

Enhancing Online Recruitment Security: A Machine Learning-Based Fake Job Detection System

Author: **GOTTAPU DEVI KALYANI**¹ (MCA student), **MS. K. RAMYA KEERTHI**² (Assistant Professor)

Master Of Computer Applications, GIET ENGINEERING COLLEGE (A), Rajahmundry, AP.

Corresponding Author: Gottapu Devi Kalyani

(email-id: devikalyani2004@gmail.com)

ABSTRACT: The internet has transformed the way people search for employment opportunities, making online recruitment platforms an essential part of modern hiring. However, the increasing popularity of these platforms has also created opportunities for fraudulent activities in the form of fake job advertisements. Such deceptive postings often mislead applicants, resulting in the loss of personal information, financial resources, and trust in digital recruitment systems. This project introduces an intelligent framework designed to identify suspicious job advertisements before they can negatively impact job seekers. By utilizing machine learning techniques, the system examines various characteristics of job postings, including textual content, company details, and recruitment patterns, to distinguish legitimate opportunities from fraudulent ones. Multiple classification models are trained and evaluated using a publicly available recruitment dataset to ensure reliable performance. The proposed solution emphasizes accuracy, scalability, and practical usability, enabling real-time prediction through a user-friendly web interface. Rather than relying solely on manual verification, the system automates the screening process and provides an additional layer of security for online hiring environments. The outcome contributes toward creating a safer recruitment ecosystem by helping users make informed decisions while reducing the influence of deceptive job advertisements in digital employment platforms.

KEYWORDS: Fake Job Detection, Machine Learning, Online Recruitment Security, Deep Neural Network, EMSCAD Dataset, TF-IDF, Text Classification, Fraud Detection, SVM, Random Forest, KNN, Naive Bayes, Recruitment Scam Prevention, NLP, Predictive Analytics.

I. INTRODUCTION:

Finding a suitable job has become easier than ever due to the widespread use of online recruitment platforms. Every day, thousands of job opportunities are published on various websites, allowing job seekers to connect with employers from different locations and industries. While this digital transformation has simplified the hiring process, it has also created new challenges. One of the most concerning issues is the growing number of fake job advertisements that are designed to deceive applicants for financial gain or to collect sensitive personal information.

Fraudulent job postings often appear convincing because they imitate legitimate companies, use attractive salary packages, and create a sense of urgency. As a result, many job seekers struggle to distinguish genuine opportunities from deceptive ones. Traditional manual verification methods are often time-consuming and may not be effective when dealing with large volumes of job

advertisements. This situation highlights the need for an automated and reliable solution capable of identifying suspicious job postings before they reach potential applicants.

The rapid advancement of machine learning has opened new possibilities for solving such real-world problems. By learning patterns from historical data, machine learning models can recognize characteristics commonly associated with fraudulent recruitment activities. These technologies enable computers to analyze job descriptions, company information, and other relevant attributes with a level of speed and consistency that is difficult to achieve through manual inspection.

This project focuses on developing an intelligent fake job detection system that utilizes machine learning techniques to classify job advertisements as genuine or fraudulent. The proposed approach analyzes recruitment data collected from a publicly available dataset and applies various classification algorithms to evaluate their

effectiveness. Through systematic data preprocessing, feature extraction, model training, and performance evaluation, the system aims to provide accurate predictions and enhance trust in online recruitment environments.

By helping users identify potentially deceptive job advertisements at an early stage, the proposed system contributes to a safer digital hiring ecosystem. It not only reduces the risk of financial fraud and identity theft but also promotes greater confidence among job seekers who rely on online platforms to build their careers.

Objectives of the Project

The primary objective of this project is to develop an intelligent and dependable system capable of identifying fraudulent job advertisements on online recruitment platforms. The system aims to assist job seekers by providing an additional layer of verification before they engage with a job opportunity, thereby reducing the risk of scams, identity theft, and financial loss.

Another important objective is to utilize machine learning techniques to analyze job posting data and automatically distinguish between genuine and suspicious advertisements. By learning patterns from previously available recruitment data, the system seeks to improve the accuracy and efficiency of fraud detection compared to traditional manual screening methods.

The project also focuses on evaluating different machine learning algorithms and determining the most effective model for fake job detection. Through proper data preprocessing, feature extraction, training, and testing, the study aims to achieve reliable classification performance while maintaining scalability for real-world applications.

In addition, the project intends to create a user-friendly platform that enables users to verify job postings quickly and conveniently. The ultimate goal is to strengthen trust in online recruitment systems and contribute to a safer digital environment where individuals can search for employment opportunities with greater confidence and security.

II.Literature Survey:

The primary objective of this project is to develop an intelligent and dependable system capable of identifying fraudulent job advertisements on online recruitment platforms. The system aims to assist job seekers by providing an additional layer of verification before they engage with a job opportunity, thereby reducing the risk of scams, identity theft, and financial loss.

Another important objective is to utilize machine learning techniques to analyze job posting data and automatically distinguish between genuine and suspicious advertisements. By learning patterns from previously available recruitment data, the system seeks to improve the accuracy and efficiency of fraud detection compared to traditional manual screening methods.

The project also focuses on evaluating different machine learning algorithms and determining the most effective model for fake job detection. Through proper data preprocessing, feature extraction, training, and testing, the study aims to achieve reliable classification performance while maintaining scalability for real-world applications.

In addition, the project intends to create a user-friendly platform that enables users to verify job postings quickly and conveniently. The ultimate goal is to strengthen trust in online recruitment systems and contribute to a safer digital environment where individuals can search for employment opportunities with greater confidence and security.

III.System Analysis And Design

3.1 Existing System

The existing fake job detection system was developed to address the growing issue of fraudulent job advertisements appearing on online recruitment platforms. The primary objective of the system is to analyze job posting information and determine whether a particular advertisement is genuine or deceptive. To achieve this, the system collects data from recruitment datasets and applies machine learning techniques to classify job postings based on predefined patterns.

In the current approach, job-related information such as job title, company profile, job description, location, salary details, and employment requirements is gathered and processed. Before classification, the collected data undergoes several preprocessing steps, including removal of incomplete records, handling of missing values, and conversion of textual information into machine-readable formats. These steps help improve the quality of the dataset and prepare it for model training.

The existing system mainly depends on the Random Forest algorithm for prediction. The model is trained using historical job posting data and learns to recognize characteristics commonly associated with fraudulent advertisements. Once training is completed, the system analyzes new job postings and predicts whether they are legitimate or fake based on the learned patterns.

Although the system provides a foundation for automated fraud detection, its effectiveness is influenced by the quality of the dataset and the capability of the selected classification algorithm. As recruitment fraud

techniques continue to evolve, the system may require frequent updates and improvements to maintain reliable performance.

1. System Design

The design of the existing system follows a straightforward workflow consisting of data collection, preprocessing, feature extraction, model training, and prediction.

2. Data Collection Layer

This layer is responsible for obtaining job advertisement data from available recruitment datasets. The collected records contain both genuine and fraudulent job postings that are used for model development and evaluation.

3. Data Preprocessing Layer

The preprocessing stage cleans and organizes the raw dataset. Duplicate records, irrelevant attributes, and missing values are handled to improve data consistency. Textual information is transformed into a suitable format for machine learning operations.

4. Feature Extraction Layer

In this phase, important characteristics are extracted from job postings. Features such as keywords, company details, and textual patterns are converted into numerical representations that can be understood by the classification model.

5. Classification Layer

The processed data is supplied to the Random Forest classifier. The algorithm builds multiple decision trees and combines their outputs to generate a final prediction regarding the authenticity of a job advertisement.

6. Prediction and Output Layer

After classification, the system displays the prediction result to the user. The job posting is categorized as either genuine or fraudulent, allowing users to make informed decisions before proceeding with the application process.

The existing design offers a practical solution for fake job detection; however, its dependence on a single classification model limits its adaptability and may affect performance when faced with diverse recruitment fraud patterns.

3.2 Proposed System

The proposed system is designed to provide a more effective and reliable solution for identifying fraudulent job advertisements in online recruitment platforms. Unlike conventional approaches that depend on a single prediction model, the proposed system adopts a multi-algorithm strategy to improve detection accuracy and adaptability. The system analyzes various attributes of job postings and intelligently determines whether a recruitment advertisement is genuine or suspicious.

The need for such a system arises from the increasing sophistication of online recruitment scams. Fraudsters continuously modify their tactics, making it difficult for traditional verification methods to identify deceptive

postings. To address this challenge, the proposed system utilizes machine learning techniques capable of learning hidden patterns from historical recruitment data and applying that knowledge to new job advertisements.

The system processes job-related information including job titles, company profiles, employment requirements, benefits, and job descriptions. Through systematic preprocessing and feature extraction, meaningful information is obtained from the raw dataset. Multiple machine learning models are then trained and evaluated to identify the classifier that delivers the most accurate and consistent performance.

By automating the detection process, the proposed system reduces human effort, minimizes the chances of overlooking fraudulent advertisements, and provides users with timely verification results. The overall objective is to enhance recruitment security and create a safer environment for job seekers who rely on digital hiring platforms.

The proposed system follows a structured architecture consisting of several interconnected stages that work together to detect fraudulent job postings efficiently.

1. Data Acquisition Layer

This layer is responsible for collecting recruitment data from the EMSCAD dataset. The dataset contains both legitimate and fraudulent job advertisements, providing a balanced environment for model training and evaluation.

2. Data Preprocessing Layer

Raw recruitment data often contains inconsistencies, irrelevant information, and missing values. In this stage, data cleaning techniques are applied to improve data quality. Unnecessary symbols, duplicate entries, and incomplete records are removed, while textual information is standardized for further analysis.

3. Feature Engineering Layer

After preprocessing, important features are extracted from the dataset. Techniques such as tokenization and TF-IDF vectorization are used to convert textual job information into numerical representations. This process enables machine learning algorithms to identify significant patterns and relationships within the data.

4. Model Training Layer

The transformed dataset is divided into training and testing subsets. Multiple machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors (KNN), and Deep Neural Networks (DNN), are trained using the prepared data. Each model learns distinctive characteristics associated with genuine and fraudulent job postings.

5. Model Evaluation Layer

The performance of each algorithm is assessed using evaluation metrics such as accuracy, precision, recall, and F1-score. Comparative analysis is performed to

determine the most suitable model for deployment within the system.

6. Prediction Layer

Once the optimal model is selected, it is integrated into the application. When a user submits job-related information, the trained model analyzes the input and instantly predicts whether the advertisement is genuine or fraudulent.

7. User Interface Layer

A user-friendly web interface developed using Python and Django allows users to interact with the system. Job seekers can enter recruitment details and receive prediction results in a simple and understandable format.

8. Overall Design Outcome

The proposed design combines advanced preprocessing techniques, effective feature extraction methods, and multiple machine learning models to create a comprehensive fake job detection framework. By focusing on accuracy, scalability, and ease of use, the system offers a practical solution for improving trust and security in online recruitment platforms while protecting users from potential fraud.

3.3 System Architecture

The system architecture of the proposed Fake Job Detection System is designed to provide an organized flow of data from user input to final prediction. The architecture integrates data processing, machine learning models, and a web-based interface to identify fraudulent job advertisements efficiently. Each component performs a specific task, ensuring that the overall system operates smoothly and produces reliable results.

The process begins when recruitment data is collected from the EMSCAD dataset or when a user enters job-related information through the application interface. The received data is forwarded to the preprocessing module, where unwanted symbols, duplicate records, missing values, and irrelevant information are removed. This step improves data quality and prepares the content for further analysis.

After preprocessing, the cleaned text is passed to the feature extraction module. Since machine learning algorithms cannot directly understand textual information, techniques such as tokenization and TF-IDF vectorization are applied. These methods convert job descriptions and company information into numerical representations while preserving their meaningful patterns.

The transformed data is then supplied to the machine learning layer. Multiple classification algorithms, including Random Forest, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors, and Deep Neural Networks, are trained using historical recruitment data. During training, the models learn the characteristics that commonly differentiate genuine job advertisements from fraudulent ones.

Once training is completed, the prediction engine evaluates incoming job postings using the selected model. The engine analyzes the extracted features and generates a classification result indicating whether the job advertisement is genuine or suspicious. The prediction process occurs within a short time, allowing users to obtain immediate feedback.

Finally, the result is displayed through a user-friendly web interface developed using Django.

The system presents the prediction outcome in a simple format that can be easily understood by job seekers and administrators. This architecture not only improves detection accuracy but also supports scalability, making it suitable for real-world recruitment platforms where large volumes of job advertisements are processed regularly.

Architecture Flow

1. User Input / EMSCAD Dataset
2. Data Preprocessing
3. Text Cleaning and Tokenization
4. TF-IDF Feature Extraction
5. Machine Learning Model Training
6. Model Evaluation and Selection
7. Prediction Engine
8. Result Generation
9. User Dashboard Display

The architecture ensures that every job advertisement passes through multiple validation stages before a final decision is produced, thereby enhancing the security and trustworthiness of online recruitment systems.

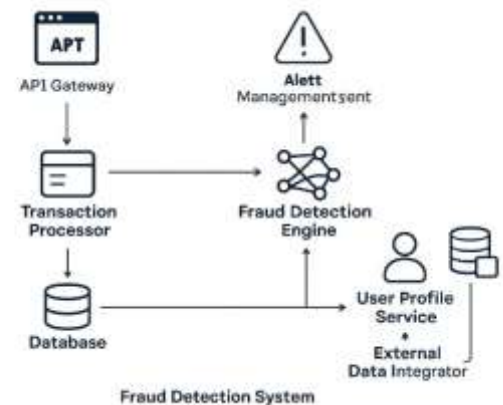


Fig1. This figure shows the architecture

IV. PROPOSED METHODOLOGY

The methodology of the proposed Fake Job Detection System is designed to systematically identify fraudulent job advertisements by applying machine learning techniques to recruitment data. The entire process follows a structured sequence of operations, beginning with data collection and ending with the generation of prediction results. Each stage plays an important role in ensuring that the final output is accurate, reliable, and suitable for practical use.

The first step involves collecting recruitment data from the EMSCAD dataset, which contains a mixture of genuine and fraudulent job postings. This dataset serves as the foundation for training and evaluating the machine learning models. Since real-world datasets often contain inconsistencies and incomplete information, the collected data undergoes a preprocessing stage before being used for analysis.

During preprocessing, unnecessary symbols, duplicate entries, missing values, and irrelevant attributes are removed. Text normalization techniques such as converting text to lowercase, removing stop words, and eliminating unwanted characters are applied to improve data quality. This stage ensures that only meaningful information is retained for further processing.

Once the data is cleaned, feature extraction is performed to transform textual information into a format understandable by machine learning algorithms. The system utilizes tokenization and TF-IDF vectorization to identify important words and phrases within job descriptions. These techniques help capture the significance of terms that frequently appear in fraudulent or legitimate job advertisements.

After feature extraction, the prepared dataset is divided into training and testing sets. The training data is used to teach the machine learning models how to recognize patterns associated with genuine and fake job postings, while the testing data is used to evaluate model performance on unseen records.

To achieve reliable classification, multiple machine learning algorithms are employed, including Random Forest, Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors (KNN), and Deep Neural Networks (DNN). Each model analyzes the extracted features and learns different relationships within the recruitment data. The performance of these models is then compared using evaluation metrics such as accuracy, precision, recall, and F1-score.

Based on the evaluation results, the most effective model is selected and integrated into the prediction module. Whenever a new job advertisement is submitted, the system processes the input through the same preprocessing and feature extraction stages before passing it to the trained model. The model then determines whether the advertisement is genuine or fraudulent and generates the corresponding prediction.

The final result is presented through a web-based interface developed using Django, allowing users to interact with the system conveniently. This methodology ensures that every job posting is analyzed through a well-defined sequence of operations, providing a dependable mechanism for detecting recruitment fraud and improving trust in online hiring

platforms.

Methodology Flow

1. Data Collection from EMSCAD Dataset
2. Data Cleaning and Preprocessing
3. Text Normalization
4. Feature Extraction using TF-IDF
5. Dataset Splitting (Training and Testing)
6. Machine Learning Model Training
7. Performance Evaluation
8. Best Model Selection
9. Job Advertisement Classification
10. Result Display through Web Application

The proposed methodology combines data preparation, intelligent learning, and automated prediction to create a comprehensive framework capable of detecting fake job postings with high efficiency and consistency.

V.IMPLEMENTATION

5.1 Technologies Used

The development of the Fake Job Detection System involves the integration of several technologies that work together to provide accurate prediction, efficient data processing, and a user-friendly experience. Each technology contributes to a specific aspect of the system, ranging from machine learning model development to web application deployment.

1.Python

Python serves as the primary programming language for the project. It is widely used in machine learning and data science because of its simple syntax, extensive library support, and flexibility. Python enables efficient implementation of data preprocessing, model training, evaluation, and web application functionalities. Its rich ecosystem makes it suitable for handling both analytical and development tasks within a single environment.

2.Machine Learning

Machine learning forms the core of the proposed system. It enables the application to learn patterns from historical recruitment data and make intelligent predictions regarding the authenticity of job advertisements. Various classification algorithms are utilized to analyze recruitment information and identify characteristics commonly associated with fraudulent job postings.

3.Deep Neural Network (DNN)

Deep Neural Networks are employed to capture complex relationships hidden within recruitment data. Unlike traditional algorithms, DNN models can automatically learn multiple levels of data representation, making them effective for detecting sophisticated fraud patterns. The neural network improves the system's ability to distinguish genuine job advertisements from deceptive ones.

4. Django Framework

Django is used as the backend framework for developing the web application. It provides a secure and organized environment for handling user requests,

managing data, and integrating machine learning models into the application. Django follows the Model-View-Template architecture, which simplifies development and improves maintainability.

5.TF-IDF Vectorization

Term Frequency–Inverse Document Frequency (TF-IDF) is utilized as a feature extraction technique. Since machine learning algorithms cannot directly process textual information, TF-IDF converts job descriptions and recruitment content into numerical values while preserving the importance of significant words. This process helps improve classification accuracy.

6.Natural Language Processing (NLP)

Natural Language Processing techniques are used to analyze and prepare textual job information for machine learning models. Operations such as tokenization, stop-word removal, and text normalization help extract meaningful information from job advertisements and improve the quality of predictions.

7.MySQL Database

MySQL is used for storing user information, prediction records, and system-related data. It provides efficient data management, reliable storage, and quick retrieval of information. The database supports the smooth functioning of the web application and ensures data consistency.

8.HTML

HTML is used to create the structural layout of web pages. It defines the arrangement of elements such as forms, buttons, navigation menus, and content sections within the user interface.

9.CSS

CSS enhances the visual appearance of the application by controlling colors, fonts, spacing, and page layouts. It improves the overall user experience by making the interface attractive and easy to navigate.

10.JavaScript

JavaScript adds interactivity to the web application. It enables dynamic page behavior, form validation, user notifications, and responsive interactions without requiring constant page reloads.

11.Scikit-learn

Scikit-learn is a powerful machine learning library used for implementing algorithms such as Random Forest, Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN). It provides efficient tools for model training, testing, and performance evaluation.

12.Pandas

Pandas is used for data manipulation and analysis. It simplifies tasks such as reading datasets, handling missing values, filtering records, and preparing data for machine learning operations.

13.NumPy

NumPy provides support for numerical computations and mathematical operations. It helps manage large datasets efficiently and improves the performance of machine learning workflows.

5.2 Modules

Service Provider Module

The Service Provider Module is the core component of the system that performs all machine learning operations. It handles data preprocessing, feature extraction, model training, and prediction generation. Various algorithms such as Random Forest, SVM, Naïve Bayes, KNN, and Deep Neural Networks are used to analyze job postings and identify whether they are genuine or fraudulent.

Admin Module

The Admin Module is responsible for managing and monitoring the entire application. It allows the administrator to view registered users, authorize access, and supervise system activities. This module helps maintain security and ensures that the platform operates smoothly and efficiently.

Remote User Module

The Remote User Module is designed for job seekers who use the system to verify job advertisements. Users can register, log in, and submit job-related information for analysis. Based on the prediction generated by the machine learning model, the system informs users whether the job posting is genuine or fake, helping them make safer decisions.

VI.RESULTS AND ANALYSIS

The Secure Student Election System Using Face Recognition was successfully implemented and tested using various functional modules. The developed system provides a secure platform for voter registration, authentication, election management, and result generation. The following results were obtained during system execution.



Fig 1: This figure shows the Home page of the Application

The Home Page serves as the entry point of the Fake Job Detection System and provides users with an overview of the application's purpose and functionality. It is designed with a simple and user-friendly interface that allows visitors to easily navigate through the system. From this page, users can access registration and login options, learn about the importance of detecting fraudulent job advertisements, and begin the verification process. The Home Page acts as a gateway that connects users to the system's services while creating a clear and

welcoming first impression.

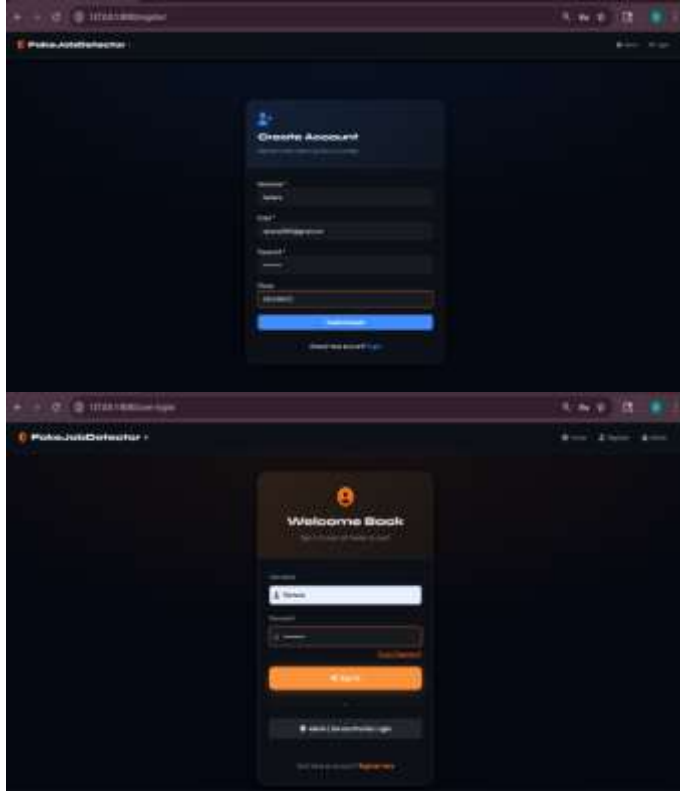


Fig 2: user Login Interface.

The User Login Page allows registered users to securely access the Fake Job Detection System using their credentials. It acts as a verification point that ensures only authorized users can utilize the application's features. Once logged in, users can submit job-related information, view prediction results, and access their personal dashboard. The page is designed to provide a smooth and secure login experience while protecting user data and maintaining system integrity.

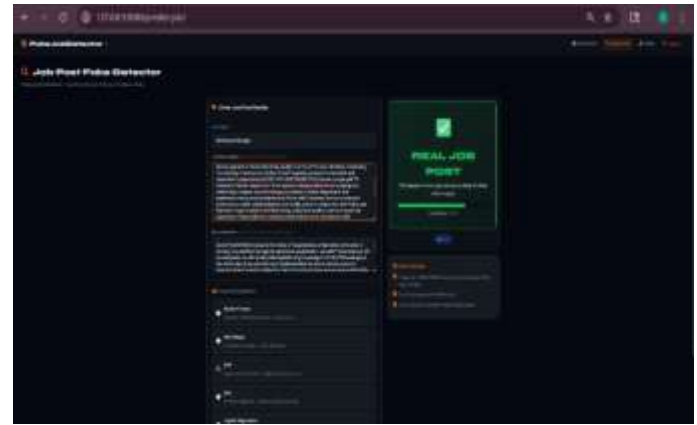
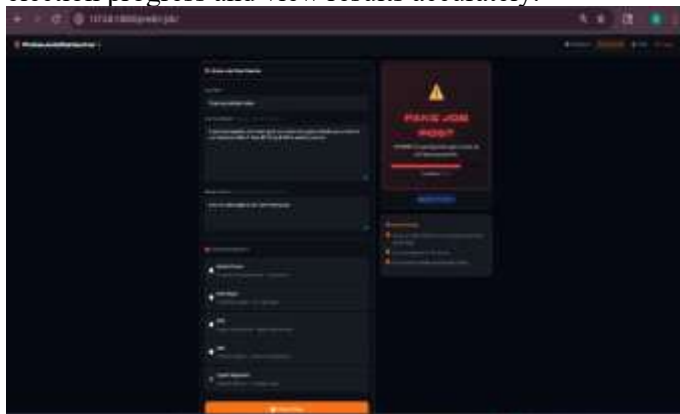


Fig 3: predict job page interface of the application

The Predict Job Page is the main working area of the system where users can enter job-related details for analysis. After receiving the input, the system processes the information through the trained machine learning model and evaluates the authenticity of the job advertisement. Within a few moments, the page displays whether the job posting is likely to be genuine or fraudulent. This feature helps users make informed decisions and reduces the risk of falling victim to recruitment scams.

VII. CONCLUSION

The completion of this project demonstrates how machine learning can be applied to address a practical challenge that affects thousands of job seekers in the digital recruitment landscape. As online hiring continues to expand, the risk of encountering misleading and fraudulent job advertisements has become increasingly common. The proposed system was developed with the intention of providing an intelligent mechanism capable of identifying such threats before they negatively impact users.

Through the application of data preprocessing techniques, feature extraction methods, and multiple machine learning algorithms, the system is able to analyze recruitment information and make informed predictions regarding the authenticity of job postings. The comparative evaluation of different models helped in identifying an effective approach for detecting fraudulent advertisements with improved reliability. Beyond its technical implementation, the project highlights the importance of combining technology with user safety. By assisting individuals in verifying employment opportunities, the system contributes to reducing the chances of financial fraud, misuse of personal information, and unnecessary inconvenience during the job search process. The developed solution not only improves trust in online recruitment platforms but also demonstrates the potential of intelligent systems in solving real-world security concerns.

In conclusion, the project represents a meaningful step

toward creating a safer and more transparent recruitment environment where job seekers can explore opportunities with greater confidence and peace of mind.

Haa

Data mining Techniques.”2ndInternationalConferenceonRobotics, Electrical and Signal Processing Techniques (ICREST) (2021)

VIII. REFERENCES

- Gulshan Shrivastava , Member, IEEE, Prabhat Kumar, Senior Member, IEEE, Rudra Pratap Ojha , Pramod Kumar Srivastava ,Senthil Kumar Mohan “Defensive Modeling of FakeNews Through Online Social Networks”,—Online social networks (OSNs) IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS
- deBeer, Dylan & Matthee, Machdel,”Approachesto Identify FakeNews: A Systematic Literature Review.” : Integrated Science in Digital Age2020.
- Bandar Alghamdi, FahadAlharby,”AnIntelligent Model for Online Recruitment Fraud Detection”. Journal of Information Security. (2019)
- Pham, Trung Tin” A Study on Deep Learning for Fake News Detection.” Journal of Information Security. (2019) n
- Manoj Kumar Balwant.” Bidirectional LSTM Based on POS tags and CNN Architecture for Fake News” 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). (2019)
- Amjad, Maaza , Sidorov, Grigoria, Zhila, Alisaa, Gómez-Adorno, Helenab, Voronkov, Iliac, Gelbukh, Alexander.” Bend the truth” Special section: Selected papers of LKE 2019
- Rami Mohawesh, Son Tran, Robert Ollington, Shuxiang Xu” Analysis of concept drift in fake reviews detection.” Expert Systems with Applications. (2021)
- Joma George; Shintu Mariam Skariah; T. Aleena Xavier.” Role of Contextual Features in Fake News Detection: AReview” International Conference on Innovative Trends in Information Technology (ICITIIT). (2021)
- In Sultana Umme Habiba; Md. Khairul Islam; Farzana Tasnim.” A Comparative Study on Fake Job Post Prediction Using Different