in the recruitment process by meeting the majority of, but not all of the job requirements.

Table III: ATS vs Smart Hire AI Comparison

Metric	ATS	Smart Hire AI	Gain
Shortlisting Time	~5 min	~30 sec	–80%
Accuracy	~50%	~60%	+10%
Bias	High	Low	Reduced
Scalability	Limited	High	Improved

B. Bias Mitigation and Equity

One of the biggest issues covered by the proposed system is biased recruitment. Conventional Applicant Tracking Systems (ATS) tends to be biased because of the rigid filtering method and accomplished keywords matching. Smart Hire AI, conversely, analyzes candidates according to the suitability of the given context and skill of a candidate, instead of the purely superficial features of the resume. The strategy will lead to a more equitable human resource recruitment process and enhance diversity in the workforce by eliminating candidates unfairly during the preliminary screening process.

C. Scalability and System Reliability

Smart Hire AI takes the form of a modular architecture which supports efficient processing of high volumes of resumes without necessarily impacting performance. The inclusion of FastAPI as the back-end processing unit and MongoDB as a flexible data storage unit will guarantee the scalability of the requests and unstructured data simultaneously. Experimental evidence shows that there is a steady performance of the system even when the number of applications rises, which clearly shows that it can work in the enterprise-level recruitment setup.

D. Limitations

Although this system has shown a good performance, a number of limitations were found after the evaluation of the system. There is accuracy of the system with the quality of the submitted resumes and their format. Very unstructured or picture-based resumes could be incorporated into the intake of the system partially or possibly incorrectly. Besides, NLP models can be very effective at capturing semantic meaning, however, the fact that they may not pose non-standard terminology, abbreviations or very domain-specific vocabulary that is not represented in the training data can lower matching performance.

E. Future Scope

In order to overcome these shortcomings, future improvements can consider:

- Optical Character Recognition (OCR) to extract images in resumes and process them.
- Formation of field-specific skill sets identifying the vocabularies to form understanding in context.
- Introduction of the multilingual NLP recruiting support.
- Transparency enhancement and enhancement of user trust through addition of explainability mechanisms.

The ethical aspects including privacy of data, transparency of the algorithm, and equity will continue to be the important aspects of the further development and implementation of AI-based recruitment systems.

F. Discussion Summary

In general, the results prove the pragmatism, scalability, and transparency of Smart Hire AI when it comes to resolving modern recruitment issues. The system is much quicker, more precise, and more equitable than conventional ATS platforms as it takes advantage of the NLP and ML approaches and has a

specific similarity level. Such findings confirm the given approach and allow developing a firm basis of the subsequent research and innovation in the intelligent Human Resource Management (HRM) systems.

V. CONCLUSION AND FUTURE WORK

This paper has shown that Smart Hire AI is an effective and strong use of AI-based automation, which may alter the contemporary recruitment practice. The system transforms into a system that automatically executes the most important hiring processes by combining Natural Language Processing (NLP) and Machine Learning (ML) capabilities to identify candidates, those skills, and potential job matches. A 60% skill match threshold will create a compromise between inclusivity and reliability so that viable candidates are not thrown out at too early a stage and yet effective shortlisting criteria are adhered to. The findings of the experiments prove that Smart Hire AI is better than the traditional Applicant Tracking Systems (ATS) in terms of processing rate, reading between the lines, and reducing bias. The system saves a lot of time on recruiting and workload during operation, which is used by the recruiters to engage in other action-oriented practices that can add value to the organization including interviews and last evaluations instead of screening resumes manually. Besides, the assessment of candidates is fairer and more consistent because semantic similarity rather than strict key matching is used. On the technological side, the platform is constructed based on lightweight and open-source technologies that include React.js, FastAPI and spaCy and can therefore be easily, scaled and tailored to organizations of any size and field. The modular architecture allows the smoother combination of advanced AI models, which means that future extensions of the system do not require a massive redesign. Therefore, both small business and large recruitment process settings are suitable to Smart Hire AI. Whereas effective, there are several ways of improvement. Sentiment analysis as a method to derive behavioral hints by examining textual reactions and deep learning-based models of personality prediction will be introduced in a later implementation and will help to more full-bodied evaluate cultural and role-specific fit assessments. An extension will involve incorporation of voice and video analysis of interviews as it will make it simpler to assess the candidates in a multimodal manner and have a more subtle assessment than the resume analysis.

Candidate profiling will be maximized through planned integration with third-party recruitment, including Indeed and LinkedIn, using API connections so that the real-time labor market information is utilized. To help resolve ethical and governance issues, future iterations of Smart Hire AI will use explainable AI (XAI), which offers clear explanation of why candidates are ranked in particular way and why specific decisions have been made. Altering pure concepts like fairness, accountability, and trust in the hiring process that involves AI will not only be supported with AI-concerned detection and mitigation systems.

Conclusively, Smart Hire AI is a strong platform of the intelligent Human Resource Management (HRM) systems of the new generation. The adoption of responsible AI, transparency and scalability make the platform superior than the typical ATS solutions with reliable and responsible environment of recruitment. The results of the paper is added to existing research relating to the field of AI UAE recruiting, and can serve as an indicator of future trends in recruiting human talent based on evidence and fairness.

REFERENCES

[1] J. Sullivan, "Talent acquisition analytics: Using data to improve recruiting effectiveness," Journal of Human Resources Management, vol. 45, no. 3, pp. 21–28, 2018.

[2] D. Jurafsky and J. H. Martin, Speech and Language Processing, 2nd ed. Pearson Education, 2019.

[3] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. O'Reilly Media, 2019.

[4] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Proc. European Conf. on Machine Learning, 1998, pp. 137–142.

[5] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[6] M. Mishra and S. K. Jain, "Automated resume screening using machine learning," International

Journal of Computer Applications, vol. 182, no. 7, pp. 20–25, 2018.

[7] A. Singh and P. Gupta, "A survey of resume parsing and matching techniques using NLP," International Journal of Advanced Research in Computer Science, vol. 10, no. 2, pp. 45–51, 2019.

[8] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring," in Proc. ACM Conf. on Fairness, Accountability, and Transparency (FAccT), 2020, pp. 469–481.

[9] H. Mehrabi et al., "A survey on bias and fairness in machine learning," ACM Computing Surveys, vol. 54, no. 6, pp. 1–35, 2021.

[10] HireVue, Artificial Intelligence in Hiring: Ethical Considerations, White Paper, 2020.

[11] LinkedIn Corporation, Talent Insights and AI-Based Recruitment Analytics, Technical Report, 2021.

[12] Indeed Hiring Lab, Machine Learning Applications in Large-Scale Recruitment Platforms, Industry Report, 2022.

[13] Y. Liu, C. Zhang, and Q. Huang, "Semantic similarity computation using TF-IDF and cosine similarity," in Proc. IEEE Int. Conf. on Big Data, 2019, pp. 394–399.

[14] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[15] European Commission, Ethics Guidelines for Trustworthy Artificial Intelligence. Brussels, Belgium, 2019.