INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 149-154

A MACHINE LEARNING APPROACHES FOR FAKE JOB IDENTICATION

K. Swapna[1]

[1]Assistant professor, Department of Computer Science and Engineering, SVS Group of Institutions, Hanamkonda, Telangana

ABSTRACT

The study suggests a mechanised method utilising machine learning-based categorization algorithms to stop fake online job ads. These days, it's common practise for companies to publish their open positions online, where they may be easily accessed by prospective employees. However, con artists may be taking advantage of people looking for work by offering those jobs in exchange for payment. Many people fall for this scam, which causes them to lose a lot of money. By performing an exploratory data analysis on the data, we can distinguish between legitimate and fake job ads. A machine learning strategy is utilised to identify fake comments, which makes use of several classification techniques. Using historical examples of both fake and real job postings, the system would teach the model how to correctly categorise future job advertisements. To begin tackling the difficulty of identifying scammers on job advertisements, supervised learning algorithms as classification techniques can be investigated. It uses two or more machine learning algorithms to determine which is best at predicting whether or not a job ad's headline is authentic.

Keywords: Fake Job, job advertisements, Machine learning,

1. INTRODUCTION

Many people have lost their jobs or been forced to take on unpaid internships because of the recession and the effects of the corona virus. Con artists are waiting for an opportunity like this to strike. Many people are falling for these con artists who are taking advantage of their desperation after a remarkable incident. The goal of the vast majority of fraudsters is to gain access to the victim's private information. Personal information includes things like a person's home address, bank account number, and social security number. Scammers tempt their victims with a dream employment offer before asking for their money. Or, they may need the job seeker to make a financial commitment in exchange for a job offer. The high rate of unemployment has led to an increase in job-related scams. There are several online resources a recruiter can use to discover a suitable candidate. Sometimes, fraudulent recruiters would post a job on a job board with the sole intention of making a profit. This is a problem for a lot of different job boards. Later on, when looking for work, they switch to a new job platform, but this one is also frequented by scam recruiters. Therefore, it is vital to identify real job openings from those that are merely scams. One of the most serious issues addressed in the field of Online Recruitment Frauds (ORF) in recent years is employment fraud. In order to facilitate the search for new employees, many companies today post their available positions online. But this might be a sort of deception that the con artist uses. However, con artists may be taking advantage of people looking for work by offering them jobs in exchange for payment. This is a perilous issue that can be addressed with machine learning and NLP techniques. A machine learning strategy is utilised to identify fake comments, which makes use of several classification techniques. In this instance, fake job postings are identified from the whole pool of job postings, and the user is alerted. To begin addressing the difficulty of identifying con artists on job advertisements, supervised learning algorithms as classification techniques are being researched. A classifier learns how to relate input variables to output classes by analysing historical data. Classifiers developed in this research help identify fake advertisements for employment. Single-classifier predictions and ensemble-classifier predictions are two types of classifier-based forecasts.

2. LITERATURE SURVEY

Little work has been done to yet in the field of detecting fraudulent activity in online recruitment. Anti-phishing methods can identify fraudulent websites, countermeasures against opinion fraud can identify the posting of dishonest and misleading fake reviews, and email spam filtering can stop users from receiving unwanted promotional emails. In the field of online fraud detection, several studies have focused on review spam detection, email spam detection, and the identification of fake news.

A. Review Spam Detection- Purchasing experiences are widely discussed in online discussion groups. It could help other shoppers make more informed decisions. Since spammers in this environment can manipulate evaluations for financial gain, it's important to devise methods to identify fake feedback. To do this, Natural Language Processing (NLP) feature extraction can be used on the reviews. Machine learning methods are applied to these characteristics.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 149-154

B. Email Spam Detection

Spam emails, or unsolicited mass mailings, are a common occurrence in people's inboxes. A storage crisis and a strain on network bandwidth could come from this. Email clients like Gmail, Yahoo Mail, and Outlook employ Neural Network-based spam filters to deal with the issue. Email spam detection strategies include content-based filtering, case-based filtering, heuristic-based filtering, memory- or instance-based filtering, and adaptive spam filtering.

C. Fake News Detection

Malicious user accounts and the echo chamber effect are hallmarks of the spread of fake news on social media. How fake news is created, how fake news travels, and how a user is connected are the three pillars upon which the identification of fake news rests. Machine learning algorithms are used to recover features associated with the news content and social context in order to detect fake news.To the best of our knowledge, no one besides Vidros. has proposed a method for identifying cases of employment fraud. Unfortunately, they only employed a balanced dataset, and it has yet to be established how well prediction algorithms work on an imbalanced dataset. This highlights the importance of testing prediction models on an imbalanced dataset. In order to identify instances of online fraud, the recommended ORF Detector employs an ensemble-based approach. They used three different voting methods (average, majority, and maximum) on three different classifiers (J48, Logistic Regression, and Random Forest) to determine a baseline. The main drawback of this method is that it is limited to balanced datasets and yields subpar results.

3. PROPOSED SYSTEM

The proposed system is simply an ML classification model based on Decision Tree and Random forest algorithm to figure out if a job posting is real or fake. The model is taught to be as effective as possible by making the dataset part of a double-blind study and taking into account the different ways that jobs are posted on professional websites and other sites. This makes it much easier to find jobs and makes it so that people don't have to worry when they look for jobs online. There is a clear study of the dataset used, which makes it very useful.

The suggested plan is made up of three main steps: (i) Preparing the data, (ii) Choosing the traits, and (iii) Learning with help. In the data preparation step, the dataset is pre-processed (to fix any problems), features are extracted and calculated, and then the data is "normalised." After that, all of the qualities are taken into account as part of the selection process. A two-step plan is set up to help choose the best group of features. Third, supervised learning is used to train and test two ML models. Standard two evaluation measures are used to judge these models. After that, we have a good template for spotting job scams and frauds.

4. PURPOSE OF THE PROJECT

The goal of the project is to figure out if a job is fake and what amount of sales risk it poses by using different machine learning techniques, such as decision trees and random forests.

AIM

The goal of the project is to find out how organisation, job description, and type of pay affect fraud detection as a stand-alone model and how they affect fraud detection when they are all taken together.

5. OBJECTIVE

The goal of this project is to show how well different machine learning techniques, like Decision Tree and Random Forest, can predict fake jobs early on. ML techniques help identify and know things early on after data analysis.

The system is proposed to have the following modules along with functional requirements.

- ADDING CORPUS
- TOKENIZATION
- FEATURE EXTRACTION AND STOP WORDS
- MODEL TRAINING AND TESTING PHASE

ADDING CORPUS: This part of the programme will load all of the email records and divide them into training and testing data. This process will accept information in the '*.txt' format for individual email (Ham and Spam). This is to help you understand how to deal with real-world problems.

TOKENIZATION: Tokenization is the process of breaking up the lines in an email into their separate words. These tokens are saved in a collection, which is then used in the testing data to find every word in an email. This will help the computers figure out if the email is spam or ham.

FEATURE EXTRACTION AND STOP WORDS: This was used to get rid of words and characters that didn't belong in each email. It also made a bag of words that the algorithms could use to compare. The Scikit-learn tool "Count Vectorizer" gives each word or token a number and shows how many times it appears in an email. The

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

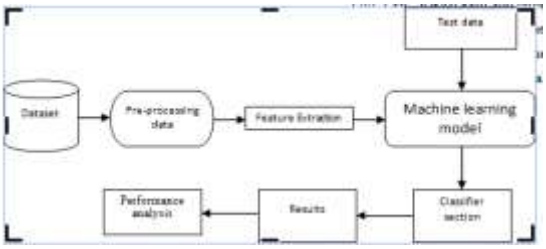
www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 149-154

instance is called to get rid of English "stopwords" like "a," "in," "the," "are," "as," "is," etc., because they don't tell much about whether an email is spam or not. The programme is then set up to learn the words in this case.

MODEL TRAINING AND TESTING PHASE: As was talked about throughout the study, supervised learning methods were used, and the model was trained with known data and tested with unknown data to predict the accuracy and other performance measures. K-Fold cross validation was used to get data that could be trusted. There are some problems with this method. For example, the trial data could be made up of all spam emails, or the training set could have most spam emails. This problem was fixed by Stratified K-fold cross validation, which splits the data and makes sure the distributed set has a good mix of spam and ham [50]. Last, Scikit-Learn and bio-inspired algorithms were used to try to improve the accuracy of ML models by adjusting the parameters. This lets you compare the Scikit-learn library with the bio-inspired algorithms.

ARCHITECTURE DIAGRAM



6. RESULTS ANALYSIS

Figure 1: Evalution of system

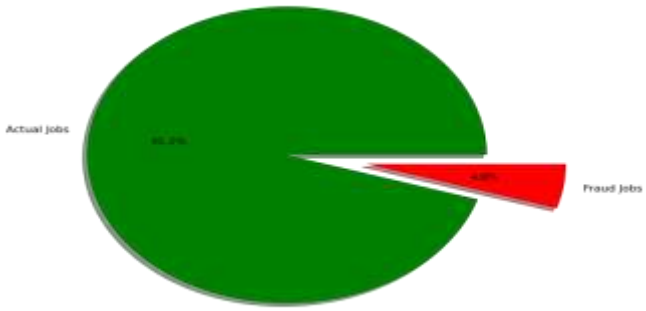


Figure 2: Distrubtion of Real and Fake Jobs

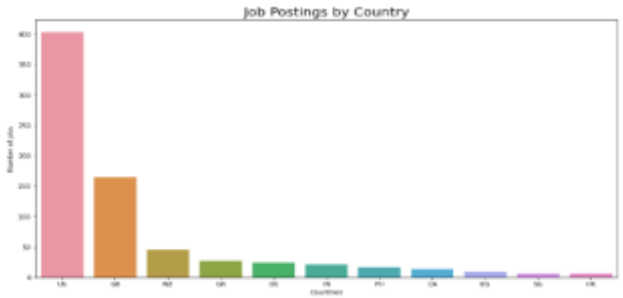


Figure 3: Countries with the most job openings

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 149-154

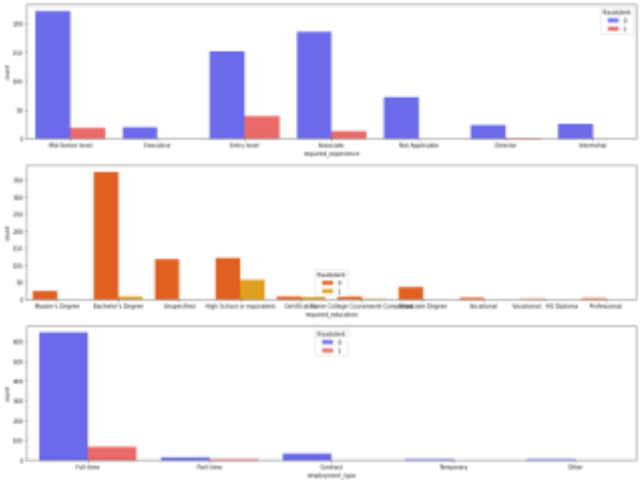


Figure 4: Observations on Required Experience, Education required & Employment type for the Fake / Real jobs

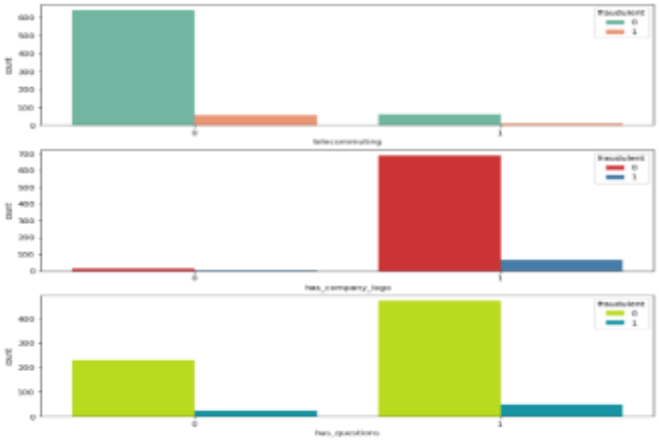


Figure 5 : Telecommuting, corporate logo, and fake/real work inquiries.

Common words in Fake Jobs

Decision Tree classifier report

	precision	recall	f1-score	support
0	0.99	0.96	0.97	5141
1	0.96	0.99	0.98	5068
accuracy			0.98	10209
macro avg	0.98	0.98	0.98	10209
weighted avg	0.98	0.98	0.98	10209

Random Forest Classifier report

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5141
1	0.99	0.99	0.99	5068
accuracy			0.99	10209
macro avg	0.99	0.99	0.99	10209
weighted avg	0.99	0.99	0.99	10209

7. CONCLUSION AND FUTURE WORK

Fake Job Postings on the Internet Prediction will make sure that companies only send job seekers real job offers. Several ways to stop Online Fake Job Postings are given as solutions in this Project. These methods use machine learning. In the end, we think that our app will be the best one out there. We worked on the problems, fixed them, and then made the changes to our project. Comparing the accuracy would help you decide if the dataset is a good representation of how hard it is to tell the difference between fake and real Jobs or not. If people are more accurate than the model, it could mean that we need to pick more fake jobs that aren't what they seem.

8. REFERENCES

[1] S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepakkumar, "Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms," Rev. GEINTECGESTAO Inov. E Tecnol., vol. 11, no. 2, pp. 642–650, 2021.

[2] Vellela, S.S., Balamanigandan, R. Optimized clustering routing framework to maintain the optimal energy status in the wsn mobile cloud environment. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-15926-5

[3] Vellela, S. S., Reddy, B. V., Chaitanya, K. K., & Rao, M. V. (2023, January). An Integrated Approach to Improve E-Healthcare System using Dynamic Cloud Computing Platform. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 776-782). IEEE.

[4] K. N. Rao, B. R. Gandhi, M. V. Rao, S. Javvadi, S. S. Vellela and S. Khader Basha, "Prediction and Classification of Alzheimer's Disease using Machine Learning Techniques in 3D MR Images," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 85-90, doi: 10.1109/ICSCSS57650.2023.10169550.

[5] VenkateswaraRao, M., Vellela, S., Reddy, V., Vullam, N., Sk, K. B., & Roja, D. (2023, March). Credit Investigation and Comprehensive Risk Management System based Big Data Analytics in Commercial Banking. In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 2387-2391). IEEE

[6] S Phani Praveen, RajeswariNakka, AnuradhaChokka, VenkataNagarajuThatha, SaiSrinivasVellela and UddagiriSirisha, "A Novel Classification Approach for Grape Leaf Disease Detection Based on Different Attention Deep Learning Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 14(6), 2023. http://dx.doi.org/10.14569/IJACSA.2023.01406128

[7] Vellela, S. S., & Balamanigandan, R. (2022, December). Design of Hybrid Authentication Protocol for High Secure Applications in Cloud Environments. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 408-414). IEEE.

[8] Vullam, N., Vellela, S. S., Reddy, V., Rao, M. V., SK, K. B., & Roja, D. (2023, May). Multi-Agent Personalized Recommendation System in E-Commerce based on User. In 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 1194-1199). IEEE.

[9] Vellela, S. S., Balamanigandan, R., & Praveen, S. P. (2022). Strategic Survey on Security and Privacy Methods of Cloud Computing Environment. Journal of Next Generation Technology (ISSN: 2583-021X), 2(1).

[10] Vellela, S. S., & Krishna, A. M. (2020). On Board Artificial Intelligence With Service Aggregation for Edge Computing in Industrial Applications. Journal of Critical Reviews, 7(07), 2020.

[11] Madhuri, A., Jyothi, V. E., Praveen, S. P., Sindhura, S., Srinivas, V. S., & Kumar, D. L. S. (2022). A New Multi-Level Semi-Supervised Learning Approach for Network Intrusion Detection System Based on the 'GOA'. Journal of Interconnection Networks, 2143047.

[12] Madhuri, A., Praveen, S. P., Kumar, D. L. S., Sindhura, S., &Vellela, S. S. (2021). Challenges and issues of data analytics in emerging scenarios for big data, cloud and image mining. Annals of the Romanian Society for Cell Biology, 412-423.

[13] Praveen, S. P., Sarala, P., Kumar, T. K. M., Manuri, S. G., Srinivas, V. S., &Swapna, D. (2022, November). An Adaptive Load Balancing Technique for Multi SDN Controllers.In 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 1403-1409).IEEE.

[14] Vellela, S. S., Basha Sk, K., & Yakubreddy, K. (2023). Cloud-hosted concept-hierarchy flex-based infringement checking system. International Advanced Research Journal in Science, Engineering and Technology, 10(3).

[15] Sk, K. B., Vellela, S. S., Yakubreddy, K., & Rao, M. V. (2023). Novel and Secure Protocol for Trusted Wireless Ad-hoc Network Creation. Khader Basha Sk, Venkateswara Reddy B, Sai Srinivas Vellela, Kancharakunt Yakub Reddy, M Venkateswara Rao, Novel and Secure Protocol for Trusted Wireless Ad-hoc Network Creation, 10(3).

[16] Rao, M. V., Vellela, S. S., Sk, K. B., Venkateswara, R. B., & Roja, D. (2023). SYSTEMATIC REVIEW ON SOFTWARE APPLICATION UNDERDISTRIBUTED DENIAL OF SERVICE ATTACKS FOR GROUP WEBSITES. Dogo Rangsang Research Journal UGC Care Group I Journal, 13(3), 2347-7180.

[17] Venkateswara Reddy, B., Vellela, S. S., Sk, K. B., Roja, D., Yakubreddy, K., & Rao, M. V. Conceptual Hierarchies for Efficient Query Results Navigation. International Journal of All Research Education and Scientific Methods (IJARESM), ISSN, 2455-6211.

[18] Sk, K. B., Roja, D., Priya, S. S., Dalavi, L., Vellela, S. S., & Reddy, V. (2023, March). Coronary Heart Disease Prediction and Classification using Hybrid Machine Learning Algorithms. In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) (pp. 1-7). IEEE.

[19] Sk, K. B., & Vellela, S. S. (2019). Diamond Search by Using Block Matching Algorithm. DIAMOND SEARCH BY USING BLOCK MATCHING ALGORITHM", International Journal of Emerging Technologies and Innovative Research (www. jetir. org), ISSN, 2349-5162.

[20] Yakubreddy, K., Vellela, S. S., Sk, K. B., Reddy, V., & Roja, D. (2023). Grape CS-ML Database-Informed Methods for Contemporary Vineyard Management. International Research Journal of Modernization in Engineering Technology and Science, 5(03).

[21] Vellela, Sai Srinivas and Chaganti, Aswini and Gadde, Srimadhuri and Bachina, Padmapriya and Karre, Rohiwalter, A Novel Approach for Detecting Automated Spammers in Twitter (June 24, 2023). Mukt Shabd Journal Volume XI, Issue VI, JUNE/2022 ISSN NO : 2347-3150, pp. 49-53 , Available at SSRN: https://ssrn.com/abstract=4490635

[22] Rao, M. V., Vellela, S. S., Sk, K. B., Venkateswara, R. B., & Roja, D. (2023). Systematic Review On Software Application Underdistributed Denial Of Service Attacks For Group Websites. Dogo Rangsang Research Journal UGC Care Group I Journal, 13(3), 2347-7180.

[23] Vellela, Sai Srinivas and Pushpalatha, D and Sarathkumar, G and Kavitha, C.H. and Harshithkumar, D, ADVANCED INTELLIGENCE HEALTH INSURANCE COST PREDICTION USING RANDOM FOREST (March 1, 2023). ZKG International, Volume VIII Issue I MARCH 2023, Available at SSRN: https://ssrn.com/abstract=4473700

[24] D, Roja and Dalavai, Lavanya and Javvadi, Sravanthi and Sk, Khader Basha and Vellela, Sai Srinivas and B, Venkateswara Reddy and Vullam, Nagagopiraju, Computerised Image Processing and Pattern Recognition by Using Machine Algorithms (April 10, 2023). TIJER International Research Journal, Volume 10 Issue 4, April 2023, Available at SSRN: https://ssrn.com/abstract=4428667

[25] Vellela, Sai Srinivas and Basha Sk, Khader and B, Venkateswara Reddy and D, Roja and Javvadi, Sravanthi, MOBILE RFID APPLICATIONS IN LOCATION BASED SERVICES ZONE (June 14, 2023). International Journal of Emerging Technologies and Innovative Research, Vol.10, Issue 6, page no. ppd851-d859, June-2023, http://www.jetir.org/papers/JETIR2306410.pdf,

[26] Vellela, Sai Srinivas and Sk, Khader Basha and B, Venkateswara Reddy, Cryonics on the Way to Raising the Dead Using Nanotechnology (June 18, 2023). INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS), Vol. 03, Issue 06, June 2023, pp : 253-257,

[27] Vellela, Sai Srinivas and D, Roja and B, Venkateswara Reddy and Sk, Khader Basha and Rao, Dr M Venkateswara, A New Computer-Based Brain Fingerprinting Technology (June 18, 2023). International Journal Of Progressive Research In Engineering Management And Science, Vol. 03, Issue 06, June 2023, pp : 247-252 e-ISSN : 2583-1062.,

[28] Gajjala, Buchibabu and Mutyala, Venubabu and Vellela, Sai Srinivas and Pratap, V. Krishna, Efficient Key Generation for Multicast Groups Based on Secret Sharing (June 22, 2011). International Journal of Engineering Research and Applications,Vol. 1, Issue 4, pp.1702-1707, ISSN: 2248-9622

[29] Vellela, Sai Srinivas and Chaganti, Aswini and Gadde, Srimadhuri and Bachina, Padmapriya and Karre, Rohiwalter, A Novel Approach for Detecting Automated Spammers in Twitter (June 24, 2023). Mukt Shabd Journal Volume XI, Issue VI, JUNE/2022 ISSN NO : 2347-3150, pp. 49-53