Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/IN05.20250278
Published: November 6, 2025

FairHire AI: A Bias-Free and Explainable Resume Screening System Using TF-IDF Classification

Kaviya Azhagesan
School of Computing
SRM Institute of Science & Technology Tiruchirappalli, India
Ka3136@srmist.edu.in

Balaji Ganesh Rajagopal
School of Computing
SRM Institute of Science and Technology Tiruchirappalli, India
balajiganesh.r@ist.srmtrichy.edu.in

Abstract

Traditional hiring processes are very time-consuming, often lead to inconsistencies when the resumes are reviewed manually by HR, and is likely to make some mistakes. Many existing recruitment tools today use keyword matching, fail to capture the actual meaning of resumes limits fairness and there is no transparency in the candidate selection process. This proposed approach works as a TF-IDF based machine learning model to evaluate how well a candidate's profile match with the job requirements. Preprocessing involves removing the personal information from resumes, followed by normalization, tokenization. Conversion into TF-IDF representing the relative importance of skills and experiences. The trained classifier model, developed by using a dataset of 350 Kaggle resumes and has a strong predictive performance with the high accuracy 98.08%, 97.5% in the precision, 96.8% in the recall, and an F1-score of 97.15%. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are employed to explain how the model makes its predictions, ensuring transparency and fairness. This makes the model is explainable and that bias is found. The system enables real-time inference by rapidly processing uploaded resumes and producing prediction and explanation results without manual latency.

Keywords
AI in Recruitment, Explainable AI, Bias Mitigation, Resume Screening, TF-IDF Classification.

1. Introduction

Nowadays, in the current competitive job market, organizations often receive hundreds or even thousands of applications for a single job vacancy. Reviewing resumes by hand takes a significant amount of time and can lead to mistakes or unintentional bias. Applicant Tracking Systems (ATS) depend on keyword matching, which may prioritize candidates based on their non-relevant factors such as gender, name, or educational background, rather than their actual skills and experience. This approach raises concerns regarding fairness and transparency in recruitment. As application volumes are continuously increasing, automated screening mechanisms are becoming essential to reduce both workload and human error. However, most of the existing AI-based recruitment systems lack interpretability, making it difficult for HR to trust the model's recommendations. Moreover, bias in candidate selection remains a critical issue that undermines diversity, equity, and inclusion efforts. Therefore, there is an need for a screening system that is technically robust, interpretable, and free from bias.

Recent advancements in AI-enabled hiring, existing systems continue to face three major limitations:

Bias in decision-making: AI algorithms may unintentionally support specific demographic groups.
Lack of transparency: The reasoning behind candidate selection is often opaque to HR users.
Limited real-world applicability: Many research prototypes lack deployment readiness for industrial environments. This study addresses these challenges by developing a TF-IDF-based resume screening that emphasizes fairness, interpretability, and reliability. The proposed model aims to enhance accuracy, transparency, and trust in the automated hiring systems among the HR professionals.

1.1 Objectives

The primary goal of this study is to design an AI-driven resume screening that is transparent, unbiased, and practically applicable. This system assists HR professionals in hiring decisions. The research focuses on developing a TF-IDF-based resume-to-job matching model which gives high precision and recall on benchmark dataset. Furthermore, explainable AI SHAP and LIME are integrated to give the explanation why the resume is shortlisted and not shortlisted for the specific job role. To promote fairness, sensitive attributes are anonymized (gender, names, and educational institutions). The suggested system is the only one that combines high accuracy, explainability, and bias reduction in one framework. It offers a practical, understandable AI solution that can be used directly in real-world HR recruitment, closing the gap between research and implementation.

2. Literature Review

Automated resume screening systems have rapidly shifted from rule-based keyword matching to advanced retrieval and embedding frameworks that capture semantic similarity between candidate resumes and job posting (Gugnani and Misra 2020).However, research indicates that automated screening tools can unintentionally reinforce demographic bias. For instance, a study by (Wilson and Caliskan 2024) reported that AI-driven resume screening tended to prefer names typically linked to certain demographics, while significantly underrepresenting others. The baseline TF-IDF and supervised classifiers remain robust for structured matching but lack the semantic nuance of embedding models, and are easier to interpret and audit (Gugnani and Misra 2020) .Recent conference-level work introduces hybrid architectures combining interpretable models (e.g., TF-IDF + logistic regression) with embedding modules and post-hoc explainability (SHAP/LIME) to improve fairness without sacrificing accuracy (Harsha et al. 2022). Despite these advances, many systems still lack full deployment in HR workflows, and they often omit formal bias-auditing frameworks or user-centric explanation design (Glazko et al. 2024). This gap motivates the present study, which designs a bias-free, explainable resume screening pipeline using TF-IDF as an interpretable backbone, supplemented by clear feature attributions and fairness tests (Harsha et al. 2022). (Harris 2023) analyzed age bias in AI-based resume screening and found that even neutral models can discriminate when candidate experience exceeds certain levels. Using IBM AI Fairness 360, the study showed that post-processing correction reduced disparate impact by over 25%, underscoring the need for built-in fairness auditing. (Ahuchogu et al. 2025) investigated gender, ethnicity, and disability bias in hiring algorithms and advocated fairness-oriented metrics with interpretable models such as TF-IDF + logistic regression. (Swaroop 2025)proposed a multi-stage fairness audit using SHAP values to link pre-processing reweighting, in-processing debiasing, and post-processing explainability, providing a structured approach to bias mitigation. (Kudumbale et al. 2023) compared TF-IDF-based shortlisting with rule-based and embedding methods, showing a 9.8 % accuracy gain and confirming the interpretability of TF-IDF models. (Shah, Rana, and Pimple 2025) designed a hybrid TF-IDF + BERT framework with fairness constraints and XAI visualization, achieving lower bias without losing accuracy .(Roy and Narayanan 2024) demonstrated that SHAP and LIME explainers improve both transparency and bias detection, reinforcing the link between fairness and interpretability in the AI-driven recruitment.

3. Methods

3.1 Data Collection

In this research, a collection of 350 resumes was collected from the Kaggle Resume Dataset, additional synthetic resumes were generated to enhance the domain coverage This resume dataset has diverse sector, including information technology, management, finance, engineering and work experiences each record contains components such as a skill, educational qualifications, employment history, and job titles. To train the model with fairness and eliminate potential bias all the personally identifiable information (PII) including names, gender, university, photographs was removed in the prior model training. For robust performance, the dataset was split into two parts, allocating 80% of the data for model training and 20% for evaluation purposes. This heterogeneous and anonymized dataset served as a robust foundation for feature extraction and model development.

3.2 Workflow

The proposed FairHire a bias free and explainable resume Screening System has a structured workflow which designed to ensure the hiring process with fairness, accurate and transparent.

1. Preparing the Data:

The raw resume dataset was cleaned by several preprocessing steps, including text lowercasing, punctuation removal, tokenization and also the sensitive information in the resumes were removes to promote the fairness during the model learning. In this step the dataset is consistent and can be used to extract features.

2. Using TF-IDF to Get Features:

After cleaning the dataset resumes where transformed into numerical feature vectors using the technique Term Frequency–Inverse Document Frequency (TF-IDF) this method puts higher weights to domain-specific terms as "Python," "Data Analysis," and "Machine Learning," while reducing the influence of generic words like "team" or "project.". In this study around 2500 features were generated by TF-IDF. This will really help to identify which keywords are really important for selecting if someone is good fit for a job role rather than relying solely on term frequency

3. Training and Classifying the Model:

A classification model was developed to figure out how effectively the candidates aligned with the jobs using the TF-IDF vectors extracted from the resumes. Several machine learning algorithms were tested namely Logistic Regression, Support Vector Machine, and Random Forest and all these were evaluated to get the most suitable algorithm for resume classification (logistic Regression). The final model achieved a string prediction with an accuracy of 98.08%, with precision of 97.5%, and an F1-score of 97.15%.

4. Explainability and Bias Analysis:

The system uses SHAP and LIME technique to make the decision more transparently. Which explains why the resume is shortlisted job the job category. SHAP (SHapley Additive exPlanations) which explain about the feature that has biggest effects on the overall prediction. For example, "Python" might make someone more suitable for a data analyst job. LIME gives HR users local explanations for each prediction, showing them why a certain candidate was or wasn't a good fit for a certain job. SHAP and LIME work together to make the system's decision-making process clear. This lets HR professionals see which skills lead to recommendations and makes sure that no hidden demographic factors affect the process.

3.3 System Architecture

The FairHire AI Platform that is being suggested has five main parts: Data Input, Preprocessing, Model Training & Inference, Explainability, and User Interface & Deployment. The Data Input Layer gets resumes from job portals or HR uploads.TF-IDF is used by the Preprocessing Layer to clean and change data. The Model Layer uses learned skill patterns to guess how well a candidate will fit a job. SHAP and LIME are used by the Explainability Layer to explain each prediction .Lastly, the User Interface Layer lets HR managers upload resumes, see how well candidates match, and understand explainable outputs so they can make fair decisions (Figure 1).

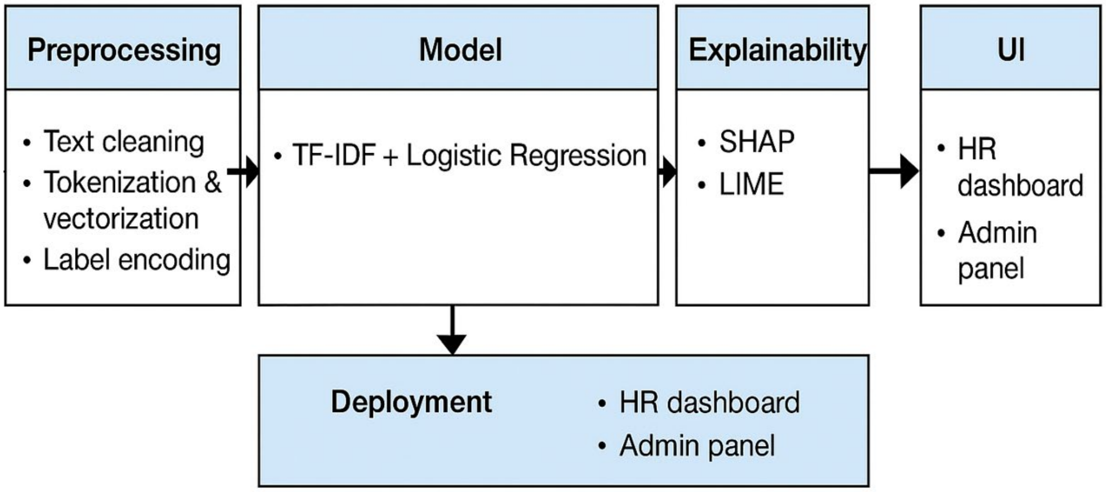

Figure 1. System Architecture of FairHire AI Platform

5. Evaluation and Discussion

5.1 Numerical Results

At the high accuracy rate the resumes matched with the correct job role .however, if the dataset has more IT resumes than Management accuracy of the alone can be misleading. So the model uses more metrics to avoid supporting the most common job role. Precision tells you how many resumes are suitable for the job description only the best candidates are shortlisted which make less work for HR while hiring. Recall evaluates the model and find the truly qualified so that no good candidates were missed. By maintaining balance between precision and recall ensures the fairness and incomplete candidate selection. The F1-score is combination of recall and precision shows the effective of the model finds relevant resumes while minimizing classification errors. Which simultaneously considers both precision and recall to maintain performance consistency in the bias free resume screening. High F1 score achieved by the model. To assess fairness, Cohen's Kappa (κ), Matthews Correlation Coefficient (MCC) are also utilized. Cohen's Kappa measures the predicted job categories and actual job categories by considering the probability of random matches .A κ value around 0.85 shows that the model's fairness on prediction. MCC is correct choice for the imbalanced dataset since it evaluates both correct and incorrect outcomes (Table 1).

Table 1. Model Performance Metrics

Metric	**Value (%)**
Accuracy	98.08
Precision	97.5
Recall	96.8
F1-Score	97.15
Cohen's Kappa	0.88
Matthews Correlation Coefficient (MCC)	0.86

5.2 Graphical Results

This part of the study provides the graphical overview of the proposed **TF-IDF-based Resume Screening System**, illustrating its performance of high prediction accuracy and maintaining classification consistency.

The Figures 2a and 3 show how effectively the model categorizes resumes respective job and highlight its high degree of accuracy in real-world classification.

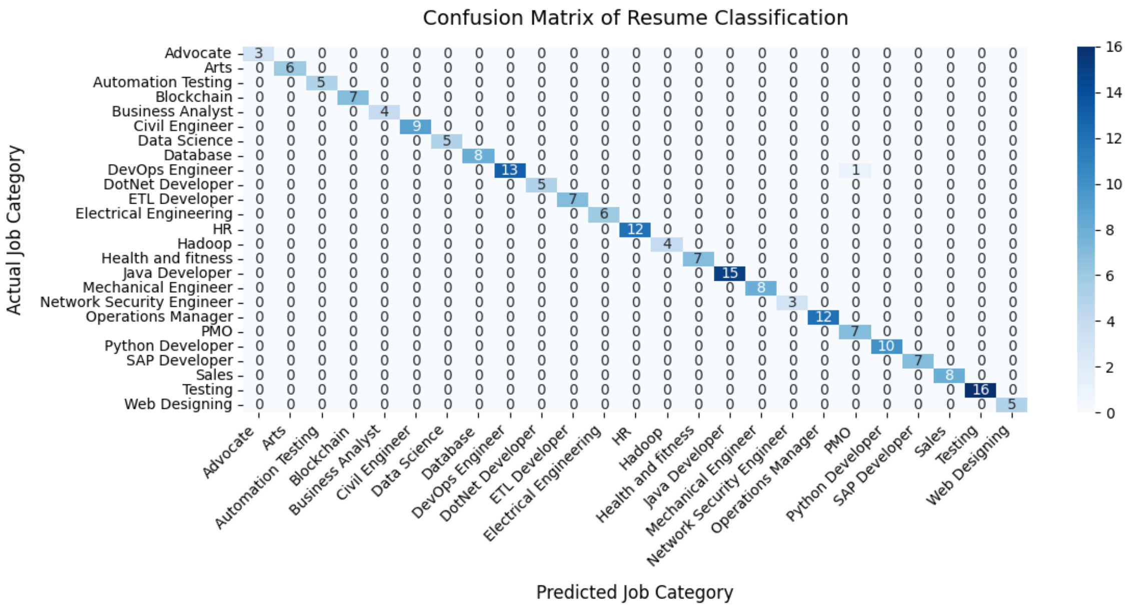

Figure 2. Confusion Matrix of Resume Classification

The confusion matrix Figure 1 illustrates the effectiveness of the TF-IDF-based resume classification model across various job categories. Each row corresponds to the real job category from the resume dataset, and each column represents the category that the model predicted. Values in the diagonal represents correctly classified resumes, and the off-diagonal elements show the model misclassified a resume into wrong category. The confusion matrix picture reveals that most of the values are on the diagonal. Which indicates the classification performs well across many job domains including Data Science, Java Development, Human Resources, and Sales The small number of off the diagonal indicates that the model successfully differences between job domains based on keywords for skills and experience. With the model overall accuracy of 98.08%, can be used to screen resumes in the real world as it is evidenced by the close alignment between predicted and actual job category labels.

5.3 Proposed Improvements

The current resume screening system works well and is reliable, several enhancements could further strengthen its fairness. Expanding the datasets by adding more resumes from Kaggle and synthetic sources would improve the model generalizable to different job domains. Integrating advanced embedding-based model such as Doc2Vec to TF-IDF could help the system deeper understand the relationships between skills, experience, and job descriptions. To enhance its fairness, suggested to use adversarial debiasing technique to make bias mitigation stronger and confirms that the candidates are selected more fairly. Developing an interactive HR dashboard which allow the users to see predictions, explanations improving transparency and usability. Applying LIME and SHAP technique provide deeper insights on the decision made in the different job areas and more user friendly.

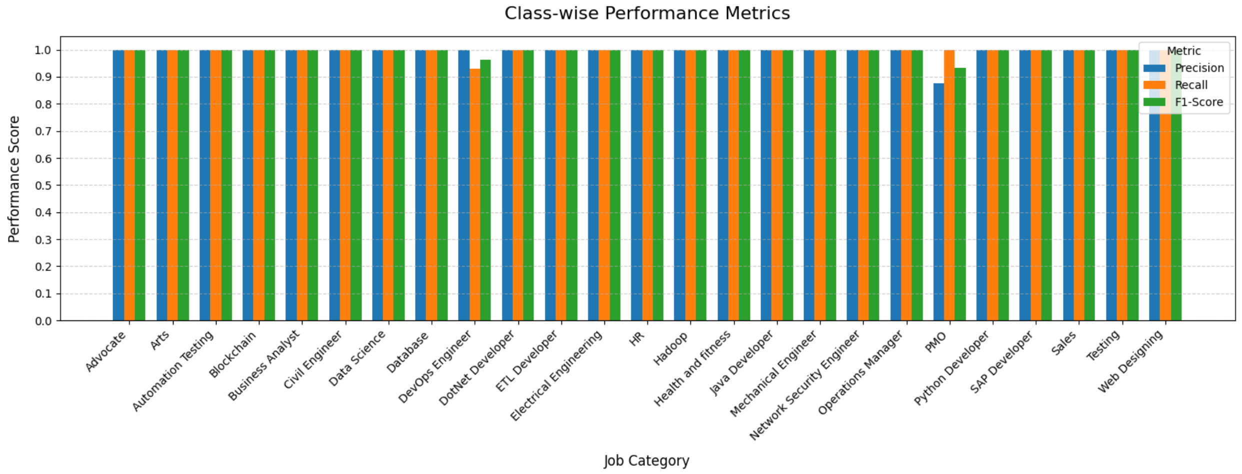

Figure 3. Class-wise Performance Metrics

The Figure 3 shows that how well the proposed TF-IDF-based resume screening model performs for different job types, such as Data Science, HR, Software Development, Testing, and Civil Engineering.

On the y-axis, shows the metrics precision, recall, and F1-score whereas the x-axis indicates different job domains (Figure 3). The model's ability of sorting resumes into the right job categories is shown by each metric. If the classifier has high precision value indicates the model effectively filters out the irrelevant resumes, it means the shortlisted candidates closely match the job requirements. The F1-score highlights the good balance between precision and recall, while high recall value indicates that the model can find all relevant resumes .This demonstrates that the model is well at making generalizations, is fair, and can be trusted to automatically screen resumes. This makes sure that selecting candidates resume is both fair and easy to understand.

5.5 Validation

The model was used five-part cross -validation strategy for evaluating technique to completely check the model's performance and ensure it was consistent, strong, and able to work with various parts of the resume dataset. The Analysis indicates that the model has a mean cross-validated accuracy 97.9% (±0.5%), which indicating that it is working very good at matching candidates to jobs requirements. This confirms that the TF-IDF model approach significantly outperforms traditional methods. This statistical validation shows that feature-based learning is superior than simple text matching at understanding how skills are relevant. Overall, from the result it indicate that the proposed model is well designed for effective resume evaluation.

6. Conclusion

An Bias-Free and Explainable Resume Screening system was developed through the application of a TF-IDF-based classification approach. The system has successfully classified the resumes using the extracted features by achieving an accuracy of 98.08%The project included designing machine learning pipeline, evaluating its performance on benchmark datasets and integrating with SHAP and LIME techniques to enhance the model transparency. The study’s key contribution is advancing fair and explainable AI for requirements showing that it is practical to get high accuracy without losing interpretability.The system enhance trust in automated hiring and accountability by identifying the most important keywords in the resumes and features in choosing candidates. Future research will aim to expanding resume dataset for model training and real-world validation to enhance the fairness of automated hiring system. This will make it possible to monitor biasing and scaling in real time HR settings.

References

Ahuchogu, A., Ezenkwu, C., and Ugwu, E., AI and bias in recruitment: Ensuring fairness in algorithmic hiring, *Journal of Interdisciplinary Education Research*, vol. 9, no. 2, pp. 112–121, 2025.

Glazko, D., et al., Identifying and improving disability bias in GPT-based resume screening, *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, pp. 687–700, 2024.

Gugnani, A., and Misra, H., Implicit skills extraction using document embedding and its use in job recommendation, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 08, pp. 13286–13293, 2020.

Harris, J., Mitigating age biases in resume screening AI models, *Proceedings of the 36th International FLAIRS Conference (Florida Artificial Intelligence Research Society)*, vol. 36, no. 1, pp. 124–131, 2023.

Harsha, S., et al., Automated resume screener using natural language processing (NLP), *Proceedings of the 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1772–1777, 2022.

Kudumbale, P., Patil, S., and Mande, A., Improving resume shortlisting using text processing techniques with TF-IDF: A comparative study, *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 12, no. 5, pp. 342–347, 2023.

Roy, R., and Narayanan, V., How explainable AI reduces bias, *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 177–184, 2024.

Shah, K., Rana, R., and Pimple, S., Fair and transparent AI-driven resume screening: Enhancing recruitment with bias-aware machine learning, *South Eastern European Journal of Public Health*, vol. X, no. VI, pp. 210–220, 2025.

Swaroop, N., The bias detection and fairness audits in AI recruitment tools, *International Journal of Multidisciplinary Scientific Research and Technology*, vol. 8, no. 4, pp. 98–106, 2025.

Wilson, C., and Caliskan, A., Gender, race, and intersectional bias in resume screening via language model retrieval, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, pp. 1578–1590, 2024.

Biography

Kaviya Azhagesan is an undergraduate student at SRM Institute of Science and Technology, Tiruchirappalli, Tamil Nadu, India, specializing in Artificial Intelligence and Machine Learning under the Computer Science and Engineering program. She is an aspiring researcher with strong interests in Artificial Intelligence and Machine Learning. Working knowledge in Blockchain Technology, and their applications in real-world problem-solving. Her research focuses on building and developing ethical systems aimed to advance the environment. She is currently working on projects titled *"Accident Prediction and Smart Ambulance Routing Using AI"* and *"AgriGuard Blockchain System for Predicting Future Paddy Prices and Monitoring Warehouse Conditions."* Her previous project, *"Prediction of Healthy Tree and Leaf Conditions,"* explored computer vision techniques for environmental monitoring. Kaviya has participated in three hackathons and presented her blockchain-based smart warehousing research at a National Conference, where she won First Prize for innovation and technical excellence. Passionate about interdisciplinary learning, she continues to explore advanced applications of AI and Blockchain for creating intelligent, transparent, and sustainable digital ecosystems. She aims to pursue research that bridges emerging technologies with societal needs in agriculture, healthcare, and intelligent transportation systems.