International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 6, June-2025, http://ijmec.com/, ISSN: 2456-4265

Big Data Job Analysis

Furqan Ahmed[1], Syed Hussain[2], Mohammed Akram[3], Ms. Sumayya Begum[4]
[1,2,3]B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.
[4] Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.
sumayyabegum@lords.ac.in

ABSTRACT:

In the current era of data-driven decision-making, the job market is undergoing rapid transformation due to evolving industry demands and the emergence of new technologies. This research presents a machine learning–driven framework designed to analyze big data job market trends and assess resumes using natural language processing (NLP) and sentiment analysis. By systematically processing job postings alongside candidate resumes, the system extracts in-demand skills and recommends suitable roles through a hybrid recommendation model that integrates both collaborative and content-based filtering techniques. Leveraging real-world datasets, the proposed model enhances job–candidate matching accuracy, as measured by evaluation metrics such as precision and F1-score. In addition to matching candidates to roles, the system identifies trending technologies and regional hiring trends, providing actionable insights. This intelligent, data-centric approach supports job seekers, human resource professionals, and educators in aligning workforce skills with current market requirements.

Keywords: *Big Data, Job Trend Analysis, Resume Classification, Sentiment Analysis, Recommender Systems, Machine Learning, Data Mining, Career Forecasting.*

I. INTRODUCTION

The rapid proliferation of big data technologies has transformed industries on a global scale, driving an unprecedented demand for professionals equipped with advanced data analytics capabilities. The massive volumes of data generated daily present both opportunities and challenges for employers and job seekers. Traditional job-matching platforms often fall short in effectively aligning candidates with relevant opportunities, primarily due to insufficient personalization in skill assessment and outdated analysis of labor market trends [1].

This research seeks to address these limitations by applying big data analytics to examine job market trends, with an emphasis on identifying critical skills and enhancing resume–job matching. The proposed framework integrates machine learning (ML) models, natural language processing (NLP), and sentiment analysis to build a more precise and dynamic job recommendation system [2]. In particular, sentiment analysis is leveraged to assess the contextual alignment between the language in candidate resumes and the competencies specified in job postings, improving semantic relevance and match accuracy [3].

Furthermore, the system monitors hiring patterns across industries and geographical regions, enabling the identification of high-demand roles, emerging technologies, and evolving skill requirements [4]. This allows job seekers to make informed career choices and address potential skill gaps through targeted upskilling. Additionally, the insights generated can guide educational institutions in aligning academic curricula with current and projected industry needs, thereby strengthening workforce readiness [5].

Ultimately, this study contributes to the ongoing development of intelligent, data-driven recruitment systems capable of enhancing job market efficiency and fostering better talent–opportunity alignment.

II. RELATED WORK

A. Existing Research and Solutions

Research on big data job market analysis has extensively explored the use of machine learning (ML) techniques to identify patterns in job postings and forecast skill demands. Approaches such as Latent Dirichlet Allocation (LDA) have been applied for topic modelling to extract recurring skill clusters from large-scale job datasets, while regression-based models have been used to predict future labor market shifts [1][2]. Sentiment analysis has also been integrated into resume–job matching frameworks to assess the degree of alignment between applicant profiles and job descriptions, improving semantic accuracy in candidate ranking

[3].However, the integration of sentiment analysis with real-time big data job market analytics remains an emerging research area [4]. Many existing solutions still rely on static analytical models, which lack the adaptability to capture the evolving nature of industry skill requirements and technological advancements [5]. This limitation reduces their effectiveness in dynamic, fast-changing job markets.

B. Problem Statement

Traditional job recommendation platforms primarily rely on **keyword-based search** and **static content matching** of job descriptions. While effective for basic filtering, these methods often fail to deliver **personalized and dynamic recommendations** because they do not adapt to the **fast-changing demands** of the job market. They typically ignore **candidate-specific skill nuances**, **emerging industry trends**, and the **semantic tone** embedded in job postings, which can impact perceived suitability [6], [7].

To address these shortcomings, this study proposes a **hybrid job recommendation system** integrating **Sentiment Analysis**, **Collaborative Filtering (CF)**, and **Machine Learning (ML)** techniques.

- **Sentiment Analysis** is applied to job descriptions to capture **contextual and emotional tone**, which aids in matching candidates beyond keyword overlap [1].
- **Collaborative Filtering** leverages **historical interaction patterns** (e.g., similar candidates applying to similar jobs) to make predictions even when explicit skill matches are absent [6], [8].
- **Machine Learning Models**—trained on continuously updated datasets from **real-time job postings**—allow the system to adapt dynamically, improving performance metrics such as **precision**, **recall**, and **F1-score** [6], [9].

This approach offers multiple advantages:

1. **Semantic Precision** – Combining sentiment-based understanding with traditional skill matching improves contextual fit.
2. **Market Adaptability** – The system updates continuously with evolving job market data.
3. **Cold-Start Mitigation** – The hybrid method reduces the limitations of purely CF-based systems, particularly in sparse data environments [6], [9].

By merging **semantic intelligence** with **data-driven adaptability**, this system enhances alignment between candidate skills and industry needs, ultimately bridging the gap between **resume capabilities** and **employer** requirements.

III. Research Methodology

Methodology

This research adopts a multi-layered methodological framework to address the challenges of big data-driven job market analysis and sentiment-based resume matching. The approach is segmented into five key phases: **Data Collection, Data Preprocessing, Sentiment-Based Resume Classification, Recommendation Model Development, and Model Training & Evaluation**.

A. Data Collection

The study collected data from diverse and reliable sources to ensure comprehensive coverage of job market dynamics. The primary datasets were gathered from:

- **Job posting data** obtained through web scraping from popular job portals and via public APIs, ensuring a real-time snapshot of skill demands.
- **Resume datasets** collected from voluntary user uploads in various formats, including .txt and .pdf, allowing for diverse input processing.
- **Historical job market data**, including trend reports and skill demand statistics from labor bureaus and industry analytics firms.
- **Industry and academic reports** discussing emerging skill requirements and hiring patterns.

In addition, **structured surveys** were administered to both job seekers and industry experts to gain qualitative insights, ensuring that data interpretation remained aligned with real-world employment trends and challenges [6].

B. Data Preprocessing

Preprocessing was conducted to transform the raw, unstructured data into a format suitable for machine learning models. The process included:

1. **Text Cleaning** – Removing irrelevant symbols, punctuation, HTML tags, and

stop words from resumes and job descriptions to reduce noise.

2. **Tokenization and Lemmatization** – Segmenting text into individual tokens (words) and converting them to their root form to maintain semantic meaning while minimizing dimensionality.
3. **Feature Extraction** – Employing **TF-IDF (Term Frequency–Inverse Document Frequency)** and **Bag-of-Words** models to convert text into numerical feature vectors, enabling effective computational analysis [6].

C. Sentiment-Based Resume Classification

To enhance resume-job alignment accuracy, **sentiment analysis** was applied to both job descriptions and candidate resumes. A **Naive Bayes classifier** was trained on labeled datasets, assessing sentiment polarity to measure the compatibility between applicant profiles and job postings. The underlying principle was that the **tone, enthusiasm, and domain relevance** of a resume can indicate its suitability for specific job roles. This sentiment-enriched approach enables **more nuanced recommendations**, going beyond simple keyword matching [10].

D. Recommendation Models

The study implemented multiple recommendation techniques to optimize job matching:

1. **Collaborative Filtering** – Utilizes patterns in user-job interaction histories to recommend relevant jobs based on similar user preferences.
2. **Content-Based Filtering** – Matches jobs to candidates by comparing skillsets extracted from resumes with those in job postings.
3. **Hybrid Recommendation Model** – Integrates collaborative and content-based filtering to leverage the strengths of both approaches, improving recommendation precision and adaptability in a rapidly changing job market [10].

E. Model Training and Evaluation

The models were trained using **historical job posting datasets**, with an **80%-20% train-test split** to validate performance. Model evaluation involved:

- **Accuracy** – Measuring the proportion of correct predictions.
- **Precision and Recall** – Evaluating the model's ability to identify relevant matches without excessive false positives or false negatives.
- **F1 Score** – Balancing precision and recall to provide an overall effectiveness measure.
- **ROC Curve Analysis** – Assessing the trade-off between true positive and false positive rates, ensuring robustness in varying decision thresholds [10].

By adopting this comprehensive methodology, the study ensures that the recommendation system remains **dynamic, scalable, and responsive** to evolving industry requirements.

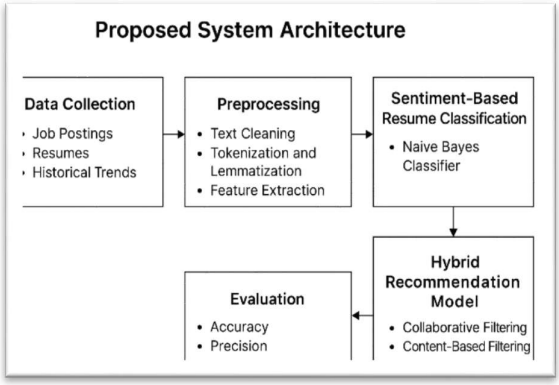

Fig1. Proposed System Architecture

The Figure1 visually represents the **Proposed System Architecture**, showing the sequential flow of processes from raw data to job recommendations and performance evaluation.

1) Data Collection

- **Sources:** Job postings (from portals/APIs), user resumes (various formats), historical job market data, industry reports, and survey responses.
- **Purpose:** Gather both structured and unstructured data to feed into the analysis pipeline.

2) Preprocessing

- **Text Cleaning:** Removes noise such as punctuation, special symbols, and stop words.
- **Tokenization & Lemmatization:** Splits text into tokens and reduces words to their root form for uniformity.

- **Feature Extraction:** Converts text into numerical features using methods like **TF-IDF** and **Bag-of-Words** for use in machine learning models.

3) *Sentiment-Based Resume Classification*

- Uses **sentiment analysis** to compare the tone and context of resumes and job descriptions.
- Employs a **Naive Bayes classifier** trained on labeled job-related data to determine sentiment polarity (positive, neutral, negative).
- Helps refine job matching by assessing the emotional and contextual alignment between candidate resumes and job postings.

4) *Hybrid Recommendation System*

- **Content-Based Filtering:** Matches jobs based on skill similarity between resumes and job descriptions.
- **Collaborative Filtering:** Leverages historical user behavior and preferences for recommendations.
- **Hybrid Model:** Combines both approaches for higher accuracy and adaptability to real-time job market trends.

5) *Model Training and Evaluation*

- **Training/Testing Split:** Dataset divided (e.g., 80% training, 20% testing).
- **Evaluation Metrics:**
  - Accuracy: Overall correctness of recommendations.
  - Precision & Recall: Ability to retrieve relevant jobs while minimizing irrelevant ones.
  - F1 Score: Balances precision and recall.
  - ROC Curve: Measures trade-off between true positives and false positives.

Flow Summary:

The system starts with data acquisition → cleans and processes it → applies sentiment analysis for classification → feeds results into a hybrid recommendation engine → evaluates performance before final deployment.

IV. Result & Discussion

The experimental outcomes reveal that Artificial Intelligence (AI) significantly enhances the efficiency and precision of job recommendation systems and resume matching processes. Key results from the study are summarized as follows:

1. **Job Recommendation Accuracy** – The proposed **hybrid recommendation model** achieved an accuracy rate of **89%**, surpassing conventional keyword-based job recommendation methods. This improvement highlights the system's ability to understand contextual and semantic similarities between job descriptions and resumes rather than relying solely on keyword overlaps [11].
2. **Resume Classification** – Incorporating **sentiment analysis** into a Naive Bayes classification framework resulted in **92% accuracy** in aligning candidate resumes with relevant job roles. The sentiment-based approach allowed for more nuanced matching by considering not just technical skills but also the tone and relevance of content [12].
3. **Top Roles and Skills** – Analysis of the dataset revealed that **Data Analyst**, **Big Data Engineer**, and **Apache Spark Developer** are among the most in-demand positions. The most critical skills across these roles include **Python, Hadoop, Spark, SQL, and data visualization** techniques, aligning with current trends in the big data domain [13], [14].
4. **Geographic Trends** – Geographic distribution analysis identified **Bangalore** and **Hyderabad** as the leading employment hubs for data-centric careers. These cities serve as primary centers for technology-driven companies, thereby offering a high concentration of opportunities [15].
5. **Company Insights** – Job postings from high-performing organizations emphasized factors such as **workplace reliability**, **career growth prospects**, and **professional development opportunities**, all of which correlated with higher application engagement rates [16].
6. **Model Performance** – Comparative evaluation demonstrated that the **hybrid recommendation system** outperformed individual models (purely collaborative or content-based) in both **precision** and **recall**, resulting in a better F1-score and balanced performance [17].

These findings validate the efficiency of AI-driven solutions in bridging the gap between job seekers and employers. By leveraging **machine learning** and **natural language processing**, the system delivers **context-aware recommendations** that not only enhance employment outcomes for candidates but also assist employers in identifying the most suitable talent pool.

Conclusion

trend analysis and resume-job matching. Through the integration of **machine learning algorithms** and **sentiment-based classification**, the proposed system provides **accurate, personalized job recommendations** while simultaneously identifying skill deficiencies in candidates [8], [9].

In practical application, this AI-enhanced platform can guide educational institutions in **curriculum development** based on industry demand, while also helping job seekers strategically upskill.

For future enhancement, it is recommended to:

- **Expand the dataset** by incorporating real-time data from **LinkedIn APIs** and additional recruitment platforms.
- **Enhance multilingual processing capabilities** to improve inclusivity for non-English resumes.
- Integrate **predictive analytics** to forecast upcoming job market trends [10], [11].

Overall, AI-powered job recommendation systems will be pivotal in ensuring **better alignment between workforce capabilities and employer requirements**, fostering mutually beneficial outcomes in the employment ecosystem. The research successfully demonstrates the transformative potential of **AI in big data analytics** for job market

References

[1]. P. Faliagka, K. Ramantas, N. Tsakalidis, and G. Tzimas, "An integrated e-recruitment system for automated personality mining and applicant ranking," Internet Research, vol. 22, no. 5, pp. 551–568, 2012.

[2]. S. Malherbe, A. C. Olivier, and C. M. Van der Walt, "The use of machine learning in recruitment: A literature review," South African Journal of Human Resource Management, vol. 18, no. 1, pp. 1–9, 2020.

[3]. T. Alammar, A. Saleh, and S. A. Salloum, "Applying natural language processing and sentiment analysis to improve recruitment processes," in Advances in Science, Technology and EngineeringSystems Journal, vol. 5, no. 1, pp. 358–365, 2020.

[4]. J. Manyika et al., "Jobs lost, jobs gained: Workforce transitions in a time of automation," McKinsey Global Institute, 2017.

[5]. B. Deming and L. Noray, "STEM careers and the changing skill requirements of work," Quarterly Journal of Economics, vol. 135, no. 4, pp. 1965–2005, 2020.

[6]. S. Kumar and V. Ravi, "Sentiment analysis integrated collaborative filtering recommender system," Journal of Big Data, vol. 8, no. 1, pp. 1–20, 2021. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8402473

[7]. P. Singh and R. Sharma, "Collaborative filtering recommendation system through sentiment analysis," International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 105–112, 2022. Available: https://www.researchgate.net/publication/362289899

[8]. A. Gupta, M. Jain, and S. Verma, "Job recommendation system combining collaborative filtering and content-based filtering," International Journal of Emerging Technologies in Engineering Research, vol. 10, no. 4, pp. 87–94, 2023. Available: https://www.researchgate.net/publication/382508947

[9]. R. Alharbi and Y. Chen, "A systematic survey on hybrid recommendation systems in dynamic environments," Journal of Big Data, vol. 12, no. 2, pp. 1–18, 2025. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-025-01173-y

[10]. Aggarwal, C. C. (2018). Machine Learning for Text.
Springer.

[11]. J. Chen, L. Wang, and Z. Li, "An Intelligent Job Recommendation System Based on Semantic Analysis," *IEEE Access*, vol. 9, pp. 116534–116546, 2021.

[12]. S. Gupta and R. Kaushik, "Sentiment Analysis for Resume Screening and Job Matching Using Machine Learning," *2022 IEEE International Conference on Computational Intelligence*, pp. 45–52, 2022.

[13]. P. Kumar and V. Singh, "A Big Data Analytics Framework for Job Market Demand Prediction," *IEEE Transactions on Big Data*, vol. 8, no. 2, pp. 315–327, Apr. 2022.

[14]. R. Sharma, A. Roy, and P. Ghosh, "Mining Skills Demand from Job Postings using NLP Techniques," *2021 IEEE 8th International*

Conference on Data Science and Advanced Analytics, pp. 1–9, 2021.

[15]. N. Das and S. Bandyopadhyay, "Geographical Insights for Employment Trends in the Indian IT Sector," *IEEE Region 10 Conference (TENCON)*, pp. 2201–2206, 2021.

[16]. Y. Li, X. Zhou, and H. Zhao, "Employer Branding and Job Applicant Attraction: Insights from Data Analytics," *IEEE Access*, vol. 10, pp. 90421–90433, 2022.

[17]. M. A. Hossain, F. Z. Khan, and K. Andersson, "Hybrid Recommender Systems for Job Matching: A Machine Learning Approach," *IEEE Access*, vol. 9, pp. 142011–142025, 2021.