As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology
Volume 10, Issue 04, pages 294-299.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue04.html
DOI: https://doi.org/10.46501/IJMTST1004043

Machine Learning-Powered Web Application for Predicting and Identifying Fake Job Listing

Dr. Y. Jayababu[1] , Vegisetti Surya Satish[2], Vasa Vanisri Naga Ratna Tejaswini[2], Manepalli Reethu[2], Gudala Praveen Kumar [2], Annabathula Venkata Satya Prabash[2]

[1]Professor, Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.
[2]Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India

To Cite this Article
Dr. Y. Jayababu , Vegisetti Surya Satish, Vasa Vanisri Naga Ratna Tejaswini, Manepalli Reethu, Gudala Praveen Kumar, Annabathula Venkata Satya Prabash, Machine Learning-Powered Web Application for Predicting and Identifying Fake Job Listing, International Journal for Modern Trends in Science and Technology, 2024, 10(04), pages. 294-299. https://doi.org/10.46501/IJMTST1004043

Article Info
Received: 06 April 2024; Accepted: 18 April 2024; Published: 26 April 2024.

ABSTRACT

To dodge false work posts on the web, the paper recommends a robotized framework that utilizes machine learning-based classification methods. Different classifiers are utilized to approve false posts on the web, and the comes about are compared to distinguish the best work trick location strategy. It makes a difference distinguish fake work notices among a huge number of posts. For identifying fake work notices, two major classifier sorts are considered: single classifiers and outfit classifiers. Be that as it may, test comes about propose that outfit classifiers outflank single classifiers in recognizing tasks.

Keywords: Fake Job, Online Recruitment, Machine Learning, Ensemble Approach

1. INTRODUCTION

Business trick is one of the genuine issues in later times tended to in the space of Online Enrolment Fakes (ORF) [1]. In later days, numerous companies favor to post their opening online so that these can be gotten to effectively and opportune by the job-seekers. In any case, this purposeful may be one sort of trick by the extortion individuals since they offer business to job-seekers in terms of taking cash from them. False work notices can be posted against a presumed company for abusing their validity. These false work post location draws a great consideration for getting an robotized apparatus for distinguishing fake employments and announcing them to individuals for maintaining a strategic distance from application for such occupations.

For this reason, machine learning approach is connected which utilizes a few classification calculations for recognizing fake posts. In this case, a classification device separates fake work posts from a bigger set of work notices and alarms the client. To address the issue of recognizing tricks on work posting, directed learning calculation as classification methods are considered at first. A classifier maps input variable to target classes by considering preparing information. Classifiers tended to in the paper for distinguishing fake work posts from the others are portrayed briefly. These classifiers-based forecast may be broadly categorized into -Single Classifier based Forecast and Gathering Classifiers based Prediction.

A. Single Classifier based Prediction

Classifiers are trained for predicting the unknown test cases. The following classifiers are used while detecting fake job posts

a) Naive Bayes Classifier-

The Naive Bayes classifier [2] is a supervised classification implementation that makes use of the Bayes Theorem of tentative liability. This classifier's judgment is relatively effective in practice, even if its probability estimates are erroneous. This classifier produces a very promising result in the following script: when the characteristics are independent or entirely functionally dependent. The delicacy of this classifier isn't related to point dependences, but rather to the quantum of information loss of the class due to the independence supposition, which is demanded to predict the delicacy.

b) Multi-Layer Perceptron Classifier -

Multi-layer perceptron's [4] can be utilized as supervised classification tools if the training parameters are tuned. For a particular problem, the number of hidden layers in a multilayer perceptron and the number of nodes in each layer may differ. The decision to choose the parameters depends on the training data and network design.

.

c) K-nearest Neighbor Classifier-

K - Nearest Neighbour Classifiers,[5] also known as lazy learners, identify objects based on their closest proximity to training samples in the feature space. When deciding on a class, the classifier evaluates the nearest k objects. The key issue of this categorization strategy is choosing the suitable value of k [5].

.

d) Decision Tree Classifier-

A Decision Tree (DT) is a classifier [6] that makes use of tree-like structures. It gets knowledge of classification. Each target class is denoted as a leaf node of DT, while non-leaf nodes of DT are utilized as decision nodes to specify a specific test. The results of those tests are determined by either branch of that decision node. Starting at the root, this tree progresses until it reaches a leaf node. It is the process of extracting categorization results from a decision tree. Decision tree learning is a technique used for spam filtering. By installing and training this model, it will be possible to forecast the target depending on specific criteria [7].

2. LITERATURE SURVEY

TITLE: "An Intelligent Model for Online Recruitment Fraud Detection,"

This work aims to prevent privacy violations and financial losses for individuals and organizations by developing a reliable model for detecting fraud exposure in online recruitment environments. This study makes a significant addition by developing a reliable detection model for online recruitment fraud (ORF) utilizing an ensemble technique based on the Random Forest classifier. The detection of online recruitment fraud differs from other types of electronic fraud detection in that it is more recent and there are fewer studies on the subject. The researcher proposed the detection model to meet the study's aims. The support vector machine approach is used for feature selection, while an ensemble classifier based on Random Forest is used for classification and detection. A freely available dataset titled Employment Scam Aegean , The model is applied using the EMSCAD dataset. Prior to selecting and classifying, a pre-processing step was performed. The results indicated an accuracy of 97.41%. Furthermore, the findings identified the primary features and important variables in identification, such as having a company biography, a corporate logo, and an industry feature [11].

TITLE: An Empirical Study of the Naïve Bayes Classifier
An empirical study of the naive Bayes classifier,

The naive Bayes classifier considerably simplifies learning by presuming that characteristics are independent within a class. Although independence is often a bad assumption, naïve Bayes frequently outperforms more advanced classifiers. Our overarching goal is to understand the data qualities that influence the performance of naive Bayes. Our approach employs Monte Carlo simulations, allowing for a systematic examination of categorization performance across multiple classes of randomly generated issues. We investigate the impact of distribution entropy on classification error and find that low-entropy feature distributions produce good naive Bayes performance. We also demonstrate that naive Bayes works well for certain nearly-functional feature dependencies, thus reaching its best performance in two opposite cases: fully independent features (as expected) and functionally dependent features.

(This is shocking). Another surprise finding is that naive Bayes' accuracy is not directly proportional to the degree of feature dependencies assessed as class-conditional mutual information between features. Instead, a better predictor of naive Bayes accuracy is the amount of class knowledge lost due to the independence assumption [7].

TITLE: Bayes's Theorem and the Analysis of Binomial Random Variables,

The author presents a fairly practical application of Bayes' theorem to the analysis of binomial random variables. Previous works (Walters, 1985; Walters, 1986a) have established the technique's reliability for one or two random variables, and an extension of the approach to multiple random variables is detailed. Two biometric examples are provided to demonstrate the method.

TITLE: Multilayer perceptron's for classification and regression,

We discuss the multilayer perceptron's theory and application. We intend to discuss a variety of issues that are relevant in terms of applying this method to real-world challenges. A variety of examples are provided to demonstrate how the multilayer perceptron compares to alternative, conventional methodologies. The application disciplines of classification and regression are particularly highlighted. Questions of implementation, such as multilayer perceptron architecture, dynamics, and associated features, are addressed. Recent work, notably in the areas of discriminant analysis and function mapping, are referenced [12].

TITLE: A Survey on Decision Tree Algorithms of Classification in Data Mining,

As computer and computer network technology advance, the volume of data in the information sector continues to grow. It is vital to examine this massive volume of data and extract relevant information from it. Data mining refers to the process of extracting meaningful knowledge from large sets of incomplete, noisy, fuzzy, and random data. The decision tree classification technique is one of the most common data mining techniques. In decision trees, the divide and conquer technique is utilized as a fundamental learning mechanism. A decision tree is a structure consisting of a root node, branches, and leaf nodes. Each internal node represents a test on an attribute, each branch represents the result of a test, and each leaf node has the class label. At the top of the tree is the root node. This study focuses on the features, difficulties, benefits, and drawbacks of the different decision tree algorithms (ID3, C4.5, and CART).

.

3. SYSTEM ANALYSIS

A. EXISTING SYSTEM

Numerous studies indicate that in the field of online fraud detection, the identification of fake news, email spam, and review spam have received particular attention.

1. **Review Spam Detection:**

Individuals frequently share their opinions on the things they buy on internet forums. It might help other buyers make product decisions. Since spammers can now manipulate reviews to increase their profits, methods for identifying these kinds of reviews must be developed. This can be accomplished by applying Natural Language Processing (NLP) to extract features from the reviews. Following that, these features are subjected to machine learning techniques. One substitute for machine learning methods that employ a corpus or dictionary to get rid of spam reviews could be lexicon-based methods.

2. **Email Spam Detection**

Unwanted bulk emails, sometimes known as spam emails, frequently end up in user mailboxes. This could result in bandwidth usage and an inevitable storage issue. Spam filters based on Neural Networks are used by Gmail, Yahoo Mail, and Outlook service providers to eliminate this issue. The following approaches to email spam detection are taken into consideration: adaptive spam filtering, content-based filtering, case-based filtering, heuristic-based filtering, memory or instance-based filtering.

3. **Fake News Detection**

Social media fake news is typified by echo chamber effects and malicious user profiles. Three viewpoints are essential to the basic research of fake news detection: the writing of fake news, the dissemination of fake news, and the relationship between a user and fake news. To identify fake news, social context and news content-related features are retrieved, and machine learning models are then used.

DISADVANTAGES OF THE EXISTING SYSTEM

The machine learning models may not generalize effectively to new, unknown data if the dataset used to train them is unbalanced, with a comparatively smaller number of positive (fraudulent) cases compared to negative (legal job listings) ones.

Difficulties in Feature Engineering: It can be difficult to design features for the machine learning model that accurately describe job ads. The performance of the model could be hampered if significant features are absent or inadequately represented.

Adaptability to Changing Scams: New strategies may arise when fraudulent operations change over time. It's possible that false negatives will result from the current system's slow ability to adjust to new scam kinds.

Explainability and Interpretability: Certain machine learning models, particularly complicated ones like ensemble classifiers, may lack transparency and interpretability. Understanding why a model generates a specific prediction is important, especially in sensitive domains such as fraud detection.

Scalability: When faced with a huge number of job listings, the existing system's performance may decline. Scalability concerns may develop if the system is not designed to manage a large volume of data efficiently.

Dependency on Training Data: The quality and representativeness of training data have a significant impact on the effectiveness of machine learning models. If the training data does not fully represent the range of fraudulent job ads, the model may not perform effectively in real-world circumstances.

Computational Resources: Complex machine learning models, particularly ensemble classifiers, may necessitate substantial computational resources for training and inference. This may be a limitation in terms of both time and hardware.

False Positives: The system may produce false positives, classifying legitimate job listings as fraudulent. This might cause user irritation and a lack of faith in the system. Regulatory Compliance: Depending on the application domain, there may be legal and ethical concerns with the use of machine learning models for fraud detection. Ensuring compliance with applicable regulations is critical.

B. PROPOSED SYSTEM

The proposed system aims to improve the ability of job placement platforms to detect and mitigate fake job postings by integrating advanced machine learning techniques. The system leverages state-of-the-art classification algorithms, improves adaptability to evolving fraud, improves the feature engineering process, and ensures superior scalability to efficiently handle large volumes of job postings. attempts to overcome the limitations of this approach. Additionally, the proposed system focuses on explainability and interpretability to provide a clearer understanding of the decision-making process behind fraud detection. To achieve optimal results, the system considers the advantages of ensemble classifiers over individual classifiers, as shown by experimental results. The overall goal is to provide a robust and reliable tool that not only effectively identifies fraudulent job advertisements, but also minimizes false positives, thereby increasing user trust in recruitment platforms. is. Additionally, the proposed system aims to maintain regulatory compliance and address legal and ethical issues related to the use of machine learning models in fraud detection applications. With these improvements, the proposed system is expected to establish a new standard in detecting fake job applications and contribute to safer and more reliable online job searches.

ADVANTAGES OF THE PROPOSED SYSTEM

Enhanced Fraud Detection Accuracy: The proposed approach uses advanced machine learning techniques, such as ensemble classifiers, to dramatically enhance the accuracy of detecting fake job postings. By using complex algorithms, the system can better distinguish patterns and abnormalities associated with frauds, resulting in a higher precision in recognizing phony job advertisements.

Adaptability to Emerging Scams: Unlike existing systems, the proposed solution is intended to respond dynamically to evolving fraud tactics. Through constant learning and upgrades, the system can efficiently recognize and combat new sorts of fraudulent activity in the ever-changing field of online job recruitment.

Improved Scalability: The system is designed for scalability, allowing for efficient processing and analysis of large numbers of job posts. This is especially beneficial for job recruitment platforms with high user engagement, allowing the system to handle larger data volumes without sacrificing efficiency.

Enhanced Explainability and Interpretability: The suggested method prioritizes transparency and interpretability, providing explicit insights into the decision-making process of machine learning models. This functionality not only helps to develop trust in the system, but also allows users and platform administrators to understand why a certain job posting is labelled as potentially fraudulent.

Reduced False Positives: The suggested method intends to reduce false positives by fine-tuning machine learning models and feature engineering techniques. This is critical for maintaining a great user experience on the job recruitment platform, preventing legitimate job ads from being wrongly labelled as fraudulent, and thereby encouraging improved user confidence in the system's authenticity.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

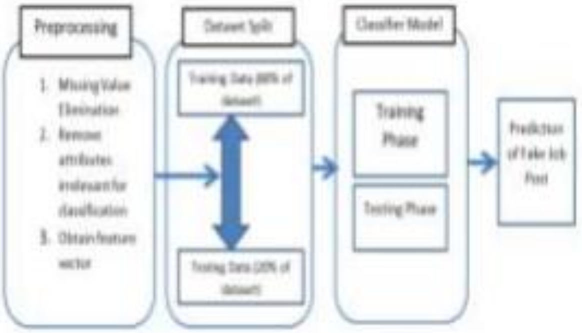

Preprocessing
1. Missing Value Elimination
2. Remove attributes irrelevant for classification
3. Obtain feature vector
Dataset Split
Training Data (80% of dataset)
Testing Data (20% of dataset)
Classifier Model
Training Phase
Testing Phase
Prediction of Fake Job Post

Fig 1. System Architecture

5. SYSTEM IMPLEMENTATION

MODULES

Data Preprocessing Module:

This module cleans and prepares raw data for analysis. It entails duties including resolving missing numbers, deleting extraneous information, and standardizing data formats. Data preprocessing ensures the quality and consistency of the input data for future machine learning model training.

Feature Engineering Module:

Feature engineering is a critical step toward improving the effectiveness of machine learning models. This module consists of choosing and manipulating significant features from the dataset in order to offer meaningful input to the classifiers. Text analysis and the extraction of key job-related attributes are used to generate a feature set that encapsulates the essential characteristics of job posts.

Machine Learning Classification Module:

This fundamental module trains and deploys machine learning classifiers. It uses both single and ensemble classifiers to assess feature-rich data and estimate the validity of job advertisements. This section includes the process of selecting classifiers, training models, and optimizing them.

Model Evaluation and Comparison Module:

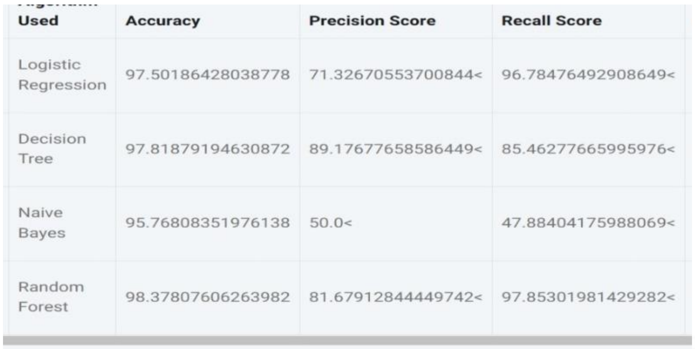
After training the classifiers, this module assesses their performance using metrics including accuracy, precision, recall, and F1-score. It also allows for a comparison analysis of various classifiers to choose the most effective model for detecting fake job postings. Model evaluation is crucial for fine-tuning parameters and choosing the best-performing method.

User Interface and Reporting Module:

This module focuses on creating a user-friendly interface for interacting with the system. It has tools that allow users to submit job listings for analysis and examine the outcomes. Furthermore, the module offers thorough reports on categorization results, indicating if a job ad is tagged as potentially fake or authentic. Clear and intuitive visualizations may also be included to assist user understanding of the system's findings.

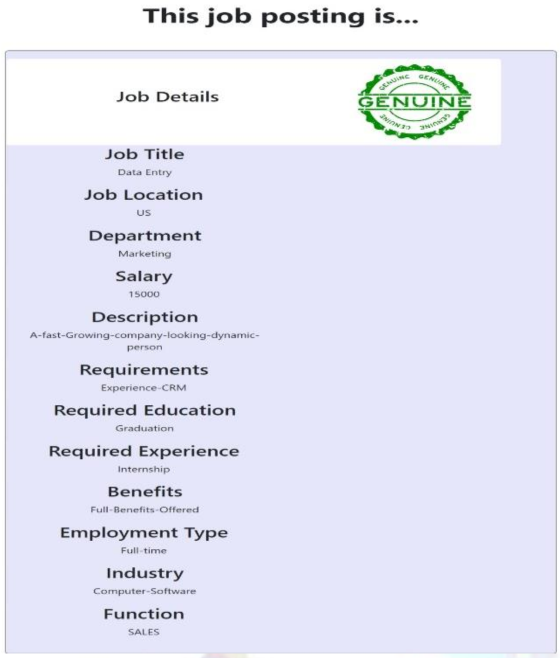
6. EXPERIMENTAL RESULTS

All of the above-mentioned classifiers are trained and tested to detect bogus job posts using a dataset that includes both false and authentic posts. The next table compares the classifiers in terms of assessing metrics, and Table 2 shows the results for classifiers that use ensemble approaches. Figure 2 shows the overall performance of all classifiers in terms of accuracy, f1-score, Cohen-kappa score, and MSE.



Algorithm Used	Accuracy	Precision Score	Recall Score
Logistic Regression	97.50186428038778	71.32670553700844<	96.78476492908649<
Decision Tree	97.81879194630872	89.17677658586449<	85.46277665995976<
Naive Bayes	95.76808351976138	50.0<	47.88404175988069<
Random Forest	98.37807606263982	81.67912844449742<	97.85301981429282<

fig 2. performance comparison chart for ensemble classifier-based prediction



This job posting is...

Job Details

Job Title
Data Entry

Job Location
US

Department
Marketing

Salary
15000

Description
A-fast-Growing-company-looking-dynamic-person

Requirements
Experience-CRM

Required Education
Graduation

Required Experience
Internship

Benefits
Full-Benefits-Offered

Employment Type
Full-time

Industry
Computer-Software

Function
SALES

Fig 3. Result For predicting Fake job posts

7. CONCLUSION AND FUTURE WORK

Employment scam identification can help job searchers acquire only authentic offers from employers. This research proposes numerous machine learning methods as countermeasures to detect employment scams. The supervised technique is used to demonstrate the utilization of several classifiers for job fraud detection. Experimental data show that the Random Forest classifier outperforms its peer classification technology. The proposed strategy obtained accuracy of 98.27%, which is significantly higher than the existing methods.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Bandar Alghamdi, Fahad Alharby, “An Intelligent Model for Online Recruitment Fraud Detection”, Journal of Information Security, 2019, pp. 155-176.

[2] Tao Jiang, Jian ping li, Amin ul Haq, Abdus labor, and Amjad al, “A Novel Stacking Approach for Accurate Detection of Fake News”, Vol. 9, 2021, pp. 22626-22639.

[3] Karri sai Suresh reddy, karri Lakshmana reddy, “fake job recruitment detection”, JETIR August 2021, Vol. 8, pp. d443-d448.

[4] Tulus Suryanto, Robbi Rahim, Ansari Saleh Ahmar, “Employee Recruitment Fraud Prevention with the Implementation of Decision Support System”, Journal of Physics Conference Series, 2018, pp.1-11.

[5] C. Jagadeesh, Dr. Pravin R Kshirsagar, G. Sarayu, G.Gouthami, B.Manasa, “Artificial intelligence based Fake Job Recruitment Detection Using Machine Learning Approach”, Journal of Engineering Sciences, Vol. 12, 2021, pp. 0377-9254.

[6] Lal, Sangeeta, Rishabh Jiaswal, Neetu Sardana, Ayushi Verma, Amanpreet Kaur, and Rahul Mourya. "ORFDetector: ensemble learning based online recruitment fraud detection." In 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. 1-5. IEEE, 2019.

[7] Samir Bandyopadhyay, Shawni Dutta, “Fake Job Recruitment Detection Using Machine Learning Approach”, International Journal of Engineering Trends and Technology (IJETT),Vol. 68, 2020, pp. 48- 53

[8] George Tsakalidis, Graduate Student Member, IEEE, and Kostas Vergidis, “A Systematic Approach Toward Description and Classification of Cybercrime Incidents”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 49, 2019, pp. 1-20

[9] Andrii Shalaginov, Jan William Johnsen, Katrin Franke, “Cyber Crime Investigations in the Era of Big Data”, IEEE International Conference on Big Data, 2017, pp. 3672-3676.

[10] Sokratis Vidros, Constantinos Kolias, Georgios Kambourakis and Leman Akoglu, “Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset”, Future Internet 2017, pp. 2-19.

[11] Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19, no. 1 (2017): 22-36.

[12] Devsmit Ranparia; Shaily Kumari; Ashish Sahani, ”Fake Job Prediction using Sequential Network”, IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp.339-343

[13] Syed Mahbub, Eric Pardede, “Using Contextual Features for Online Recruitment Fraud Detection”, 27th International Conference on Information Systems Development, 2018.

[14] Najma Imtiaz Ali, Suhaila Samsuri, Muhamad Sadry, Imtiaz Ali Brohi, Asadullah Shah, “Online Shopping Satisfaction in Malaysia: A Framework for Security, Trust and Cybercrime”, 6th International Conference on Information and Communication Technology for The Muslim World, 2016, pp. 194-198.

[15] Vidros, Sokratis; Kolias, Constantinos; Kambourakis, Georgios, “Online recruitment services: another playground for fraudsters”, Computer Fraud & Security, 2016, pp. 8-13.