



A Comparative Study Of Machine Learning Classifiers For Fake Job Posting Detection

Bhanushali Shiv G, Dr. P. Srinivasa Rao

Department of Computer Science and Engineering Jain (Deemed-to-be University)
Bengaluru, India

Abstract

Online job portals have grown rapidly over the past few years, making it easier for companies to reach a large number of applicants. However, this growth has also led to an increase in fraudulent job postings, which can result in financial loss and privacy risks for job seekers while also reducing trust in genuine recruiters. Identifying such postings manually becomes difficult and inconsistent, especially when the number of listings is high. Recent studies have explored the use of machine learning techniques to detect fraudulent job postings from textual data [1]–[3]. In this work, different machine learning classifiers are tested and compared using a publicly available Kaggle dataset that includes job descriptions and related attributes [4]. The textual data is converted into numerical form using a TF-IDF representation, after which a combination of linear, tree-based, and ensemble models is trained and evaluated under a consistent experimental setup to ensure fair comparison [5], [6]. Due to the imbalanced nature of the dataset, the evaluation focuses on precision, recall, and F1-score rather than accuracy alone. The results show that ensemble and boosting-based models perform better, particularly in terms of recall and F1-score for the minority class [7], [8], indicating that these approaches are more suitable for real-world fraud detection scenarios where missing a fraudulent posting can have serious consequences.

Index Terms -

Fake Job Posting Detection, Machine Learning, TF-IDF, Ensemble Learning

1. Introduction

In recent years, online recruitment platforms have become one of the main ways to publish and access job opportunities. The growth of web-based hiring portals allows organizations to reach a large number of applicants efficiently. However, this increased accessibility has also led to a rise in fraudulent job postings. These postings are often designed to closely resemble legitimate advertisements, which makes them difficult for job seekers to identify through simple inspection.

Fake job listings commonly involve scams that request money, personal documents, or other sensitive information, exposing individuals to financial and privacy risks. Such activities not only harm job seekers but also reduce trust in online recruitment platforms and affect the reputation of genuine employers. While manual identification of suspicious postings may work at a small scale, the large volume of listings makes reliable verification difficult.

Recent studies have explored machine learning techniques to detect fraudulent job postings by analyzing textual data and identifying hidden patterns [1]–[3]. At the same time, fraudsters are becoming more sophisticated, often designing job descriptions to closely match the structure, tone, and content of genuine postings. As a result, basic rule-based or surface-level filtering methods are often not sufficient. This has led to the use of computational approaches that can capture deeper linguistic and structural differences between real and fake job advertisements.

Machine learning-based text classification provides a scalable solution for automated fake job detection [6], [7]. By analyzing textual content from job descriptions and related attributes, these models can identify patterns that are not easily noticeable through manual review. This helps improve detection compared to manual screening alone. In this work, multiple classical and ensemble machine learning classifiers are evaluated using features derived from job-related text data.

Since fraud detection datasets are typically imbalanced, this study focuses more on precision, recall, and F1-score rather than relying mainly on accuracy. This is mainly because accuracy alone does not reflect how well the model detects fraudulent postings. The aim of this study is to compare different classifiers under a consistent experimental setup and identify models that perform well for the minority class. In this context, higher recall helps reduce the chances of fraudulent postings going undetected, thereby limiting potential risks to users. Overall, the findings are intended to support practical screening mechanisms for online recruitment platforms.

The main contributions of this study are as follows:

- A comparative evaluation of classical and ensemble machine learning classifiers under severe class imbalance conditions.
- A unified preprocessing and TF-IDF-based framework to ensure consistent evaluation across models.
- An analysis of precision–recall trade-offs for fake job posting detection in a realistic setting.

2. Methodology

2.1. Dataset Description

The experiments in this study are carried out using a publicly available fake job posting dataset obtained from online recruitment platforms [4]. The dataset includes 17,880 job postings with 18 attributes, where each entry is labeled as either legitimate or fake. Out of these, 17,014 postings correspond to real jobs and 866 are marked as fake, resulting in a highly imbalanced class distribution.

The dataset contains both textual and structured attributes. Textual fields such as job description, company profile, requirements, benefits, and job title act as the main sources of information for classification. In addition, categorical and numerical attributes related to employment type, industry, and organizational details are also included [4].

From the initial analysis, it can be observed that several non-textual attributes, including salary range, department, and required education, contain a large number of missing values. In contrast, core textual fields, especially the job description, are mostly complete and therefore form the primary basis for feature extraction. The target variable, labeled as fake, is binary and is used to distinguish between legitimate and fraudulent postings during model training and evaluation.

Since the dataset is highly imbalanced, model performance is evaluated mainly using precision, recall, and F1-score, as these metrics better capture minority-class detection performance. The results reported in this study are based on a single 80–20 train–test split. Cross-validation is not included, and further statistical stability analysis is considered for future work.

2.2. Data Characteristics

The dataset has several characteristics that influence the design and evaluation of fake job posting detection models. One of the main challenges is the severe class imbalance, as fake postings make up only a small portion of the data. Even though they are fewer in number, fraudulent listings pose a higher

practical risk, making their detection more important. For this reason, evaluation metrics such as precision, recall, and F1-score provide a more meaningful assessment of performance than accuracy alone [1], [2].

Another important aspect is the uneven availability of structured information across job postings. Many non-textual attributes contain a high number of missing values, while textual fields remain largely complete. This suggests that fraudulent postings may omit structured or optional details while keeping descriptive text to appear legitimate. To preserve dataset size and reflect real-world conditions, records were retained and handled during preprocessing rather than removed [1].

Based on these observations, textual data is treated as the most consistent and reliable source for classification. Patterns within job descriptions, company profiles, and requirement sections can capture subtle indicators of fraudulent behavior. Focusing on text-based features also improves applicability across different job categories while relying on the most stable attributes in the dataset. This supports the use of TF-IDF representations with unigram features, default normalization, and no manual feature selection [1], [2].

2.3. Data Preprocessing

Before training the models, several preprocessing steps were applied to ensure consistency and suitability for machine learning classification. These steps help prepare the data for more effective model training. Textual fields with missing values were replaced with empty strings so that all samples could be retained without reducing the dataset size [1].

The text data was standardized by converting all content to lowercase and removing non-informative characters that do not contribute meaningful semantic value. Relevant text fields for each job posting were then combined to form a single unified textual representation for every instance [1].

Feature extraction was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) approach, which converts textual data into numerical vectors by assigning weights to terms based on their frequency within a document and across the dataset [3]. This helps highlight important terms while reducing the impact of very common words. The TF-IDF representation was constructed using unigram features with default normalization settings, without applying task-specific vocabulary filtering or manual feature selection. Although the dataset is imbalanced, no resampling techniques were applied. Preserving the original class distribution allows evaluation under more realistic conditions and avoids introducing artificial patterns that may affect interpretability. Instead, model performance is evaluated using precision, recall, and F1-score, which better capture minority-class detection performance in imbalanced settings [1].

2.4. Feature Extraction

To convert textual data into numerical form suitable for machine learning models, the Term Frequency–Inverse Document Frequency (TF-IDF) technique is used. TF-IDF assigns weights to words based on their importance within individual job postings while reducing the influence of terms that frequently appear across the dataset and provide limited discriminative value. This helps emphasize more meaningful terms in the text.

The TF-IDF representation transforms the preprocessed text into high-dimensional feature vectors that capture term relevance across the corpus. These vectors provide a useful representation for text classification tasks and allow the model to focus more on meaningful terms rather than very common ones. A standard implementation with unigram features and default normalization settings is applied, without using task-specific vocabulary filtering or manual weighting adjustments.

Due to its simplicity, interpretability, and computational efficiency, TF-IDF remains a widely used baseline in text classification and provides a consistent foundation for comparing traditional machine learning models [3].

$$TF-IDF(t, d) = TF(t, d) \times \log \frac{N}{DF(t)} \quad (1)$$

where $TF(t, d)$ represents the frequency of term t in document d , $DF(t)$ is the number of documents containing term t , and N denotes the total number of documents in the dataset.

The resulting TF-IDF vectors are used as the primary textual features for all classifiers evaluated in this study. This representation provides a reliable baseline for comparing different models.

2.5. Machine Learning Models

In this study, multiple machine learning classifiers are evaluated to analyze their performance in detecting fake job postings. Both traditional linear models and ensemble-based approaches are included to allow a structured comparison under consistent experimental conditions [1], [2].

The classifiers considered include Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting, XGBoost, and LightGBM. These models are selected because they represent a mix of simple and more complex approaches commonly used in text classification tasks.

In addition, a stacking ensemble model is implemented by combining selected base classifiers to examine whether aggregating predictions leads to more stable precision–recall performance.

All models are trained using the same dataset, TF-IDF feature representation, and preprocessing pipeline to maintain consistency and ensure fair comparison across experiments [1], [3], [4].

2.6. Experimental Setup

All experiments were carried out using the Google Colab environment, which is a cloud-based platform for implementing and evaluating machine learning workflows. The same publicly available fake job posting dataset described earlier is used throughout the study to maintain consistency across all experimental stages [4].

The dataset is divided into training and testing sets using a fixed 80–20 split to evaluate model generalization. The same preprocessing steps and TF-IDF feature extraction process are applied to all classifiers to ensure uniform experimental conditions and allow fair comparison.

The models are implemented using standard Python libraries commonly used in text classification tasks. Each classifier is trained on the training data and evaluated on the test set using multiple performance metrics. Since the dataset is imbalanced, the evaluation focuses more on precision, recall, and F1-score, while accuracy is treated as a secondary metric.

A fixed random seed is used during dataset splitting and model training to improve reproducibility. This helps reduce variability and ensures that performance differences are mainly due to the models rather than changes in the setup.

3. Related Work

Recent research on fake job posting detection has explored machine learning and transformer-based approaches using textual features extracted from job advertisements [1], [2]. These methods show that textual data can provide useful signals for identifying fraudulent postings. However, performance still depends on dataset characteristics and how effectively the models handle class imbalance.

Earlier studies mainly relied on traditional machine learning techniques such as Logistic Regression, Support Vector Machines, and Naïve Bayes due to their simplicity and interpretability. These approaches typically use text representation methods like Bag-of-Words or Term Frequency–Inverse Document Frequency (TF-IDF) to capture patterns in job descriptions. While effective to some extent, their performance is often limited by severe class imbalance and the restricted expressive capability of linear models [3].

More recent work has explored ensemble and boosting-based methods to improve detection performance under imbalanced conditions. Techniques such as Random Forest, Gradient Boosting, XGBoost, and Light-GBM combine multiple learners and focus on previously misclassified samples during training. These models are well suited for high-dimensional TF-IDF feature spaces and can capture non-linear relationships that simpler models may miss. As a result, ensemble-based approaches have gained attention in fraud detection tasks involving unstructured textual data [1], [2].

Despite these advancements, direct comparison across studies remains challenging due to differences in datasets, preprocessing methods, evaluation metrics, and experimental setups. Performance is often reported mainly in terms of accuracy, which can be misleading for highly imbalanced datasets. This makes it difficult to objectively compare model performance across different studies. To address this, the present

study conducts a structured comparison of classical and ensemble-based classifiers using a common dataset, preprocessing pipeline, feature representation, and evaluation strategy, with a focus on metrics that better reflect minority-class detection performance.

4. Results and Discussion

This section presents the results of the comparative evaluation of multiple machine learning classifiers for fake job posting detection. Model performance is measured using accuracy, precision, recall, and F1-score, with greater emphasis on recall and F1-score due to the imbalanced class distribution. The results are presented in both tabular and visual form to allow clear comparison across different models.

Baseline classifiers such as Logistic Regression and Support Vector Machine show stable and interpretable performance, making them useful as reference models for text classification tasks. While these models achieve high precision, their ability to detect the minority class varies depending on how they handle the imbalanced data.

Tree-based models, including Random Forest and Gradient Boosting, demonstrate improved minority-class recall and F1-score compared to linear models. This suggests that non-linear approaches are better able to capture complex patterns within the high-dimensional TF-IDF feature space. This behavior is generally consistent across most of the evaluated models.

Boosting-based methods such as XGBoost and LightGBM show similar or slightly better performance in terms of recall and F1-score. The stacking ensemble model achieves a more balanced precision–recall trade-off, indicating that combining predictions from multiple classifiers can help improve overall stability under class imbalance. This leads to more reliable performance across different scenarios.

The comparison of evaluation metrics shows that no single model performs best across all criteria. Instead, each model has different strengths depending on whether precision or recall is prioritized. Overall, ensemble-based approaches provide more consistent performance for detecting the minority class, while simpler baseline models remain useful due to their stability and lower computational cost.

Table 1: Classification performance of evaluated machine learning models

Model	Precision	Recall	F1-Score	Accuracy
Logistic Regression	1.0000	0.0116	0.0229	0.9522
Support Vector Machine	1.0000	0.4682	0.6378	0.9743
Random Forest	0.9848	0.3757	0.5439	0.9695
Gradient Boosting	0.9744	0.4393	0.6056	0.9723
XGBoost	0.9492	0.6474	0.7698	0.9813
LightGBM	0.9758	0.6994	0.8148	0.9846
Stacking Ensemble	0.9453	0.6994	0.8040	0.9835

Accuracy is reported for completeness but is not considered the primary evaluation metric due to the severe class imbalance in the dataset. Table 1 summarizes the performance of the evaluated classifiers in terms of precision, recall, F1-score, and accuracy. This provides a clearer comparison across different models and shows clear differences in model behavior, indicating that the choice of classifier directly affects fake job detection performance.

Given the imbalanced distribution, no single metric can fully represent overall performance. Therefore, a combined evaluation using precision, recall, and F1-score provides a more reliable assessment than accuracy alone. Recall becomes especially important, as false negatives correspond to fraudulent postings that remain undetected. Although Logistic Regression and Support Vector Machine achieve high accuracy, their recall for the minority class is significantly lower, which limits their effectiveness in detecting fake postings.

The perfect precision observed for Logistic Regression reflects a conservative prediction strategy, where the model identifies very few instances as fraudulent. This reduces false positives and results in a precision of 1.0000, but at the cost of extremely low recall. This highlights the limitation of relying only on accuracy or precision in imbalanced classification problems. This trade-off becomes more noticeable when detecting rare classes.

Tree-based and boosting models show improved minority-class recall and higher F1-scores under the same conditions. Among the evaluated methods, LightGBM and the stacking ensemble achieve relatively better F1-scores, indicating a more balanced trade-off between precision and recall. Accuracy remains similar across models and is therefore treated as a secondary metric. From a practical perspective, higher recall is often preferred, as it reduces the risk of fraudulent postings going undetected, even if it leads to a slight increase in false positives. This trade-off is generally acceptable in real-world fraud detection scenarios.

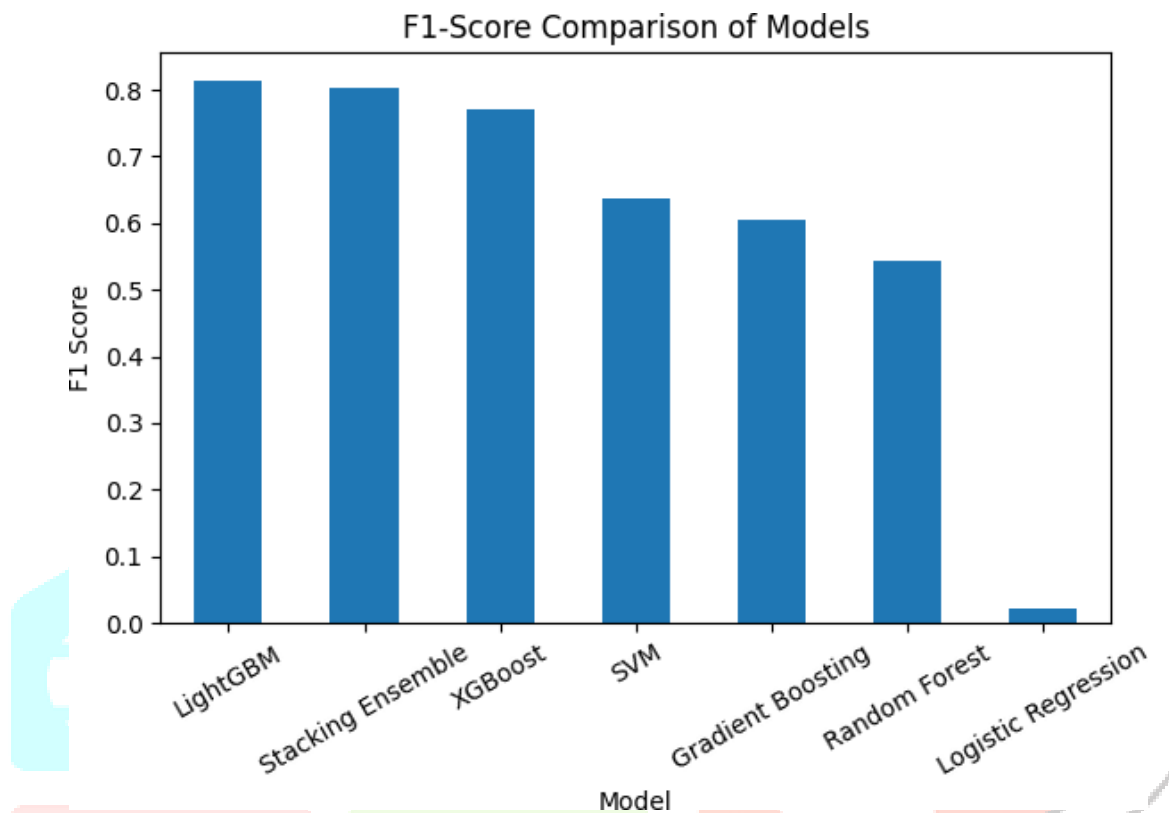


Figure 1: Comparison of F1-scores across evaluated machine learning classifiers for fake job posting detection. Figure 1 illustrates the comparative F1-score performance of the evaluated classifiers, providing a visual comparison of how different models perform.

5. Conclusion and Future Work

Machine learning techniques are increasingly being used to detect fake job postings on online recruitment platforms. Earlier studies mainly focused on traditional text classification models such as Logistic Regression, Support Vector Machines, and Random Forest. These approaches showed that textual features of job postings can provide useful signals for distinguishing fraudulent listings from legitimate ones.

More recent work has explored ensemble and boosting-based methods to address the limitations of baseline models. Techniques such as Gradient Boosting and related ensemble approaches have been shown to improve recall and F1-score in certain cases. However, performance often varies depending on the dataset, preprocessing methods, and evaluation setup, making it difficult to directly compare results across different studies.

In this study, multiple machine learning classifiers were evaluated under a consistent experimental setup using a common dataset, preprocessing pipeline, and evaluation framework. This helps ensure that differences in performance are mainly due to the models rather than variations in the setup. The results show that traditional models provide stable baseline performance, while ensemble and boosting-based methods achieve better minority-class detection under the same conditions. Overall, this suggests that ensemble approaches can offer a more balanced precision-recall trade-off in imbalanced classification problems.

Despite these findings, some limitations should be noted. The analysis is based on a single publicly avail-

able dataset and relies on TF-IDF features with classical machine learning models. As a result, the findings may not fully generalize to other datasets or real-world scenarios. This limitation is common in similar studies. Future work can explore deep learning-based models, including recurrent neural networks and transformer-based architectures, which are capable of capturing more complex semantic and contextual information from job descriptions. Comparing these approaches with classical methods may provide further insights into performance improvements. In addition, evaluating models on larger and more diverse datasets, along with considering practical factors such as computational efficiency and real-world deployment, would help improve the applicability of fake job detection systems.

References

- [1] M. Naudé, K. J. Adebayo, and R. Nanda, "A machine learning approach to detecting fraudulent job types," *AI & Society*, 2023.
- [2] K. Taneja, J. Vashishtha, and S. Ratnoo, "Fraud-BERT: Transformer-based context-aware online recruitment fraud detection," *Discover Computing*, 2025.
- [3] S. Dutta and S. Bandyopadhyay, "Text-based fraud detection using machine learning classifiers," *Procedia Computer Science*, 2020.
- [4] Kaggle, "Fake Job Postings Dataset," 2025.
- [5] A. Lal et al., "Fake job posting detection using ensemble machine learning techniques," *International Journal of Computer Applications*, 2021.
- [6] A. Sharma and S. Kumar, "Fake job detection using machine learning approaches," *International Journal of Information Technology*, 2021.
- [7] R. Singh and P. Sharma, "Detection of online recruitment fraud using text classification techniques," *Journal of Big Data*, 2022.
- [8] M. K. Gupta et al., "An efficient approach for detecting fraudulent job postings using ensemble learning," *Applied Intelligence*, 2022.
- [9] H. Alghamdi and A. Alotaibi, "Detecting fake job advertisements using supervised learning algorithms," *IEEE Access*, 2021.
- [10] S. Patel and D. Shah, "Comparative analysis of machine learning models for fraud detection in online recruitment systems," *Expert Systems with Applications*, 2023.