INTELLIGENT RECRUITMENT SCAM DETECTION

Mrs. A.Shobana, Assistant Professor,

Department of Computer Applications,

Sri Krishna Arts and Science College, Coimbatore-641 008

Negha K, Department of Computer Applications,

Sri Krishna Arts and Science College, Coimbatore-641 008

Abstract

Online recruitment platforms have made job searching easier, but they have also increased the number of fraudulent job postings. Recruitment scams often target job seekers by offering fake job opportunities and requesting money or sensitive information. Identifying such scams manually is difficult because fraudulent messages are written in a convincing way. This paper proposes an Intelligent Recruitment Scam Detection System that uses Machine Learning techniques to classify job descriptions as Scam or Real.

The proposed system uses Count Vectorizer for text feature extraction and the Naive Bayes algorithm for classification. The model learns patterns from historical job description data and predicts the authenticity of new job postings. The system is implemented using Python and deployed with a Streamlit interface to provide an interactive and user-friendly experience. The results demonstrate that the Naive Bayes classifier performs efficiently for text-based scam detection tasks. The proposed solution offers a reliable, fast, and practical approach to protect users from recruitment fraud.

Keywords: Recruitment Scam Detection, Online Job Fraud Detection, Machine Learning, Supervised Learning, Naive Bayes Classifier, Probabilistic Classification, Count Vectorizer, Bag of Words Model, Text Classification, Natural Language Processing (NLP), Text Mining, Feature Extraction, Spam Detection, Fraud Detection System, Online Security, Cyber Security, Job Portal Safety, Data Preprocessing, Classification Model, Artificial Intelligence.

1.Introduction

The internet has transformed the recruitment process by allowing companies to post job openings online and enabling candidates to apply from anywhere. However, along with these benefits, online recruitment scams have become increasingly common. Fraudsters create fake job advertisements that appear genuine and attract job seekers. These scams may request application fees, training charges, or personal information such as bank details. Many job seekers lack the technical knowledge to identify fake job postings. Traditional detection methods rely on manual checking or fixed keyword-based filtering.

These approaches are not reliable because scammers continuously modify their techniques and language to avoid detection.To overcome these limitations, this project introduces a machine learning-based Intelligent Recruitment Scam Detection System. The system automatically analyzes job descriptions and classifies them into two categories: Scam or Real. By using the Naive Bayesalgorithm and Count Vectorizer, the system learns from training data and detects patterns associated with fraudulent job postings. The objective of this research is to design a system that is accurate, fast, simple to use, and effective in reducing recruitment fraud risks.

2.Literature Review

Text classification is widely used in spam detection, sentiment analysis, and fraud detection. Researchers have applied various machine learning algorithms to classify text data. Among these algorithms, Naive Bayes has been frequently used due to its simplicity, efficiency, and good performance in handling textual data.Previous studies in spam email detection show that Naive Bayes performs well when combined with Bag of Words or Count Vectorizer techniques. These techniques convert text into numerical features that machine learning models can process. Several research works have also explored online fraud detection using supervised learning models.Compared to complex deep learning models, Naive Bayes requires less computational power and is easier to implement. For student-level projects and real-time applications, Naive Bayes provides a good balance between accuracy and performance. Based on these findings, the proposed system uses Naive Bayes with Count Vectorizer for recruitment scam detection

3.Proposed Methodology

The proposed system follows multiple stages to detect recruitment scams effectively.

1. Data Collection

A dataset containing both genuine and fraudulent job descriptions is collected. The dataset includes labeled examples that help the model learn patterns associated with each class.

2. Data Preprocessing

The collected job descriptions are cleaned to remove unnecessary characters, numbers, and symbols. The text is converted into lowercase format to maintain uniformity. This step improves the quality of input data and ensures better model performance.

3. Feature Extraction using Count Vectorizer

Count Vectorizer converts text data into numerical form by counting the frequency of words. Each unique word in the dataset becomes a feature. The output is a matrix representing word frequency in each job description. This numerical representation is used as input to the classifier.

4. Model Training using Naive Bayes

The Naive Bayes classifier is trained using the processed dataset. It calculates the probability of a job description belonging to either the Scam or Real category. The model learns patterns such as frequently occurring words in scam messages.

5. Prediction and Output

When a new job description is entered, the system preprocesses the text and transforms it using the saved Count Vectorizer. The trained Naive Bayes model then predicts the class label and provides a confidence score.

4. System Architecture

The system architecture consists of several interconnected components:

- User Interface (Streamlit Application)
- Text Preprocessing Module
- Feature Extraction Module (Count Vectorizer)
- Classification Module (Naive Bayes)

- Output and Visualization Module

The process begins when the user enters a job description in the Streamlit interface. The input text is sent to the preprocessing module where cleaning operations are performed. The cleaned text is converted into numerical features using Count Vectorizer. These features are passed to the Naive Bayes classifier, which predicts whether the job description is Scam or Real. The result is then displayed to the user along with a confidence score.

5.Implementation

The system is implemented using Python programming language. The Scikit-learn library is used to implement Count Vectorizer and Naive Bayes classifier. The trained model is saved as a file and reused during prediction to reduce processing time.TheStreamlit framework is used to create a simple and interactive web interface. Users can paste job descriptions into the text area and click the “Detect Scam” button to receive instant results. The interface also displays prediction result, confidence percentage, and highlighted suspicious words.The implementation ensures that the system works efficiently and provides quick responses. Since the model is pre-trained, predictions are generated within seconds

6.Results and Discussion

The model was evaluated using test data containing both real and scam job descriptions. The Naive Bayes classifier showed consistent performance in correctly classifying most job messages. The use of Count Vectorizer allowed the model to identify frequently used scam-related words such as “registration fee,” “urgent hiring,” and “easy income.”The system provides clear output with prediction result and confidence score. This helps users understand the reliability of the prediction. The results indicate that Naive Bayes is suitable for text-based scam detection tasks, especially when computational simplicity and speed are required.Although the system performs well, accuracy can be further improved by increasing the size of the dataset and including more diverse job examples.

7. Advantages of the Proposed System

- Automatically detects fake job postings
- Uses machine learning instead of fixed keyword rules
- Provides fast and accurate results
- Simple and user-friendly interface

- Can be retrained with updated data
- Low computational cost compared to complex models.

8. Conclusion

This research presents an Intelligent Recruitment Scam Detection System using Naive Bayes and Count Vectorizer. The system effectively classifies job descriptions into Scam or Real categories based on learned patterns. The implementation using Python and Streamlit makes the system practical and easy to use.The results demonstrate that Naive Bayes is an efficient algorithm for text classification problems such as recruitment scam detection. The proposed system provides a useful tool to help job seekers avoid fraudulent job offers and enhance online safety.

9.Future Scope

In the future, the system can be enhanced by increasing the dataset size to improve model accuracy. Additional preprocessing techniques such as stop-word removal and stemming can be applied for better feature representation. The system can also be integrated with job portals to provide real-time scam detection. Support for multiple languages and database integration can further improve usability and scalability.

Appendices

References

1.T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill Education, 1997.

2. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

3. S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.

4. K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.

5. F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

6. J. Rennie, L. Shih, J. Teevan, and D. Karger, "Tackling the poor assumptions of Naive Bayes text classifiers," in Proceedings of the 20th International Conference on Machine Learning (ICML), 2003, pp. 616–623.

7. Scikit-learn Developers, "Scikit-learn: Machine Learning in Python," Available: https://scikit-learn.org

8. A. McCallum and K. Nigam, “A comparison of event models for Naive Bayes text classification,” in AAAI Workshop on Learning for Text Categorization, 1998.

9. S. Raschka and V. Mirjalili, Python Machine Learning. Birmingham, U.K.: Packt Publishing, 2017.