ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

IDENTIFYING FAKE JOBS USING MACHINE LEARNING TECHNIQUES

Dr.Nazimunissa[1], A.Vivek, S.Jatin Sagar[2], Akhil[3], K.Manideep[4]

Associate Professor, Department of Computer Science and Engineering[1]

Student, Department of Computer Science and Engineering[2,3,4]

Sree Dattha Institute of Engineering and Science, Sheriguda, Telangana. [1,2,3,4,5]

ABSTRACT

To avoid fraudulent post for job in the internet, an automated tool using machine learning based classification techniques is proposed in the paper. Different classifiers are used for checking fraudulent post in the web and the results of those classifiers are compared for identifying the best employment scam detection model. It helps in detecting fake job posts from an enormous number of posts. Two major types of classifiers, such as single classifier and ensemble classifiers are considered for fraudulent job posts detection. However, experimental results indicate that ensemble classifiers are the best classification to detect scams over the single classifiers.

Index-Terms: Employment scam detection, machine learning, classification techniques, ensemble classifiers, single classifiers, fraudulent job posts.

I.INTRODUCTION

Employment scams are a serious issue within the domain of Online Recruitment Frauds (ORF). Many companies now prefer to post their vacancies online for easy and timely access by job-seekers. However, this convenience can be exploited by fraudsters who offer fake employment opportunities in exchange for money. Fraudulent job advertisements can tarnish the reputation of legitimate companies and deceive job-seekers. Detecting these fraudulent job posts has become crucial, necessitating an automated tool to identify and report fake jobs. To achieve this, a machine learning approach employing various classification algorithms is used to recognize fake posts. This classification tool isolates fraudulent job posts from a larger set of advertisements

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

and alerts users. To tackle the problem of identifying scams in job postings, supervised learning algorithms as classification techniques are initially considered. A classifier maps input variables to target classes using training data. The classifiers discussed in the paper for identifying fake job posts are categorized into two main types: Single Classifier-based Prediction and Ensemble Classifiers-based Prediction. Each type is briefly described and compared for their effectiveness in detecting fake job posts.

II.LITERATURE SURVEY

- Bandar Alghamdi and Fahad Mohammed Al-harby proposed research attempts to prohibit privacy and loss of money for individuals and organizations by creating a reliable model which can detect fraud exposure in online recruitment environments. This research presents a major contribution represented in a reliable detection model using an ensemble approach based on a Random Forest classifier to detect Online Recruitment Fraud (ORF). The detection of Online Recruitment Fraud is characterized by its modernity and the scarcity of studies on this concept. The researchers proposed the detection model to achieve the objectives of this study. For feature selection, the support vector machine method is used, and for classification and detection, an ensemble classifier using Random Forest is employed. A freely available dataset called the Employment Scam Aegean Dataset (EMSCAD) is used to apply the model. A pre-processing step had been applied before the selection and classification adoptions. The results showed an obtained accuracy of 97.41%. Furthermore, the findings presented the main features and important factors in the detection process, including having a company profile feature, a company logo feature, and an industry feature.

- Irina Rish proposed the naive Bayes classifier greatly simplifies learning by assuming that features are independent given the class. Although independence is generally a poor assumption, in practice, naive Bayes often competes well with more sophisticated classifiers. Our broad goal is to understand the data characteristics which affect the performance of naive Bayes. Our approach uses Monte Carlo simulations that allow a systematic study of classification accuracy for several classes of randomly generated problems. We analyze the impact

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

of distribution entropy on classification error, showing that low-entropy feature distributions yield good performance of naive Bayes. We also demonstrate that naive Bayes works well for certain nearly-functional feature dependencies, thus reaching its best performance in two opposite cases: completely independent features (as expected) and functionally dependent features (which is surprising). Another surprising result is that the accuracy of naive Bayes is not directly correlated with the degree of feature dependencies measured as the class-conditional mutual information between the features. Instead, a better predictor of naive Bayes accuracy is the amount of information about the class that is lost because of the independence assumption.

- Dr. D. E. Walters proposed a very practical application of Bayes's theorem for the analysis of binomial random variables. Previous papers (Walters, 1985; Walters, 1986a) have already demonstrated the reliability of the technique for one or two random variables, and the extension of the approach to several random variables is described. Two biometrical examples are used to illustrate the method.

- Finn Murtagh proposed a review of the theory and practice of the multilayer perceptron. We aim to address a range of issues which are important from the point of view of applying this approach to practical problems. A number of examples are given, illustrating how the multilayer perceptron compares to alternative, conventional approaches. The application fields of classification and regression are especially considered. Questions of implementation, i.e., of multilayer perceptron architecture, dynamics, and related aspects, are discussed. Recent studies, which are particularly relevant to the areas of discriminant analysis and function mapping, are cited.

- Himani Sharma and Sunil Kumar proposed that as computer technology and computer network technology develop, the amount of data in the information industry is getting higher and higher. It is necessary to analyze this large amount of data and extract useful knowledge from it. The process of extracting useful knowledge from a huge set of incomplete, noisy, fuzzy, and random data is called data mining. The decision tree classification technique is one of the most popular data mining techniques. In decision tree, divide and conquer technique is used as

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

a basic learning strategy. A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. This paper focuses on the various algorithms of decision trees (ID3, C4.5, CART), their characteristics, challenges, advantages, and disadvantages.

III.EXISTING SYSTEM

According to several studies, Review spam detection, Email Spam detection, Fake news detection have drawn special attention in the domain of Online Fraud Detection.

A.Review Spam Detection

People often post their reviews online forum regarding the products they purchase. It may guide other purchaser while choosing their products. In this context, spammers can manipulate reviews for gaining profit and hence it is required to develop techniques that detects these spam reviews. This can be implemented by extracting features from the reviews by extracting features using Natural Language Processing (NLP). Next, machine learning techniques are applied on these features. Lexicon based approaches may be one alternative to machine learning techniques that uses dictionary or corpus to eliminate spam reviews.

A. Email Spam Detection

Unwanted bulk mails, belong to the category of spam emails, often arrive to user mailbox. This may lead to unavoidable storage crisis as well as bandwidth consumption. To eradicate this problem, Gmail, Yahoo mail and Outlook service providers incorporate spam filters using Neural Networks. While addressing the problem of email spam detection, content based filtering, case based filtering, heuristic based filtering, memory or instance based filtering, adaptive spam filtering approaches are taken into consideration.

A.Fake News Detection

Fake news in social media characterizes malicious user accounts, echo chamber effects. The fundamental study of fake news detection relies on three perspectives- how fake news is written, how fake news spreads, how a user is related to fake news. Features related to news content and social context are extracted and a machine learning models are imposed to recognize fake news.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

IV. PROPOSED SYSTEM

The target of this study is to detect whether a job post is fraudulent or not. Identifying and eliminating these fake job advertisements will help the job seekers to concentrate on legitimate job posts only. In this context, a dataset from Kaggle is employed that provides information regarding a job that may or may not be suspicious.

A.Implementation of Classifiers

In this framework classifiers are trained using appropriate parameters. For maximizing the performance of these models, default parameters may not be sufficient enough. Adjustment of these parameters enhances the reliability of this model which may be regarded as the optimised one for identifying as well as isolating the fake job posts from the job seekers.

B.Performance Evaluation Metrics

While evaluating performance skill of a model, it is necessary to employ some metrics to justify the evaluation. For this purpose, following metrics are taken into consideration in order to identify the best relevant problem-solving approach. Accuracy is a metric that identifies the ratio of true predictions over the total number of instances considered. However, the accuracy may not be enough metric for evaluating model‘s performance since it does not consider wrong predicted cases. If a fake post is treated as a true one, it creates a significant problem. Hence, it is necessary to consider false positive and false negative cases that compensate to misclassification. For measuring this compensation, precision and recall is quite necessary to be considered.

V.SYSTEM ARCHITECTURE

For better understanding of the target as a baseline, a multistep procedure is followed for obtaining a balanced dataset. Before fitting this data to any classifier, some pre-processing techniques are applied to this dataset. Pre-processing techniques include missing values removal, stop-words elimination, irrelevant attribute elimination and extra space removal. This prepares the dataset to be transformed into categorical encoding in order to obtain a feature vector. This feature vectors are fitted to several classifiers. The following diagram Fig. 1 depicts a description of the working paradigm of a classifier for prediction.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

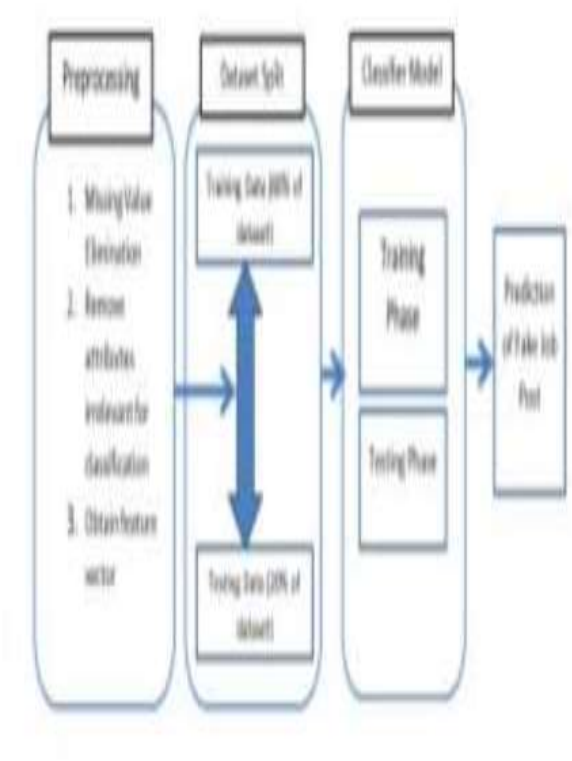

Figure .1 System Architecture

VI.IMPLEMENTATION

MODULES

- **Admin Module:**

The Admin module allows administrators to manage the system efficiently. The first functionality is the login, which provides secure access to the admin panel. Once logged in, administrators can handle user management, including viewing pending user registrations and accessing a list of all registered users. Another critical feature is the Fake Job section, where admins can upload datasets related to fraudulent job listings and view the existing datasets.

The Algorithm section is essential for analyzing data using various machine learning algorithms. Administrators can choose from several algorithms, including the Support Vector Machine (SVM) Algorithm, Decision Tree Algorithm, Naïve Bayes Algorithm, K-NN Bayes Algorithm, and Random Forest Algorithm. Additionally, the Graph Analysis section offers a comparison graph to visualize and compare the performance and results of these algorithms.

- **User Module:**

The User module is designed for candidates who interact with the system. The first step for users is registration, where they provide necessary information to create an account. After registering, users can log in to the system using their credentials. The primary functionality for users is the Predict feature, which allows them to input data and receive predictions based on the algorithms implemented in the system. This feature helps users identify potential fraudulent job listings and make informed decisions.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Both modules are integral to ensuring the system's functionality, providing administrators with robust tools for data management and analysis, and offering users valuable insights to protect against online recruitment fraud.

VII.RESULTS

Figure .2

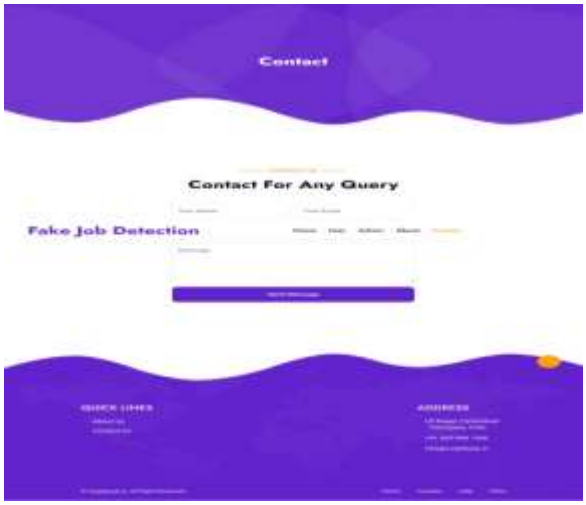

Figure .3

Figure .4

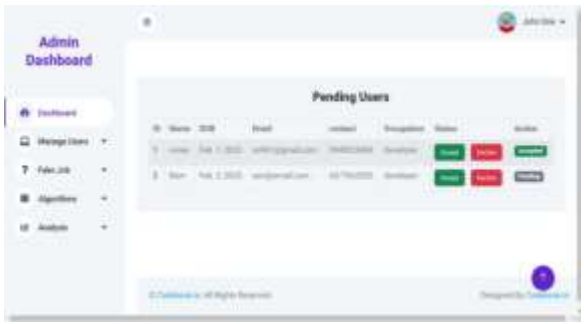

Figure .5

Figure .6

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

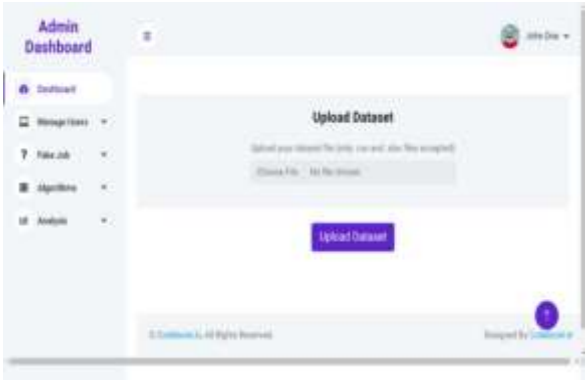

Figure .7

Figure .8

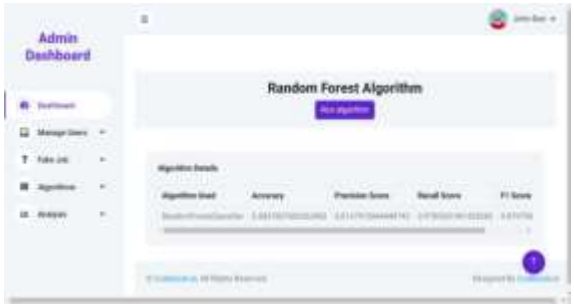

Figure .9

Figure .10

Figure .11

Figure .12

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Figure .13

VIII.CONCLUSIONS

Employment scam detection will guide job-seekers to receive only legitimate offers from companies. To tackle employment scam detection, this paper proposes several machine learning algorithms as countermeasures. A supervised mechanism exemplifies the use of various classifiers for this purpose. Experimental results indicate that the Random Forest classifier outperforms its peers, achieving an accuracy of 98.27%, which is significantly higher than existing methods.

IX.FUTURE ENHANCEMENT

Future work can enhance the system by integrating more sophisticated machine learning algorithms and deep learning techniques to further improve accuracy and detection speed. Additionally, expanding the dataset to include more diverse and updated examples of employment scams will enhance the model's robustness. Implementing real-time detection and alert systems can provide immediate protection for job-seekers. Collaboration with job portals and integrating the system into their platforms can offer broader protection and reach. Lastly, exploring multilingual support can make the system more accessible to non-English speaking users, providing a more comprehensive solution to employment scam detection.

X.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.

[16] Sultana Umme Habiba, Md. Khairul Islam, Farzana Tasnim, "A Comparative Study on Fake Job Post Prediction Using Different Data mining Techniques", 2nd International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST), 2021, pp. 543-546.

[17] Sarvesh Tanwar, Thomas Paul, Kanwarpreet Singh, Mannat Joshi, Ajay Rana, "Classification and Impact of Cyber Threats in India: A review", 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020,pp. 129-135.

[18] Veena, K., and P. Visu. "Detection of cyber crime: An approach using the lie detection technique and methods to solve it." In 2016 International Conference on Information Communication and Embedded Systems (ICICES), pp. 1-6. IEEE, 2016.

[19] Gunjan, Vinit Kumar; Kumar, Amit; Avdhanam, Sharda, "A survey of cybercrime in India", 15th International Conference on Advanced Computing Technologies (ICACT), 2013, pp. 1–6.

[20] Thangiah, Murugan; Basri, Shuib; Sulaiman, Suziah, "A framework to detect cybercrime in the virtual environment", International Conference on Computer & Information Science (ICCIS), 2012, pp. 553–557