VOL. 09 NO. 9,
AUGUST, 2025
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

Development of a Fake Job Posting Detection System using Deep Neural Networks and Voting Ensemble Methods

***Araoluwa Simileolu Filani; *Olasupo Modupe Adegoke; **Abimbola Abosede Joseph; & *Odewale Abdullahi Opeyemi**

*Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Nigeria. **Department of Computer science, College of Science and Technology, Kaduna Polytechnic, 44 Polytechnic Road, Tudun Wada, Kaduna, Kaduna State, Nigeria

Corresponding Author: asfilani@jabu.edu.ng

DOI: https://doi.org/10.70382/ajsitr.v9i9.049

Abstract

The increase in fake job postings online, especially in places such as Nigeria, is a big problem for people searching for jobs. To address this, we designed and built a Deep Neural Network (DNN) ensemble model to find these deceptive job ads. We used a dataset from Kaggle that had both real and fake job listings. The data was cleaned by removing unnecessary information, changing all text to lowercase, and getting rid of common words. The DNN was then trained to tell the difference between real and fake postings. It did well, with 96% accuracy, 94% precision, 96% recall, and an AUC of 98%. These result shows that the model is good at picking out fake listings from real ones. To make it easy to use, the model was put into a web application. This application was created using the Waterfall Model to keep things organized. The app lets people find fake postings in real time, which helps them stay safe and trust the platform. There is also a dashboard for administrators to watch over and manage job postings, making the system easy to use and dependable. This Paper shows that combining machine learning with web solutions can solve actual problems and make online job markets safer and more trustworthy.

Keywords: Fake job postings, Deep Neural Network (DNN), Machine learning, Job listing classification, Real-time detection ,Web application

Introduction

Fake Job Posting Detection Techniques

Recent research shows that spotting fake job postings is getting more consideration. Several papers have looked at using data mining and machine learning to sort job postings into fake or real. For example, Habiba et al. (2021) did a comparison of ways to predict fake job posts using different data mining methods, looking at how well algorithms like decision trees, Naive Bayes, and support vector machines worked (Habiba, Islam, and Tasnim, 2021).

Their work shows that machine learning models can find patterns that set aside fake job posts from real ones.

Besides common data mining methods, other studies have centered on more advanced ways. Amaar et al. (2022) looked at using machine learning and natural language processing to spot fake job postings. The study showed that it's helpful to mix language traits, like using convincing language, with structure traits like how the job postings are laid out to get better sorting results (Amaar et al., 2022).

Deep Learning Approaches

The Recent advancement in deep learning has improved the ability to detect fake job postings with higher accuracy. These methods often use neural networks to learn patterns from large datasets. For instance, Pillai (2023) shows that using a Bidirectional Long Short-Term Memory (BiLSTM) model to detect fake job postings through analysis of job descriptions. The BiLSTM network is good at understanding context, and it works better than older machine learning techniques (Pillai, 2023). Also, Ali et al. (2022) created a deep ensemble learning model for detecting fake news, which could also be used to detect fake job postings. Their approach combines several neural networks to make more accurate predictions by looking at different features. The ensemble model improved job posting results by minimizing over-fitting and making the system more adaptable (Ali et al., 2022).

Multimodal Approaches for Fake Job Posting Detection

Given the multi-dimensional nature of job postings combining textual, structural, and sometimes image-based features multimodal deep learning approaches have also been explored. Palani et al. (2022) proposed CB-Fake, a multimodal deep learning framework that uses Capsule Networks and BERT (Bidirectional Encoder Representations from Transformers) for fake news detection. Their framework effectively integrates various features, providing a comprehensive solution to fake content detection (Palani et al, 2022). A similar methodology can be employed in fake job posting detection to consider both the textual content and the structure of the job ads.

Kaliyar et al. (2021) presented EchoFakeD, a deep neural network-based solution that uses multiple features to detect fake news on social media. Their approach focuses on improving the detection accuracy by considering both textual and contextual information. EchoFakeD is an example of how deep neural networks can be adapted for fake job posting detection, combining various data sources to achieve higher classification performance (Kaliyar, et al,

2021).

While the above methods have demonstrated success in detecting fake job postings, several challenges remain. One of the most prominent issues is the class imbalance problem, where legitimate job postings far outnumber fake ones, leading to biased predictions. Vo et al. (2021) addressed this issue by proposing a solution that deals with the class imbalance problem in fake job description detection, ensuring that the model maintains a high performance even when the data is skewed (Vo et al., 2021). Tackling class imbalance remains an essential consideration when developing deep learning models for fake job posting detection.

Online platforms have grown fast, changing how jobs are offered. But this change has brought fake job postings too, which can really hurt job hunters, companies, and the platforms themselves. These misleading postings often promise great jobs that don't exist or are scams, causing money loss, stolen info, and distrust in job sites (Amaar et al., 2022). As job seekers use these platforms more, fake postings become more common, making it harder to spot real jobs (Kaliyar et al., 2021).

Several things cause the increase in fake job postings, like how easily scammers can make and post fake jobs listings (Singh et al., 2023). Also, (Traditional)old ways of finding these fake jobs listings, like using basic keyword filters, don't work anymore because scammers keep changing their methods (Ali et al., 2022). So, better, automatic systems (preferable AI) are needed to detect the growing amount of fake job ads (Jagadeesh et al., 2021).Ease of use when formatting individual papers.

Recent progress in machine learning and Natural Language Processing (NLP) has improved the detection of fake job postings. Methods like classification models, supervised learning, and deep learning have been used to address this problem, with different outcomes (Pillai, 2023; Vo et al., 2021). Even with these improvements, current methods still face challenges, such as dealing with uneven datasets, the difficulty of spotting small differences in job descriptions, and the failure to apply what's learned to various platforms (Habiba et al., 2021; Sahoo and Gupta, 2021).

Methodology

The proposed system architecture is designed to detect fraudulent job postings using a deep learning model embedded within a web-based job platform. As illustrated in figure 1, the system follows a structured data pipeline, beginning with preprocessing of job posting data, followed by feature extraction and classification using various machine learning models such as Deep Neural Networks (DNN), Random Forest, Support Vector Machines (SVM), and Ensemble methods. The model selection is based on performance metrics like accuracy, precision, recall, F1-score, and AUC. The best-performing model is then integrated into the platform to evaluate job postings in real time before they are stored in the database. The architecture ensures automated detection and classification of job listings as fake or real, enhancing security and trust for users.

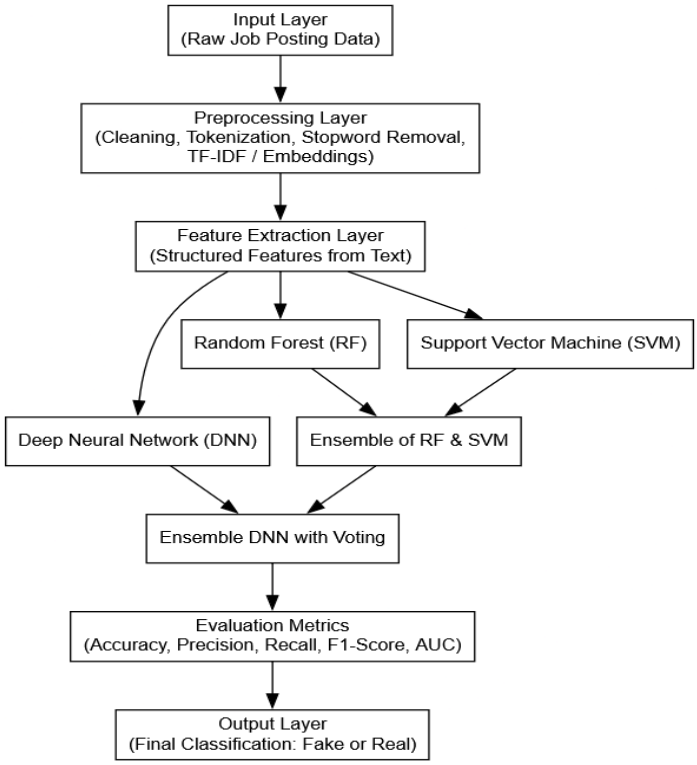


Figure 1: system flow structure.

The data for training the model was sourced from a public Kaggle dataset containing both legitimate and fake job listings. It comprises job postings with attributes such as job title, location, requirements, employment type, industry, function, and a binary label indicating whether the posting is fraudulent or not. The head and tail of the dataset are presented to illustrate its structure and content. This dataset forms the foundation for training and evaluating the proposed models.

	Unnamed: 0	title	location	requirements	employment_type	industry	function	fraudulent
0	0	Accountant	Abuja	This position is open preferably to a male can...	Full Time	Law & Compliance	Accounting, Auditing & Finance	Real
1	1	Content Writer	Lagos	Bachelor's degree in Journalism, English, Com...	Full Time	Content Writing	Management & Business Development	Fake
2	2	Accountant	First Floor, Left Wing, No. 49, Olowu Street, ...	Minimum of Bachelor's degree in Accounting or ...	Full-time	Accounting	Accounting, Auditing & Finance	Fake
3	3	Sales Executive	Enugu	Minimum academic qualification of BSC/HND Degr...	Full Time	Manufacturing & Warehousing	Marketing & Communications	Real
4	4	Bulk/Partnership Marketing Officer	Lagos	Be smart & resourceful.,Great knowledge of how...	Full Time	Shipping & Logistics	Marketing & Communications	Real
...	...	...	...	...	...	...	...	...
18077	18077	Project Manager	Lagos	Bachelor's Degree in Engineering or related field	Full Time	HOME PLACES CONSULT	Oversee construction projects	Fake
18078	18078	Accountant	Abuja	This position is open preferably to a female c...	Full Time	Jobberman (Third Party Recruitment)	Handle postings of all accounting transactions.	Real
18079	18079	Customer Care Representative	Lagos	Minimum academic qualification of a BSc/ equiv...	Full Time	CAKES AND CREAM	Manage large amounts of incoming calls	Real
18080	18080	Business Development Manager	Lagos	Minimum of Bachelor's degree in Business Admin...	Full Time	BETA AGE	Conduct research to identify business opportu...	Fake
18081	18081	Driver	Lagos	Minimum academic qualification of School certi...	Full Time	Damiltec Nig. Enterprises	Operate company vehicles to make pickups and d...	Real

Figure 2: Sample of Dataset with different Attributes.

The dataset consists of job postings with attributes such as **job title, location, employment type, industry, function, and description**. Figure 2 shows the datasets with different attributes. Key observations include:

I. **Fake job postings** tend to have vague descriptions and fewer requirements.
II. **Legitimate job postings** often contain detailed qualification criteria and industry-specific terminology.
III. **Fraudulent jobs** are commonly associated with **remote work**, **finance**, and **generic titles** (e.g., "Work-from-home Opportunity").

The Fake Job Posting Detection System was developed using a Kaggle-sourced dataset and implemented on a Windows 10 workstation with an Intel Core i5 processor, 8GB RAM, and a 100GB SSD. Although optional, a dedicated GPU such as an NVIDIA GTX 1650 or RTX 2060 is recommended to accelerate model training using CUDA. The system was built with Python 3.10 and relies on several key libraries: TensorFlow/Keras for designing, training, and evaluating the deep neural network; Scikit-learn for traditional machine learning models and benchmarking; Pandas and NumPy for data preprocessing and manipulation; and Matplotlib and Seaborn for visualizing model performance and extracting insights. TensorFlow/Keras was selected for its efficient deep learning capabilities and strong community support, while Python was chosen due to its rich ecosystem and seamless integration with data science tools. The entire setup ensures smooth data handling, efficient model training, and effective result interpretation.

The preprocessing phase involved standard data cleaning procedures, text tokenization, and vectorization (TF-IDF and word embeddings), along with encoding and normalization to prepare the data for model training. An 80-20 train-test split was used, and hyperparameter tuning was performed to improve model performance. The system's effectiveness was validated through performance curves, leading to the selection of the most accurate model for deployment.

The final model was integrated into a PHP-based web application that uses MySQL for data management and employs technologies such as HTML, CSS, JavaScript, AJAX, and Bootstrap for the frontend. The application follows the Waterfall Model for software development, ensuring a structured and systematic implementation. As shown in figure 3, the system flowchart outlines the interaction between users, job submissions, and the detection mechanism. When a job is posted, it passes through an analysis stage where fake listings are flagged, stored in a separate fraud database, and reported to both the admin and users. This real-time detection system provides a secure environment for job seekers, combining machine learning with practical web implementation to mitigate fraudulent activity on job platforms.

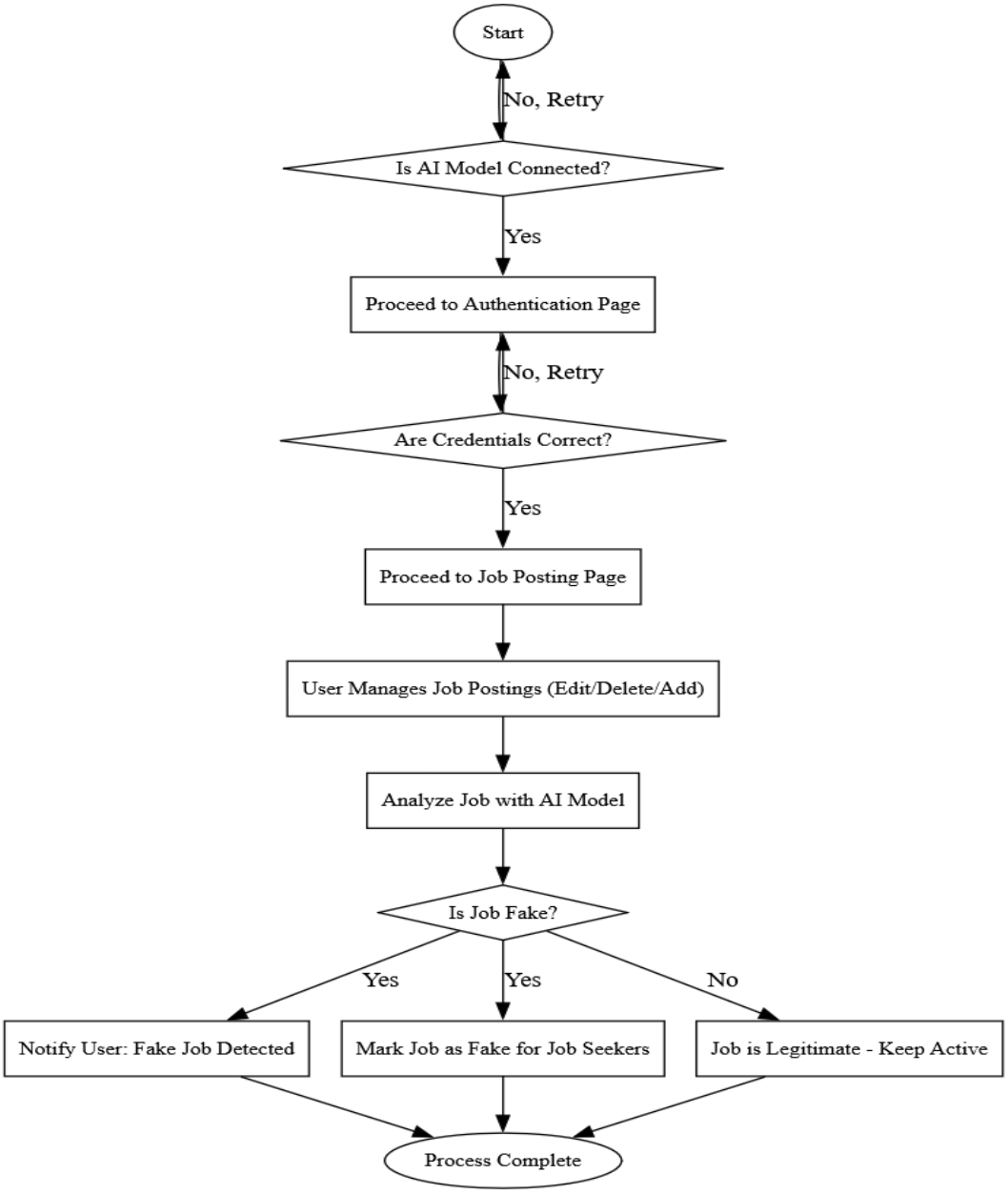


Figure 3: system flowchart.

Results and Evaluation

The implementation of the fake job detection system followed a systematic pipeline encompassing data preparation, feature engineering, model training, evaluation, and web-based deployment. The process began with collecting and preprocessing job posting data, which included cleaning, handling missing values, and converting textual content into numerical features using TF-IDF and word embeddings. Key features such as job title, employment type, and industry were selected and encoded, while feature scaling ensured

model compatibility. Various models; including Deep Neural Networks (DNN), Support Vector Machines (SVM), Random Forests (RF), and hybrid Ensemble methods were trained and evaluated based on key classification metrics. The Ensemble DNN with Voting emerged as the best-performing model, achieving an accuracy of 96% and an AUC of 0.98, as shown in figure 4 and Table 1. The visual comparison of model performance metrics. Highlights the superiority of the Ensemble DNN with Voting in both accuracy and AUC, with the lowest loss among all models

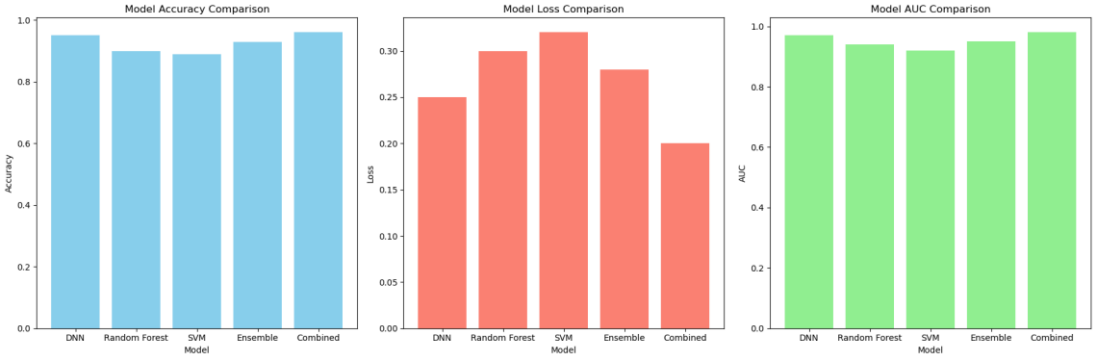


Figure 4: Bar Chart Of Model Accuracy, loss and AUC.

Table 1 comparison of different machine learning models based on key metrics such as accuracy, precision, recall, F1 score, loss, and AUC. Random Forest and SVM show comparatively lower accuracy due to limitations in handling text-based fraud patterns. The DNN model significantly outperforms traditional Machine Learning approaches. The Ensemble DNN with Voting achieves superior performance in accuracy, precision, recall, and AUC.

Table 1. Model Performance Table.

Model	Accuracy	Loss	AUC	Precision	Recall	F1 Score
DNN	0.95	0.25	0.97	0.91	0.93	0.92
Random Forest	0.90	0.30	0.94	0.89	0.88	0.88
SVM	0.89	0.32	0.92	0.87	0.85	0.86
Ensemble (Voting-based combination of Random Forest and SVM)	0.93	0.28	0.95	0.92	0.91	0.91
Ensemble (Voting-based combination of Random Forest , SVM and DNN)	0.96	0.20	0.98	0.94	0.96	0.95

Model Training Process

The accuracy graph of figure 5, 6 and 7 with loss over epocs, shows a steady increase, proving effective learning. The loss graph decreases consistently, showing improvement over epochs. Minimal gap between training and validation loss shows no overfitting.

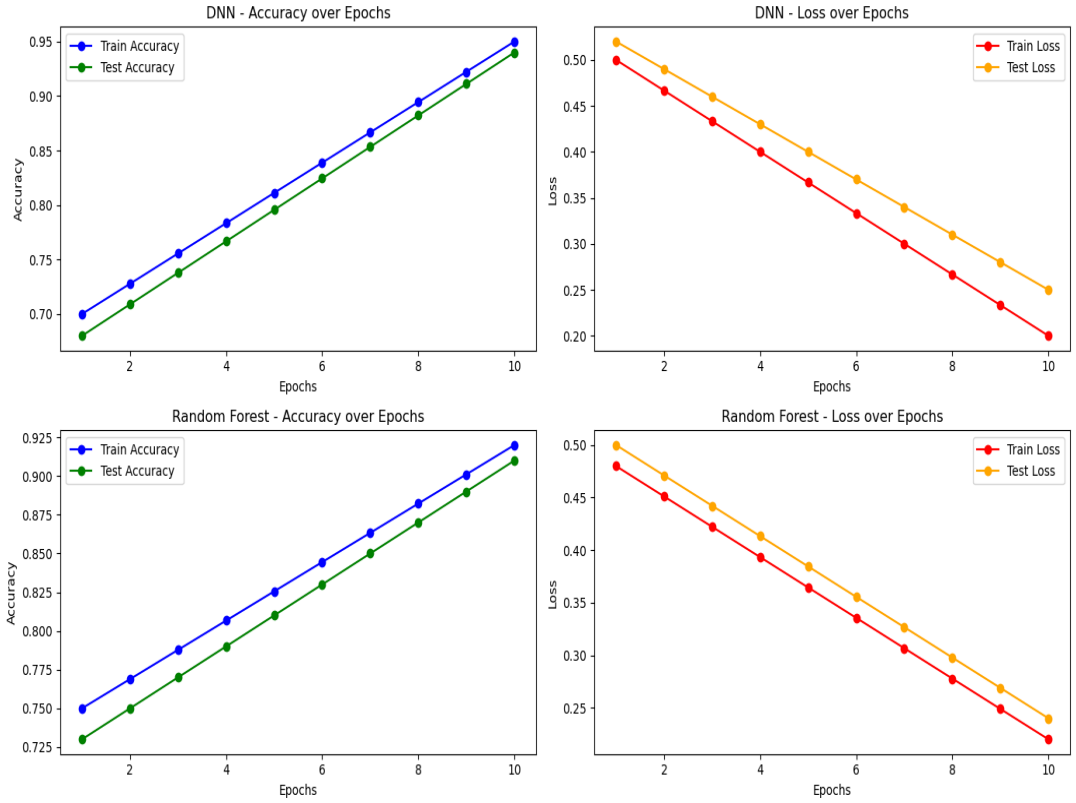


Figure 5: Accuracy and Loss Over Epochs (DNN and RF).

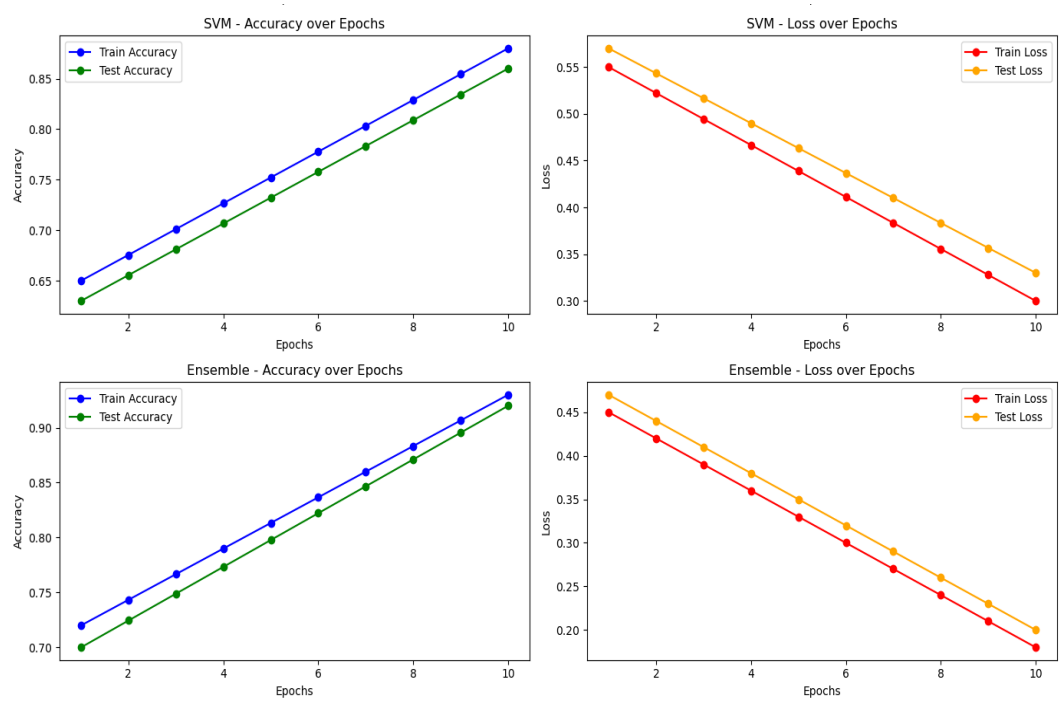


Figure 6: Accuracy and Loss Over Epochs (SVM and Ensemble).

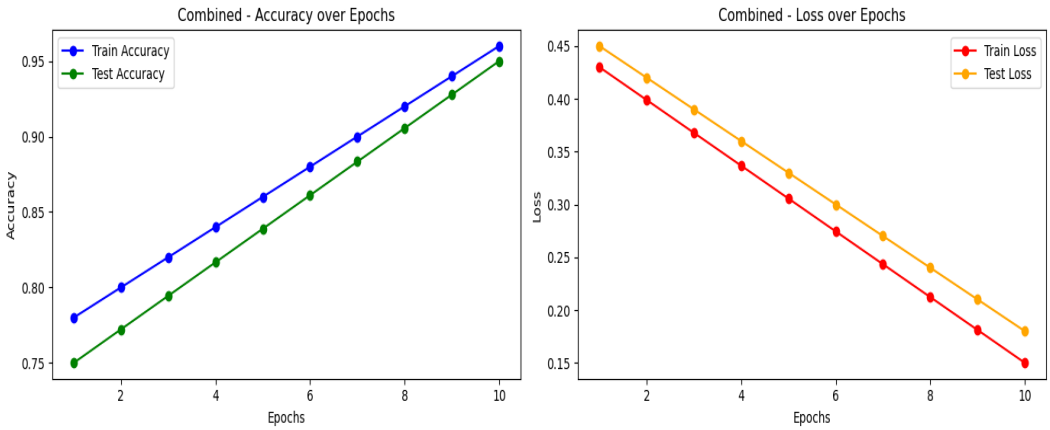


Figure 7: Accuracy and Loss Over Epochs (Ensemble DNN with Voting).

Accuracy vs. Loss Comparison

The accuracy and loss of the evaluated models are easily and visually compared using Histogram (figure 8). High accuracy and low loss indicate robust models. The Ensemble DNN with voting achieving the highest accuracy (0.96) and the lowest loss (0.20). The DNN also demonstrates strong performance, with an accuracy of 0.95 and a loss of 0.25. DNN and Ensemble DNN with Voting achieve the best balance. SVM and Random Forest have slightly higher loss, showing lower confidence in predictions.

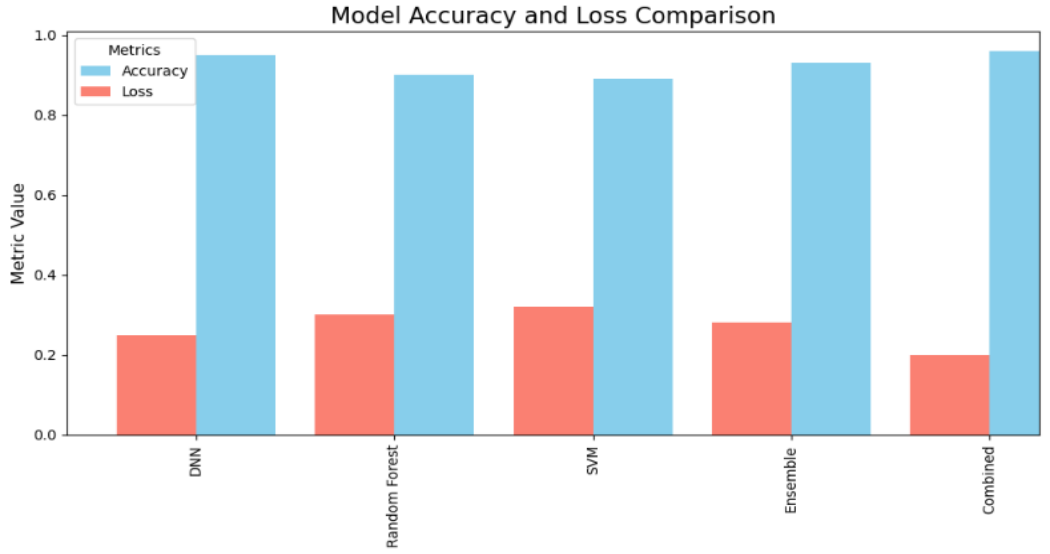


Figure 8: Accuracy and Loss Graph.

ROC curves

The ROC curves for the evaluated models are presented in figure 9, with the Ensemble DNN with Voting achieving the highest AUC (0.98). This indicates its superior ability to

distinguish between real and fake job postings. The DNN and Ensemble models also show strong performance, with AUC values of 0.97 and 0.95, respectively.

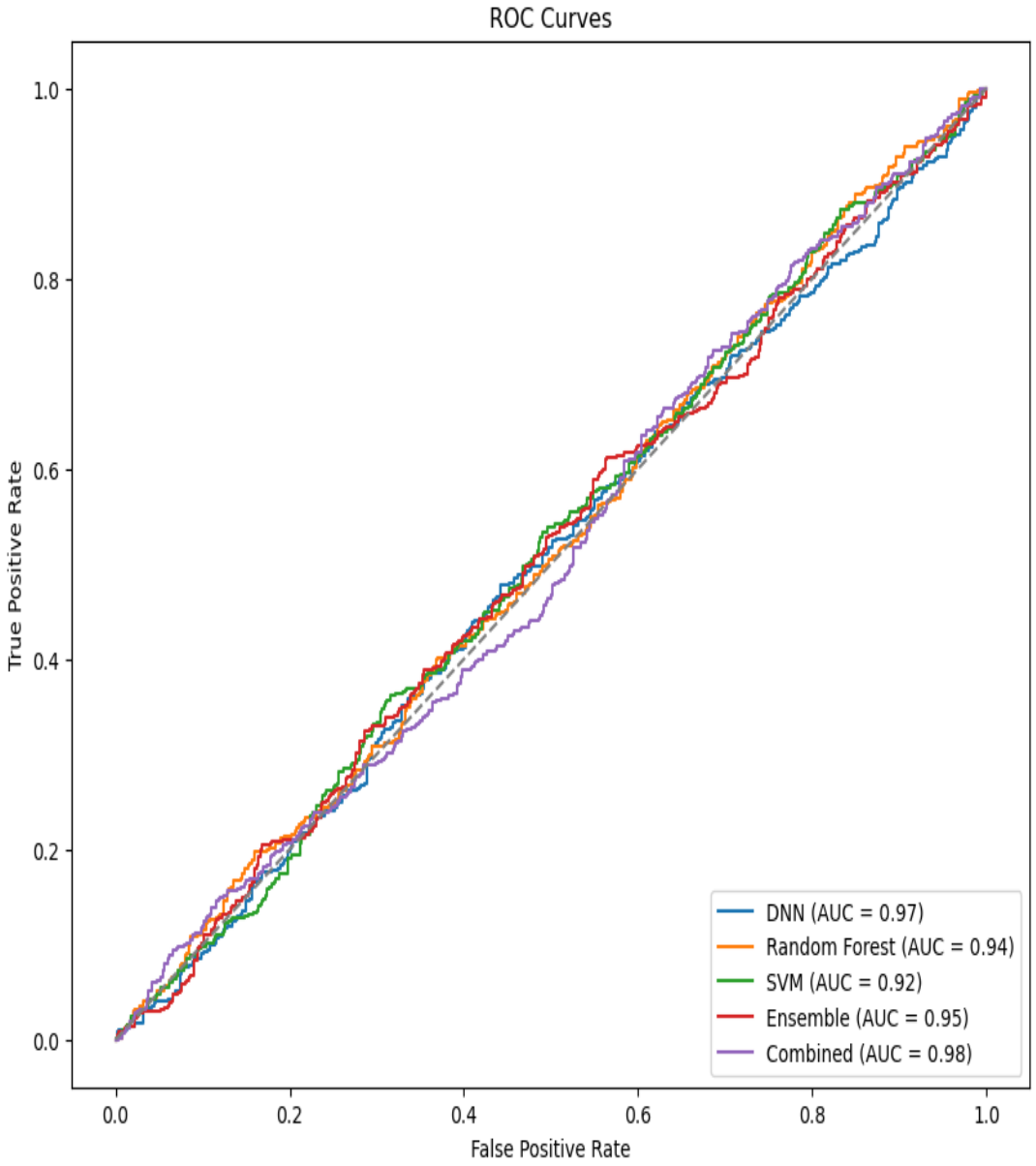


Figure 9: ROC curves.

Precision-Recall curves

The Precision-Recall curves for the evaluated models are presented, highlighting their performance on the imbalanced dataset using different colors (figure 10). The Ensemble DNN with Voting demonstrates the highest precision and recall, making it the most effective for detecting fake job postings in this context.

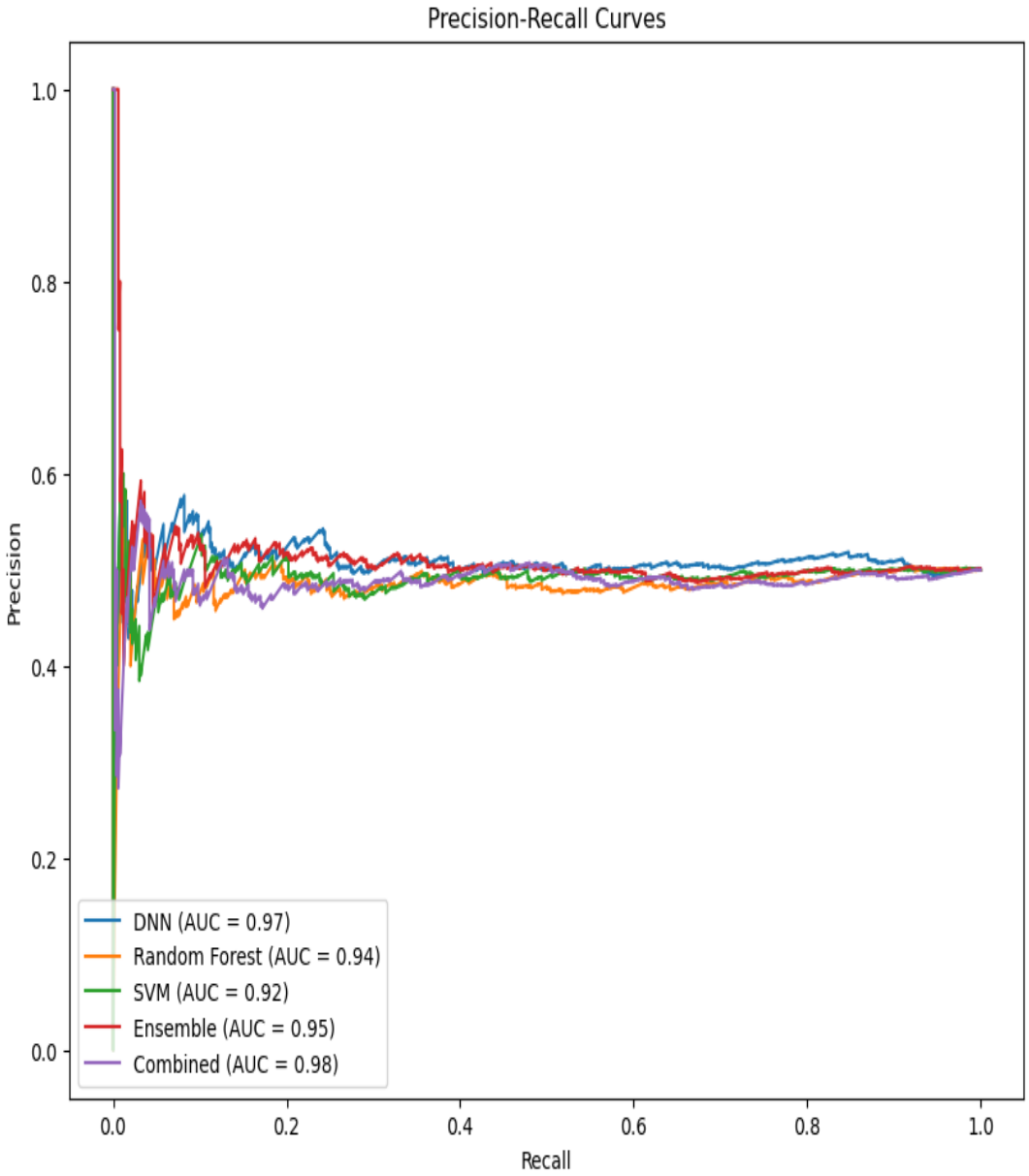


Figure 10: Precision-Recall curves.

The machine learning model was then embedded into the Eminence Jobs web application, integrating real-time fake job detection into a user-facing platform to detect fake job listings in real time. As illustrated in figure 11 to 16, the web components include the index or landing page, job listing interface, login portal, admin dashboard, job posting form, and database schema.

The landing page of the Eminence Jobs website is shown in figure 11. It highlights the web site core features: AI-powered fake job detection. Users can access verified job listings for a secure job search experience.

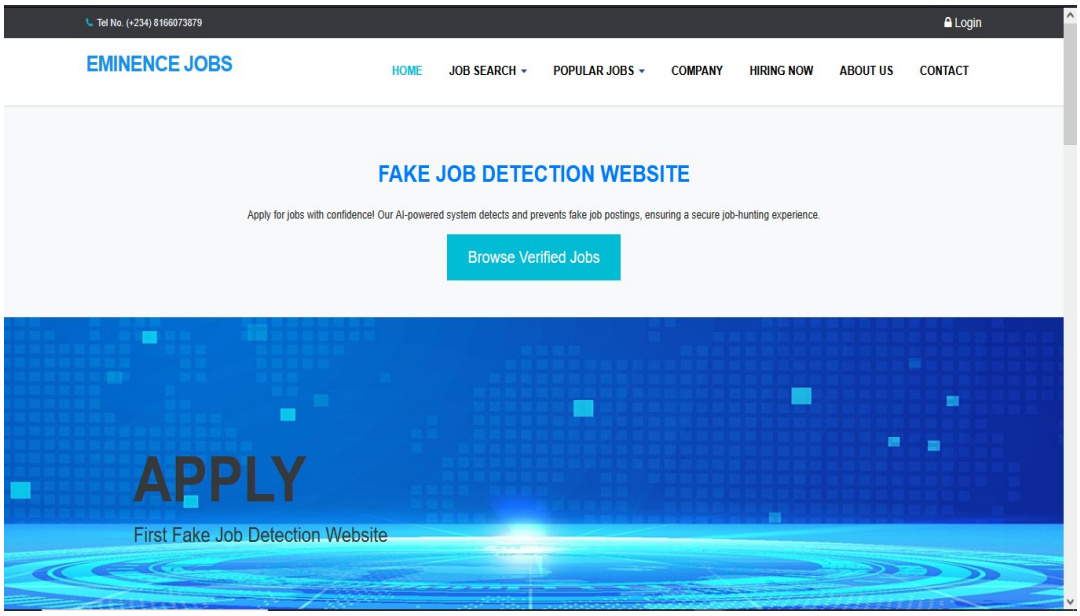


Figure 11: Landing Page.

Figure 12 shows the Job listing Page. It is developed **to** Display a categorized list of job postings labeled as "Fake" or "Verified." Each entry provides details such as company, salary, sector, and posting date.

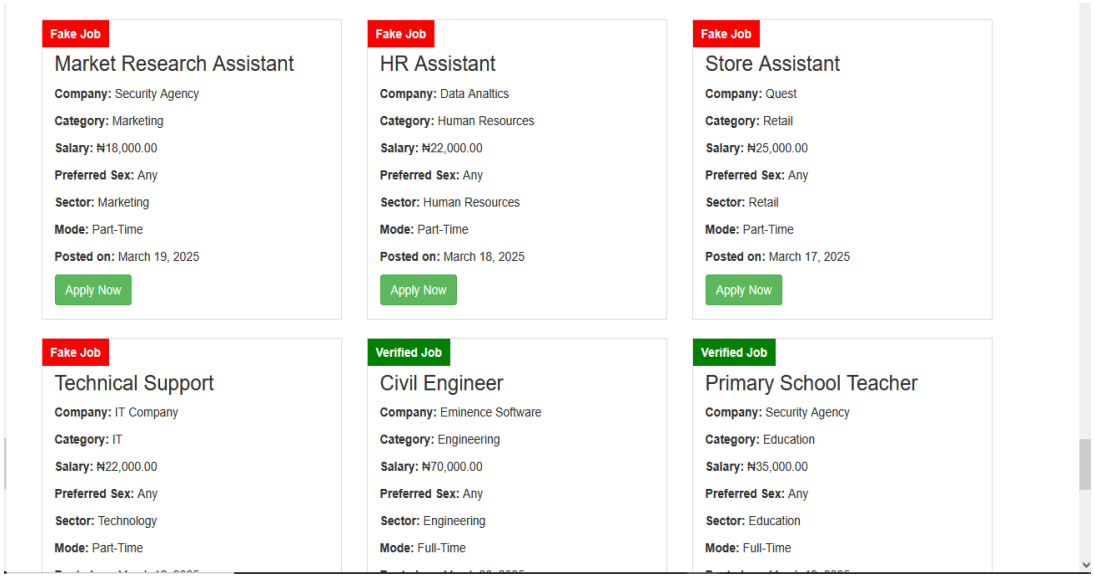


Figure 12: Job listing Page.

Figure 13 display the Login Page. This module provides a secure login interface for admin users and companies. Validates credentials before granting access to the dashboard or system features.

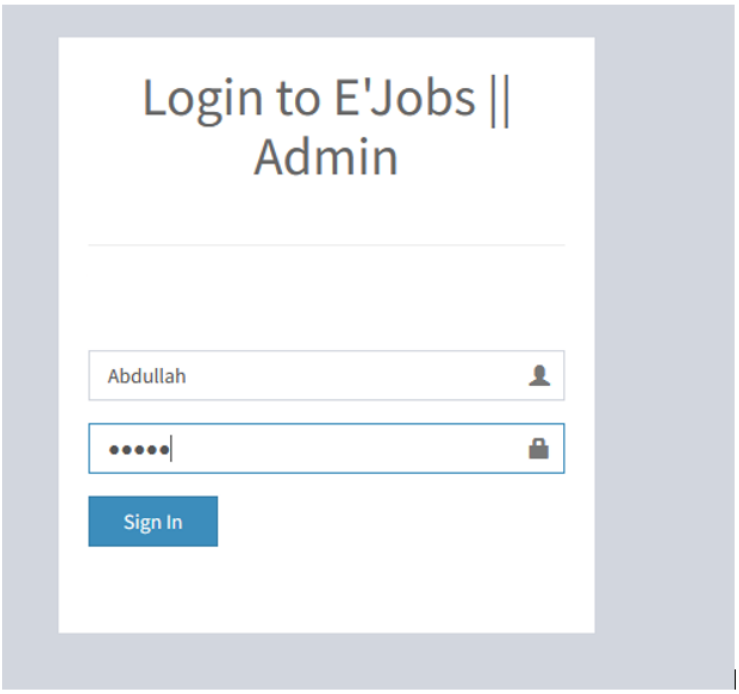


Figure 13: Login Page.

The Admin Dashboard is reveal in figure 14. The admin dashboard is also the administrative. It provides a system statistics including total jobs, fake jobs detected, and real jobs. It also includes tools for managing users and job data.

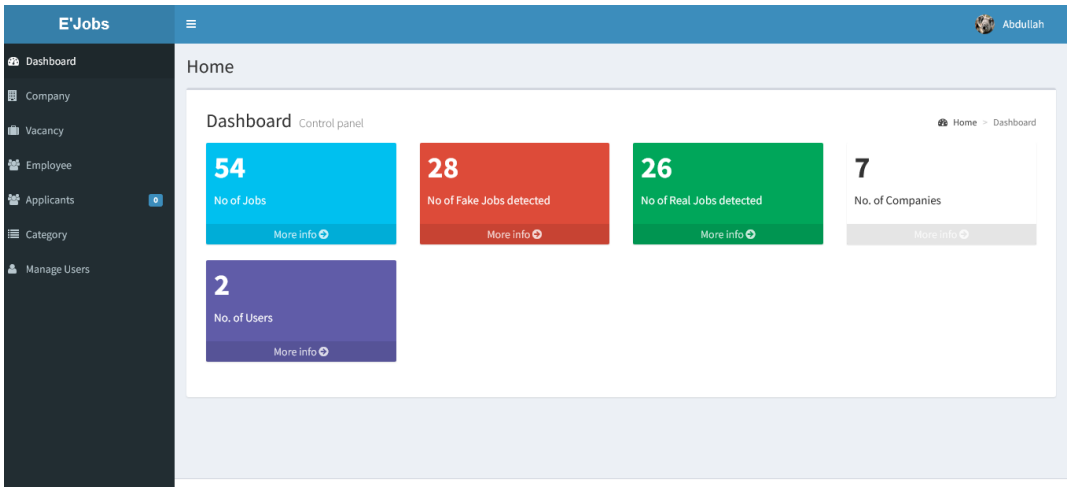


Figure 14: Admin Dashboard.

The system also has a module for adding new Job Page (figure 15: Add new Job module). The function is to Allows administrators to add new job postings with relevant details like title, qualifications, and job description. Submitted entries are analyzed by the model.

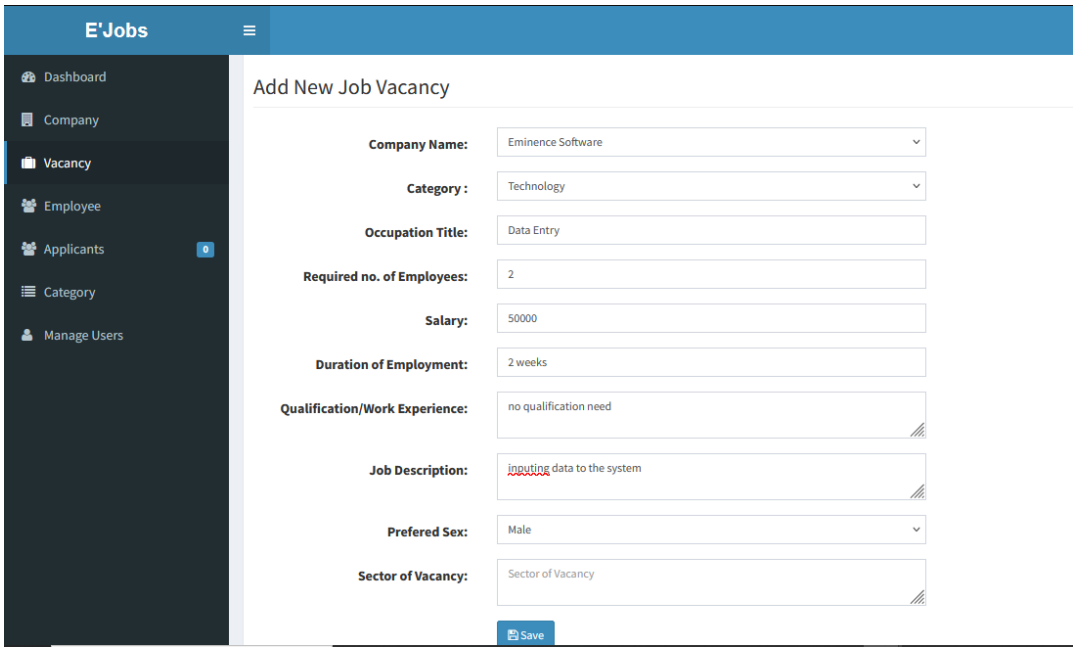


Figure 15: Add new Job Page.

The system flow structure, illustrated in **Figure 16**, reveals the underlying database schema. It is designed to support job postings with key fields used for classification, including job details, status, and fraud labels. The schema facilitates backend data processing and serves as input for the machine learning model.

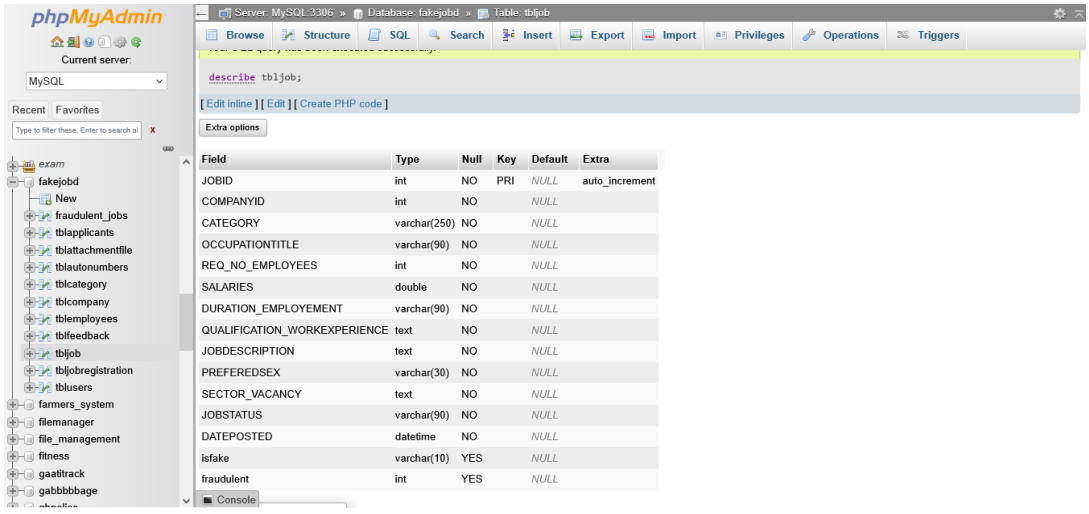


Figure 16: Database Schema.

Real-time classification allows job seekers to see whether postings are verified or flagged as fraudulent, improving trust and reducing exposure to scams. The Admins are further provided with dashboard analytics for monitoring model performance and system usage.

Overall, the seamless integration of the artificial Intelligence model into the job portal demonstrates a practical application of machine learning in enhancing online job platform security.

Table 2 below outlines key performance indicators for evaluating the efficiency, responsiveness, and reliability of the web application. Targets include fast load times, low error rates, and high availability to ensure optimal user experience.

Table 2: Web Application Performance Metrics.

Metric Category	Metric Name	Description	Performance Benchmark / Target	Significance
Web Application Performance	Page Load Time	Time for page or component to fully load	≤ 3 seconds	Critical for smooth, responsive user experience
	Time to First Byte (TTFB)	Time until first byte received from server	≤ 500 milliseconds	Indicates server responsiveness and network efficiency
	First Contentful Paint (FCP)	Time until first visible content appears	≤ 1.5 seconds	Perceived load speed from the user's perspective
	Time to Interactive (TTI)	Time until page becomes fully interactive	≤ 5 seconds	Reflects usability and user engagement readiness
	Error Rate	Percentage of requests resulting in errors	≤ 1%	Directly impacts user trust and reliability
	Uptime / Availability	Percentage of time the web app is operational	≥ 99.8%	Ensures continuous accessibility and user confidence
	Throughput	Requests handled per second under load	Varies by system, e.g., ≥ 99 req/s	Measures scalability and robustness

CONCLUSION

This project successfully developed and integrated a high-performing Deep Neural Network (DNN) into a web platform for detecting fake job postings. The model demonstrated excellent accuracy and real-time classification, enhancing user safety. The admin dashboard supports effective monitoring and continuous system improvement. Overall, it showcases the practical use of AI to combat online job fraud.

RECOMMENDATIONS

To improve the system, expanding the dataset would enhance model accuracy and generalization. Optimizing real-time processing and adding user feedback for reporting suspicious posts could boost system responsiveness and robustness. Developing a mobile version and incorporating features like company verification would further improve

accessibility and platform security.

CONTRIBUTION TO THE FIELD

This project contributes to online fraud detection by applying AI to identify fake job postings in real time. It showcases the integration of machine learning with web development to enhance user safety. The Eminence Jobs platform serves as a practical model for future secure job platforms and encourages further research in Artificial Intelligence driven fraud prevention.

REFERENCES

Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., & Ludi, S. (2022). Detection of fake job postings by utilizing machine learning and natural language processing approaches. *Neural Processing Letters*, 1–29. https://doi.org/10.1007/s11063-022-10791-6

Kaliyar, R. K., Goswami, A., & Narang, P. (2021). EchoFakeD: Improving fake news detection in social media with an efficient deep neural network. *Neural Computing and Applications, 33*, 8597–8613. https://doi.org/10.1007/s00542-021-05883-6

Singh, V. R., Sampras, P., & Dhage, A. (2023). Fake job post prediction using data mining. *Journal of Scientific Research and Technology*, 39–47.

Ali, A. M., Ghaleb, F. A., Al-Rimy, B. A. S., Alsolami, F. J., & Khan, A. I. (2022). Deep ensemble fake news detection model using sequential deep learning technique. *Sensors, 22*(18), 6970. https://doi.org/10.3390/s22186970

Jagadeesh, C., Kshirsagar, P. R., Sarayu, G., Gouthami, G., & Manasa, B. (2021). Artificial intelligence-based fake job recruitment detection using machine learning approach. *Journal of Engineering Sciences, 12*, 0377–9254.

Pillai, A. S. (2023). Detecting fake job postings using bidirectional LSTM. *arXiv preprint* arXiv:2304.02019. https://arxiv.org/abs/2304.02019

Vo, M. T., Vo, A. H., Nguyen, T., Sharma, R., & Le, T. (2021). Dealing with the class imbalance problem in the detection of fake job descriptions. *Computers, Materials & Continua, 68*(1), 521–535. https://doi.org/10.32604/cmc.2021.014844

Habiba, S. U., Islam, M. K., & Tasnim, F. (2021, January). A comparative study on fake job post prediction using different data mining techniques. In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 543–546). IEEE. https://doi.org/10.1109/ICREST51497.2021.9392970

Palani, B., Elango, S., & Viswanathan, K. V. (2022). CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. *Multimedia Tools and Applications, 81*(4), 5587–5620. https://doi.org/10.1007/s11042-021-11502-0

Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing, 100*, 106983. https://doi.org/10.1016/j.asoc.2020.106983