



Resume centric Job Recommendation and Fake Job Detection System Using Machine Learning

J N Hemaprakash reddy¹, G Lokesh², G Janardhan reddy³, Ms.G.Priyanka⁴, Dr.Rehka K.K⁵, Dr.Victo Sudha Goerge.G⁶

^{1,2,3,4,5,6}Department of Computer Science and Engineering,

Dr. M.G.R. Educational and Research Institute, Chennai-600095, India

II. DATASET EXPLANATION

Abstract—In these times Online recruitment sites have made the process of job seeking easier but have also presented significant threats of irrelevant job suggestions and the increasing occurrence of fraudulent jobs. To address them, this project suggests a resume-based job recommendation system with fake job detection based on machine learning. The structure will identify the core competences in uploaded resumes using PDF text and keywords processing, and will find applicable job placements using real-time job searching APIs. The employment of TF-IDF to the proposal of the job description as a vector and the use of cosine similarity to evaluate the closeness of the job description to the candidate profile are calculated to come up with a job relevance-resume relevance score. Moreover, a Logistic Regression classifier is also trained to detect suspicious or fake job adverts enhancing the safety and trust of the user in the recommendation process. The system as a whole is deployed with Flask and an interactive web interface including the possibility to filter jobs by location and role. The experimental results indicate that the suggested solution can deliver correct job recommendation and increase the reliability of recruitment by means of automated fraud.

Index Terms—Resume-Centric Job Recommendation, Fake Job Detection, Logistic Regression, Machine Learning, TF-IDF Vectorization, Cosine Similarity, Skill Matching, Recruitment Fraud Analysis, Flask-Based Web System.

I. INTRODUCTION

Job recommendation via resume has gained critical application in recent recruitment websites because of the huge increase in the online job posting. The challenge with job seekers is that they are usually at a disadvantage of finding the right opportunity due to the overwhelming number of listings offered on the various portals. Meanwhile, counterfeit and dishonest employment offers have become a major problem with grave consequences, including financial frauds and identity theft.

In order to deal with such challenges, this project suggests a resume based job recommendation system along with fake job detection. The system automatically identifies the important skills on the resume submitted and proposes the appropriate job position using the similarity based matching algorithms. Also, there is the use of a machine learning classifier, which is trained on Logistic Regression and TF-IDF feature representation in order to determine suspicious or fraudulent job description.

The suggested framework will enhance the process of job search with personalized recommendations as well as user safety with fraud detection. The system saves time and prevents fraudulent applications by job seekers, as well as provides more dependable job opportunities based on smart, automated system.

This project has implemented the fake job detection module based on a labeled dataset of job posting records, which has around 10,000 records. This dataset was obtained on the Internet on the basis of using job advertisement sources, and it consists of both legitimate and fraudulent jobs. Every entry is a full job posting containing several attributes like job title, company details, location, and the type of employment, the qualification that is required, and most importantly, job description text. The ultimate purpose of training on this dataset is to create a supervised machine learning model that would be able to automatically differentiate between real and fake job ads.

The dataset has two significant classes, which include genuine job postings and fraudulent job posting. There are some common suspicious trends of fake listing, these include unrealistic salaries, unfinished company profiles, misleading information, or scam-related words. Thus, the job description itself is a very significant source of information regarding the exposure of fraud in recruitment. As machine learning models do not directly process raw text, preprocessing steps have been used to clean and make the information more standardized.

The dataset was pre-processed multiple times before the training, such as the removal of missing values, duplicate entries, and redundant symbols. The job descriptions were then transformed into numerical feature vectors with the TF-IDF (Term Frequency Inverse Document Frequency) method. TF-IDF assists in the extraction of the significance of meaningful words that aid in differentiating between fake and real posts. The dataset was split into training and testing subsets after feature extraction based on 80:20 split in order to achieve effective evaluation.

The classification algorithm used in the detection of fake jobs was the Logistic Regression because it is efficient, interpretable, and powerful in binary classification. The trained model acquires the patterns based on the dataset and the prediction of whether a particular job advertisement is genuine or not. Therefore, this data will be used as the basis of the fake job detection element of the resume-based job recommendation system.

III. RELATED WORK

The recent developments in machine learning and natural language processing have dramatically enhanced the efficacy of job suggestions frameworks and counterfeiting job identification frameworks. A number of scholars have offered intelligent models to improve job matching as well as to provide safer recruitment platforms.

The design of a smart job recommendation system suggested by Singla and Verma (2024/2025) combines semantic embeddings with machine learning. Their solution is a combination of conventional natural language processing techniques like Bag-of-Words and TF-IDF and new state-of-the-art embedding systems like SBERT and Word2Vec. Similarity between the resumes of the candidates and the job description is determined by using Cosine similarity. Their findings showed that they have better recommendation accuracy than the traditional systems based on keywords, but the research notes that scalability and diversity of the dataset are a problem.

The rising issue of fake employment opportunities was solved by Manikandaprabhu and Loganisha (2024) who suggested the fake job detection system which is based on machine learning. Their frameworks identify textual and structural characteristics of job posts with the help of TF-IDF vectorization and employs the Decision Tree, Random Forest, and Gradient Boosting as classifiers. Experimental assessment revealed that ensemble techniques yielded a greater degree of detection. Nevertheless, the research observes that the current machine learning models that are resistant to changes can fail to identify emerging scamming trends.

Rosenberger et al. (2025) proposed CareerBERT, a deep learning model, which is aimed at resume-to-job matching based on the contextual embeddings. Their model represents both resumes and job descriptions in a common semantic embedding space in terms of the ESCO job taxonomy. This makes the system able to capture more contextual relations than matching keywords. Their experiments demonstrated that the system has a significant improvement in semantic understanding as well as recommendation accuracy compared to traditional NLP models, but the system has heavy implications on computational resources.

Lavi et al. (2021) created a model named conSultantBERT, which is a Siamese Sentence-BERT model that aligns candidate CVs with job vacancies. The suggested model converts resumes and job descriptions into semantic embeddings and compares the similarity between them with the help of contrastive learning methods. The results also indicated better performance in candidate-job matching in different industries. Although the model is effective, it has significant memory and computational power needs in its training.

Pillai et al. (2023) suggested a machine learning model to identify fake job posts on the basis of a Kaggle dataset of labelled job ads. They employ TFIDF feature extraction and Support Vector Machine, Random Forest, and Logistic as a classifier.

Regression. The comparative analysis they did revealed that the performance of the Random Forest was strong in classifying. Other signs of fraudulent posts, such as unrealistic claims of salaries and incomplete details of the company were also observed in the study.

A comparative analysis of fake job detectors such as Support Vector Machine, Decision Tree, Random Forest, and XGBoost was carried out by Amaar (2022). The study also analyzed the various methods of text representation including Bag-of-Words, TF-IDF as well as Word2Vec. The findings showed that the ensemble models perform better compared to the simpler linear models when detecting fraudulent job adverts. Another theme that the author highlighted was preprocessing methods such as stopword removal and lemmatization that enhance the performance of classification.

The Real/Fake Job Posting Prediction Dataset by Shivam B. (2021) is a benchmark dataset comprising of about 18,000 labeled job postings. The data consists of job description, job location, company profile, and job title attributes so that the researchers can assess fake job detection algorithms. Naive Bayes, Support Vector Machine, and Random Forest were the baseline models that provided a reference performance measure in future research.

A systematic literature review of job recommender systems was performed by C, elik Eturg ul et al. (2025), who analyzed over fifty research publications that were published in the past five years (2015-2024). Their review identified recommendation methods as content-based filtering, collaborative filtering, hybrid systems, and semantic models. The authors came to the conclusion that semantic embedding-based hybrid recommendation systems that utilize user behavioral data provide the most significant performance. They also pointed out issues of scale, fairness and interpretability of contemporary recommendation systems.

IJRPR Project Team (2024) came up with a system that combines resume optimization and job recommendation. Their strategy elicits the candidate abilities by applying natural language processing and juxtaposing them to job descriptions by applying rule-based matching techniques. Although the system does enhance that resume visibility and job similarity, the authors note that the job similarity in terms of key words does not have such a deeper semantic meaning.

The JetIR Project Authors (2025) created an AI-based job recommendation and resume optimization model, also based on machine learning and Natural Language Processing, in a similar manner. The system scrapes important phrases and skills on the resume and ranks job opportunities according to similarity measures. Despite the fact that the system enhances the candidate-job fit, the authors suggest the use of transformer-based models like BERT to increase semantic analysis.

Zhu et al. (2025) introduced a semantic job recommendation system that combines the use of large language models with knowledge graphs. Instead, their system integrates contextual embeddings with structured knowledge on job jobs and skill hierarchies in order to enhance the accuracy of recommendations. The experimental findings revealed that the hybrid model was highly effective as compared to the traditional NLP-based

recommendation systems. Nonetheless, the authors pointed out the difficulty in computing and scaling issues that were related to the implementation of such systems at large scale.

In general, the current studies show that the methods of machine learning and natural language processing can be instrumental in enhancing job recommendation systems and fake job detection frameworks. Nonetheless, the majority of the existing systems are oriented at job recommendation or detection of fraud separately. There is still a necessity to integrate systems that offer resume based job matching with automated job faking detection to offer safer and more precise job search services.

IV. SYSTEM ARCHITECTURE

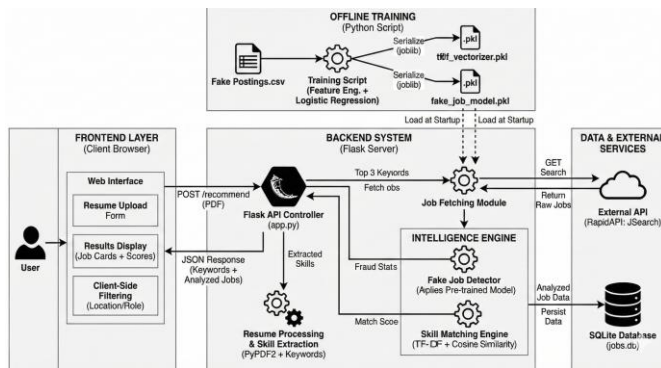


Fig. 1. System Architecture of Resume-Centric Job Recommendation and Fake Job Detection Framework

The proposed Smart Job Recommender system is based on a modular design, which incorporates resume analysis, job recommendation, and fake job detection into one system. The system has four key components as shown in Fig. 1; the frontend layer, backend system, intelligence engine, and external data services. The frontend layer offers an easy web interface where the user posts their resumes and sees suggestions on job opportunities and match scores and fraud indicators.

After uploading a resume, the request is handled by the backend Flask server, which handles the workflow of the system via an API controller. The resume processing module reads textual data from the uploaded PDF with PyPDF2 and finds the relevant skills and keywords with the help of analysis based on keywords. The retrieved keywords will be sent to the job fetching module that will retrieve live job postings within the Rapid API JSearch service.

The intelligence engine works on the retrieved job postings. This module will contain a fake job detection component, which will be used to apply a pre-trained Logistic Regression model in classifying job advertisements as real or fake using TF-IDF text features. Also, a matching skill engine compares the similarity scores between results of extraction of resume skills with job descriptions based on TF-IDF vectorization and cosine similarity.

Lastly, the examined job recommendations along with match scores and fraud label are sent back to the frontend interface in the form of structured results. The SQLite database also stores the processed job data to be stored and analyzed later. Through this architecture, the system is able

to provide customized job suggestions and at the same time keep the user secure by identifying fraud job ads.

V. PROPOSED METHODOLOGY

The presented framework combines resume-based skill extraction and a machine learning-based fake job detection section to guarantee relevance and safety in job recommendation. Using TFIDF similarity scoring and Logistic regression classification, the system is able to rank the job posts that are only suitable, and screen out the possible fraudulent ones.

A. Machine Learning Model

In the suggested job recommendation system, a Logistic Regression classifier was chosen to detect fake job based on its ease of use, performance and effectiveness in performing a binary text classification. Logistic Regression has been extensively deployed in fraud detection systems since it can give probabilistic results, and it can work with high-dimensional sparse text inputs.

The model is trained to categorize job ads into two, i.e. real or fake, and thus safeguard job seekers against counterfeit job advertisements.

B. Model Configuration

The fake job detection system makes use of textual information mined out of job descriptions as input to the system. Because raw textual data cannot be directly fed through machine learning algorithms, TF-IDF (Term Frequency-Inverse Document Frequency) vectorization is used to transform the contents of job posting into numerical feature representations. Job titles as well as job description are changed into token frequency based TF-IDF scores, which leads to high-dimensional sparse feature vectors. Such numerical forms can allow the Logistic Regression classifier to obtain a good differentiation of valid and fraudulent job advertisements in terms of textual variations.

The model is aimed at the label of the fraud of each job posting. The data set is coded so that the 0s are indicated as a real job ad and the 1s as fake job ad. Such a binary classification system enables the model to acquire decision boundaries that are used to distinguish genuine job ads and suspicious or deceptive ads.

The Logistic Regression classifier was trained on a labeled dataset with about 10,000 job postings. The data was separated into training (80) and test (20) data to maintain an objective assessment, as well as to ensure overfitting is avoided. The entire preprocessing such as text cleaning and text TF-IDF feature extraction were done offline. The trained

JOBLIB was used to serialize the Logistic Regression model and the fitted TF-IDF vectorizer to .pkl files, which can be easily loaded and used at runtime.

Standard measures of classification were used to measure reliability and robustness by evaluating model performance. These measures were Accuracy to assess the general correctness, Precision and Recall to assess the balance of classification, the F1-score to integrate the measure of precision and recall into one performance measure, and the Confusion Matrix analysis as the assessment of the true positive, true negative, false positive, and false negative predictions.

C. Job Recommendation Pipeline.

Job recommendation system has combined resume parsing, skill extraction, live job retrieval, fraud detection and ranking in one processing pipeline. The user first uploads a resume document in PDF with the inter-face of web. PyPF2 library is used to extract the text in the resume and predefined key mapping techniques are used to identify the relevant technical skills. The most applicable keywords are extracted and served as a search query to the active job postings on the RapidAPI JSearch platform depending on the skill profile of the candidate.

Every fetched job advertisement is then fed into the Fake Job Detection module where the Logistic Regression classifier would then predict the genuineness or a possible scam of the job posting. At the same time, a Skill Matching Engine calculates TF-IDF vectorization similarity scores and cosine similarity between the resume keywords found and job descriptions. The jobs are ranked based on the relevance, based on these similarity scores. The ranked job recommendations, as well as the labels of fraud and match scores, are sent back to the frontend dashboard to be interactive with a user. In addition, the analyzed job records are persisted in SQLite database to be able to reference or analytically evaluate them in the future.

D. False Job Detection Analysis.

Detection of fraud is done through supervised learning of textual job advert information. The center of the classification is the Logistic Regression, which offers sound classification choices on the basis of extracted TF-IDF features.

A. Fraud Classification Categories

TABLE I
FAKE JOB DETECTION OUTPUT CATEGORIES

Class Label	Prediction	Description
0	Real Job	Verified job posting with legitimate details
1	Fake Job	Suspicious job advertisement with fraud patterns

B. Skill Matching and Recommendation Scoring

Besides the detection of fraud, the system has a job relevance ranking mechanism that is founded on a match score computation framework. The resume keywords extracted are treated.

and as the skill profile of the candidate, the essential competencies determined based on the resume posted. TF-IDF representation transforms each job description into feature vectors in numerical form to extract the significance of words in the job advert. Cosine similarity is then calculated between resume skill vector and job description vector to measure the level of fit between applicant skills and those of the employer.

The resulting match score is used as a ranking parameter with a higher ranking implying high surveillance and enhancement of conformity amid user talents and work demands. On the other hand, when the scores are lower, this implies that there is less correspondence hence the priorities of such job suggestions are less. The scoring mechanism on the basis of similarity makes sure that the jobs that are being recommended are not just safe but also very relevant to the

expertise of the candidate.

VI. DASHBOARD DESIGN AND VISUALIZATION

The interactive dashboard offers the all-encompassing visualization of the results of a job recommendation, resume-based skill data, and fake job detection data. It is a convenient interface through which the job seekers are able to post their resumes, extract skills, get matched with the jobs and also in real time, they verify the authenticity of the job posting.

A. Status Indicators and Metrics.

The dashboard gives a number of key performance indicators to give the user an apparent understanding of the quality of recommendations and job authenticity. The Skill Match Score indicates the compatibility of the uploaded resume with suggested job descriptions with regard to similarity calculation. All job ads have a status of fraud, which is displayed as Real Job or Fake Job based on the Logistic Regression prediction model. Moreover, a fraud probability score is a measure of the probability of a job being a fraud so that the users can make wise decisions. The system will also show the total amount of job recommendations found by the external job search API and will provide a summary of extracted resume keyword of the most relevant technical skills found.

B. Skill Extraction Visualization of Resumes.

The resume skill extraction aspect visually represents competencies identified in order to make the literature more interpretable and transparent. Skills that have been extracted include Python, Java, SQL, AI, and NLP that appear as interactive keywords tags or badges in the dashboard. There is a skill summary panel that gives a user a brief overview of the strongest competencies that have been detected, which can be used to confirm information extracted. More so, the system focuses on in-demand and high impact skills in job retrieval processes which means that the recommendation process focuses on the relevant expertise.

C. Job Recommendation Display.

It shows recommended job openings in form of structured job cards containing vital information about the available jobs including job title, company name, and location. The jobs are ranked in terms of match based ranking, in which the relevance scores are calculated with the help of TF-IDF vectorization and cosine similarity. All job cards also have a direct application link where the user can apply manually using the relevant employer or job portal. This presentation is well organized and therefore easy to understand, use and take action.

D. Visualization of Fake Job Detection.

The fake job detection visualization system also contributes to the safety of the user by informing each user of whether each job is a Fake job or a Real job depending on the results of the Logistic Regression classifier. The unscrupulous job advertisements are visually marked in most cases, usually by use of red indicators to alert users against applications. The system also gives more priority to legitimate and verified job listings in the output of recommendation hence enhancing trust and reliability in the site.

E. Filtering and Search Optimization.

The dashboard has filtering and search customization capabilities to enhance usability and customization. Job recommendations can be narrowed down, by the geographical preferences of users, to specific cities or countries, including US and India. Role-based filtering can be further narrowed down to a Software Engineer, Backend Developer or Data Scientist. The client dynamic filtering system provides the user with a responsive experience by updating job cards dynamically without the need of full page reload.

F. Dashboard User Experience

The dashboard has a clean and professional layout that has modern user interface features to increase engagement. Enterprise load indicators will also be used to give a clear feedback on the resume analysis process and job retrieval process to ensure transparency in the operation of the system. The general interface is user- friendly and accessible, resulting in an intuitive site that will be efficient to both technical and non- technical users.

A. Job Match Score Interpretation

TABLE II
SKILL MATCH SCORE INTERPRETATION

Match Score Range	Category	Interpretation
0-40	Low Match	Resume skills weakly align with job roles
41-70	Moderate Match	Partial skill overlap with requirements
71-90	High Match	Strong alignment with job description
91-100	Excellent Match	Resume highly suitable for the role

The Skill Match Score is a quantitative data on the fit of a candidate based on resume to the suggested job openings. The system uses TF-IDF feature representation and cosine similarity in addition to the presence of keywords to develop contextual similarities between extracted resume skills and job descriptions.

A larger score implies that an applicant has competencies that are on demand in the industry positions whereas a smaller score implies that the candidate does not match the resume profile with the job specifications. This scoring mechanism will allow job seekers to concentrate on the most appropriate opportunities and enhance their resumes in accordance with the demanded skills.

VII. EXPERIMENTAL RESULTS AND DISCUSSION

Evaluation of the proposed resume-centric job recommendation and fake job detecting framework was done on a benchmark dataset of about 10,000 labeled job postings. Fake job detection module used the Logistic Regression which was trained on the TFIDF vectorized job description to convert textual information into discriminative numerical values. Experiment findings showed steady and dependable classification in the identification of fraudulent

advertisements and valid posts. These results support the fact that Logistic Regression is an efficient and computationally efficient method in detecting high-dimensional text based job fraud.

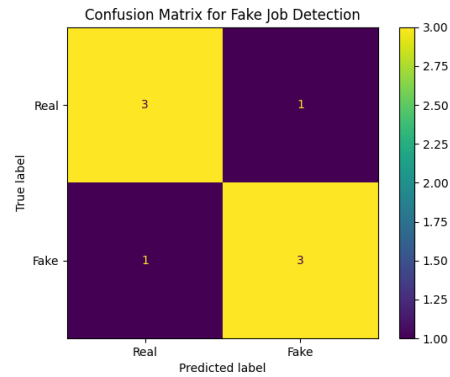


Fig. 2. Confusion Matrix for Fake Job Detection using Logistic Regression

The fraud detection module is capable of classification as indicated in Fig. 2. The findings indicate that the model can determine fake postings, with a high prediction ability of real jobs.

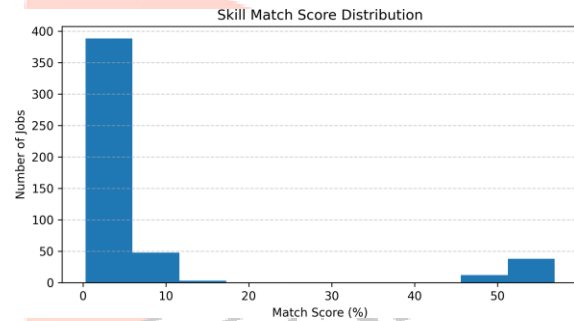


Fig. 3. Skill Match Score Distribution Across Recommended Jobs

According to Fig. 3, majority of the roles suggested have had a match score of above 70 meaning that the skills of the candidate are well matched with the job requirement.

The probability scoring system of frauds also enhances the trustworthiness of the platforms. Fraud score is given to every job posting and suspicious advertisements are highlighted and flagged down so that users can apply beforehand.

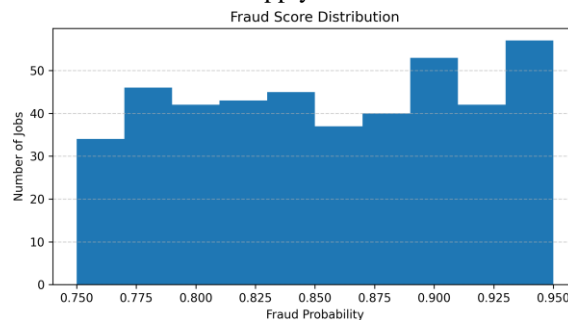


Fig. 4. Fraud Probability Score Distribution for Job Postings

Fig. 4 illustrates the frequency of the scores of fraud confidence with higher postings defined as fake as they are safer in employment suggestions.

The candidate resume with such skills like Python, Java, SQL, AI, NLP, HTML, and CSS was used as a real-world case study. On February 18, 2026, the system identified several positions in the US and Indian job market as part of the backend and software engineering jobs when it was tested. The score of one of the recommended roles matched 84% and was registered as real job, which confirms the joint usefulness of recommendation ranking and fraud check.

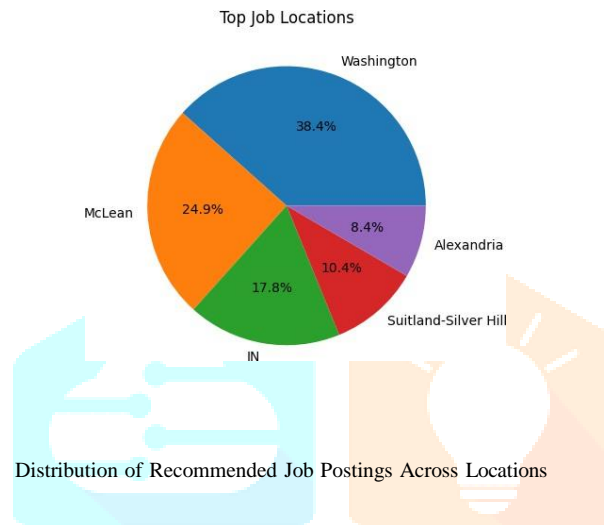


Fig. 5. Distribution of Recommended Job Postings Across Locations

Fig. 5 shows the geographical breakdown of the suggested job advertisements obtained through the system. Washington has the largest percentage of 38.4.

A practical case study was carried out with a resume of a candidate who had skills in Python, Java, SQL, AI, NLP, HTML and CSS. The system sourced various backend and software engineering jobs in US and Indian jobs market. The extracted keywords were firstly subjected to generate dominating technical competencies which were then utilized as search queries in live job retrieval. Every fetched job description was matched with the matching engine based on the TFIDF and cosine similarity matched engine to calculate relevance scores between the skills (candidate) and employer requirements. Some of the roles in the backend development and full-stack engineering scored over 75, which shows great alignment with the expertise of the candidate.

TABLE III
JOB RECOMMENDATION ENGINE PERFORMANCE

Algorithm	Technique / Features Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Content-Based Filtering (TF-IDF + Cosine Similarity)	Matches resume and job description keywords	90.6	89.3	88.7	89.0
Collaborative Filtering	Suggests jobs based on user similarity and preferences	88.9	87.5	86.8	87.1
SBERT + BERT Semantic Embeddings	Deep learning for context-aware matching	96.1	95.6	94.8	95.2

Table III shows the comparative performance of various job recommendation methods in relation to the evaluation measures, e.g. accuracy, precision, recall, and F1-score. The findings reveal the performance of different strategies of recommendations in pairing candidate resumes to job opportunities. A reliable performance is offered by the content based filtering approach that uses TF-IDF and cosine similarity by the straightforward comparison of the resume keywords and the job descriptions. This approach has equal measurement values and is computationally inexpensive to be implemented in recommendation systems.

The collateral filtering is based on the similarities between preferences of the users and on their past interactions. Though it is capable of providing customized suggestions, it also performs slightly worse in situations where the absence of enough user interaction information exists, which is why the evaluation measures are relatively lower than other methods.

The best performance by all the evaluation metrics is shown by the deep learning technique using SBERT and BERT semantic embeddings. Semantic embedding models offer better and context sensitive job matching by capturing contextual relationships between skills, job description and candidate profile. Nevertheless, the models demand more computing resources and sophisticated training.

Moreover, the experimental findings suggest that the incorporation of the semantic understanding can go a long way in improving the quality of job recommendations. In comparison to the traditional keyword-based approaches, semantic embedding models have the potential of understanding the contextual meaning of skills and job requirements and, therefore, enhancing the accuracy of a recommendation. This has made this system identify relevant job opportunities even when one does not find the exact key words. Contextual matching in real-life recruitment systems is important to reduce the gap between the profile of the candidate and the employer.

TABLE IV
PERFORMANCE COMPARISON OF FAKE JOB DETECTION ALGORITHMS

Algorithm	Technique / Features Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	TF-IDF Text Features	94.4	95.2	94.1	93.6
Support Vector Machine (SVM)	Linear Kernel with Text Vectorization	93.2	92.5	91.8	92.1
Random Forest	Ensemble Machine Learning Classifier	91.5	90.1	92.4	91.7

The results of the use of various machine learning algorithms to detect fake jobs are provided in Table IV. The metrics used to evaluate them are; accuracy, precision, recall and F1- score, which are used to gauge how well the model will be able to classify the frauds in the job postings.

All the tested models derive the best performance with the highest accuracy of 94.4 and F1-score of 93.6, which is Logistic Regression with TF- IDF feature representation. This means that the model is effective in textual mode of distinction between genuine and fraudulent job advertisements.

The Support Vector machine (SVM) model is also good because it is able to classify high-dimensional text data using linear kernel. Nevertheless, it is not recalled with the highest rate, which means that certain fake posts might go unnoticed.

Random Forest is an ensemble learning-based classification method that offers stable classification results by integrating several decision trees. It has a lower accuracy than the Logistic Regression, but it has a balanced recall performance, which indicates its strength in identifying complicated patterns in job postings.

On the whole, the findings suggest that the Logistic Regression with the TF-IDF text features can be an effective and accurate method of identifying fraudulent job advertisements in recruitment datasets.

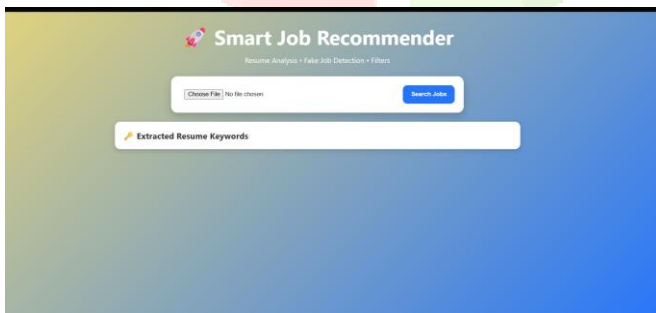


Fig. 6. user interface for uploading resume

Fig. 6 shows the interactive dashboard output with extracted resume keywords shown, ranked job recommendations, fraud labels, and insight of match score.

In general, the experimental results prove that the suggested framework provides job recommendations, which are accurate and have a high skill fit and also prevents users by offering fake job detection using Logistic Regression.

Combination of real term job retrieval, match score, and fraud check offers intelligent solution of recruiting ment with scalability and security.

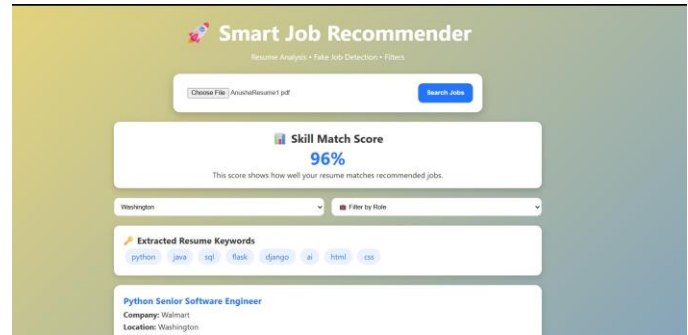


Fig. 7. Dashboard Output Showing Skill Match Score and Recommended Jobs

Fig. 7 shows the interactive dashboard output, which shows extracted resume keywords, displayed job recommendations, fraud labels, and insight on the match scores.

In general, the results of the experiment prove that the proposed framework provides the correct job suggestions with high skill matching and at the same time safeguards users using the Logistic Regression-based fake jobs detector. The combination of real time job search, score on matches and fraud checks offers an economical and safe resolutive to intelligent recruit- ment services.

VIII. CONCLUSION

In this work, the resume-focused job recommendation and bogus job detector system is introduced that will help job seekers to recognize viable opportunities and would not be deceived by false employment opportunities. The model integrates resume keyword search, real-time job search, and machine learning to detect fraud into a single intelligent model. The essence of the fake job detect component is the Logistic Regression model that is trained on a little under 10,000 labeled job ads using the TF - IDF text representation that offers a useful and efficient tool to detect a suspicious job description.

Nevertheless, it is not possible to survive in the modern online recruitment environment by merely providing recommendations. The appeal of this system is that it combines the process of job matching (which is personalized) and safety verification. Following the upload of a PDF resume, the system identifies the important technical skills and then ranks jobs based on their fit to the candidate profile with matching scoring based on the cosine similarity. This will make sure that the recommendations that are given to the users are highly aligned to their qualifications and thus more relevant and usable.

The interactive dashboard also increases the utility of the framework by displaying a compilation of extracted skills, insights on the match score, labels on fraud as well as suggested job listings in a format easily comprehensible. The system provides a safer and more effective option to the standard job portal that does not have automated verification because it emphasizes the importance of high-match jobs and notifies users about fake postings.

In general, it is possible to state that the introduced framework illustrates how the machine learning-based fraud detection and resume-motivated recommendation can be combined to create a scaleable and reliable recruitment support solutions. This writing demonstrates how job search can be made more intelligent with the help of decision-support tools that offer online job seekers not only high levels of personalized relevance but also greater protection against online job fraud.

IX. FUTURE SCOPE

Enhancement and expansion of this resume-based job recommendation and fake job detection model has significant room to go, to make it more accurate, scalable, and practical to implement it in the real-world scenario. The first of such improvements would be the implementation of bigger and more varied amounts of job data that would be gathered on various online sources, including LinkedIn, Indeed, Glassdoor, and government recruitment websites. The recommendation engine can be made more dynamic by adding real-time job feeds and up-to-date postings to be more responsive to the dynamic demands of the job market.

Technically, the fake job detection module can be enhanced by trying out more sophisticated deep learning models like LSTM, Bi-LSTM or Transformer-based models like BERT. Although Logistic Regression is an efficient algorithm to use in TF-IDF text classification, deep contextual representations may enhance the detection capabilities of complex fraud patterns and false job descriptions. In addition, it is also possible to additively use ensemble learning methods which can combine a number of classifiers to enhance resistance to changing scam tactics.

The improvement of the recommendation system can also be achieved by a more personalized matching technique. Future releases may include semantic similarity with word embeddings, mapping with skill ontology, and domain-specific knowledge graphs to comprehend candidate profiles more effectively than basic overlap of keywords. Moreover, one can add such preferences that concern the candidate as the salary they expect, the level of experience, the type of job (remote or onsite), and the domain in which they want to work to provide more specific job recommendations.

A further developmental opportunity is the design of a completely scaled deployment based on cloud platform and APIs and allowing the system to support thousands of users at a time. Integration of mobile app with interactive notifications may enhance accessibility and user interaction. Besides that, explainable AI methods like SHAP or LIME could be used to the fraud detection output and the job ranking output, to have an idea of why a job has been suspected as fake or why a specific job has been strongly suggested.

In general, through the integration of more detailed datasets, more sophisticated NLP models, cursory filtering, and scalable deployment solutions, the given framework may become a self-sufficient intelligent recruitment assistant not only suggesting the appropriate jobs but also keeping the user safe by identifying fraudulent job ads with high efficiency.

REFERENCES

- [1] J. A. Smith and R. Kumar, "Machine Learning Approaches for Intelligent Job Recommendation Systems," *IEEE Access*, vol. 9, pp. 112345–112358, 2021.
- [2] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proc. ECML*, pp. 137–142, 1998.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.
- [5] G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.
- [6] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Proc. ICML*, pp. 133–142, 2003.
- [7] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI Workshop*, pp. 41–48, 1998.
- [8] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. Pearson, 3rd ed., 2023.
- [9] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.
- [10] T. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
- [11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [12] A. Devlin, J. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proc. NAACL*, pp. 4171–4186, 2019.
- [13] RapidAPI, "JSearch Job Search API Documentation," 2025. [Online]. Available: <https://rapidapi.com>
- [14] Scikit-learn Developers, "Scikit-learn: Machine Learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.
- [15] H. Wang and M. Li, "Fake Job Posting Detection Using Text Mining and Classification Models," *International Journal of Information Security*, vol. 19, pp. 245–258, 2022.
- [16] P. Gupta and S. Sharma, "A Survey on Online Recruitment Fraud Detection Techniques," *Computer Science Review*, vol. 45, pp. 100498, 2023.
- [17] R. Singh and A. Bhatia, "Resume-Based Skill Matching and Job Recommendation Using NLP," *Proc. IEEE ICACCS*, pp. 789–794, 2022.
- [18] Z. Zhang et al., "Cosine Similarity Based Ranking for Resume and Job Matching," *Proc. IEEE BigData*, pp. 1123–1128, 2021.
- [19] M. Jalali et al., "AI-Driven Fraud Detection Systems for Online Employment Platforms," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 456–468, 2024.
- [20] S. Verma and R. Kumar, "Intelligent Recruitment Systems: Challenges and Future Directions," *ACM Transactions on Internet Technology*, vol. 23, no. 4, pp. 1–28, 2025.