



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

REAL AND FAKE JOB PREDICTION USING MACHINE LEARNING TECHNIQUE

¹Mrs M. Vidhya, ²Malarvizhi S, ³Reshma R,

¹Assistant Professor, ²Student, ³Student

¹Department of Computer Science and Engineering

¹Panimalar Institute of technology, Chennai, Tamil Nadu, India

Abstract : Now that everything is online, consumers can save time and effort by reducing their manual labour. All job posts are now made available online, providing the employer with a large pool of candidates to choose from, as well as providing information to job seekers. Most organizations can now post jobs directly on their websites. There are job advertisements that are not true, and there are also fake job postings. As a result, we attempt to qualify the bogus posting as genuine. The goal is to anticipate actual or false job prediction outcomes with the highest accuracy using machine learning-based techniques. The dataset is evaluated using the supervised machine learning approach and capture numerous pieces of information such as variable identification, univariate, bivariate, and multivariate data validation, data cleaning/preparation, data analysis, and data visualization.

Keywords- Machine Learning, highest accuracy, fake job postings, bogus, supervised machine learning approach.

I. INTRODUCTION

The proposed solution uses a machine learning-based system to assist non-technical individuals in analyzing these fake scams and securing their jobs in the modern era. The rise of fake jobs today exposes not only the hazards of the fake job's consequences but also the difficulties in distinguishing false employment from actual jobs. However, technological advancements and the distribution of news via various forms of social media have boosted the spread of bogus jobs today. As a result, the consequences of bogus jobs have grown enormously in recent years, and some precautions must be made to avoid this from happening again. As a result, our goal is to employ machine learning to distinguish between real and bogus jobs at least as well as humans.

II. RELATED WORK

In large-scale data centers, early detection of project failures and appropriate discarding mechanisms could considerably enhance resource use efficiency. Existing machine learning-related prediction algorithms typically use offline work patterns that are inconsistent with online forecasting in real-world tasks where data arrives in proper sequence. To solve this issue, this research uses the Digital Sequential Extreme Learning Machine to develop a novel technique for detecting online job termination. This technique, which collects real-time data based on the order in which jobs occur, can be used to forecast the job status, and the following the proposed can be changed as a result of the data. The method employs an online adaptive control mechanism that enables rapid learning and generalization. According to a comparison analysis using Google trace data, the suggested strategy has an accuracy rate of 93 percent of the total and can update the model in 0.1s. The method presented in this study has a number of advantages over state-of-the-art methodologies including less time spent attempting to establish and customize the model, higher predictive accuracies, and better falsified efficiency. The research proposed an automatic solution based on machine learning-based classification approaches to prevent bogus job postings on the web. For checking fraudulent posts on the web, many classifiers are utilized, and the outcomes of those classifiers are contrasted in order to determine the optimum employment scam detection model. It aids in the detection of bogus job postings among a large number of candidates. For the identification of fake job postings, two basic types of classifiers are used: single classifiers and ensemble classifiers. However, experimental data show that ensemble classifiers are superior to single classifiers in detecting frauds.

In this study, they looked at the machine learning classifiers in a text categorization problem. They used both actual and fraudulent job postings in their research. They scrubbed and pre-processed their sample before extracting features with TF-IDF. After the classifiers were implemented, they were trained and evaluated. Accuracy, remembrance, f-measure, and accuracy are the evaluation criteria. The findings of each classifier were compiled and evaluated to those of other classifiers. Public-key cryptography is vital for data communication systems in the realm of cryptography. The public key is made up of two separate keys that are linked together in such a way that the public key can safeguard data while the secret key can reverse it. RSA is one of the most good public cryptography schemes available. The success of the RSA Algorithm is largely determined by the distribution of public key components such as modulus n and public key exponent e . If these components are penetrated using a

mathematical attack, the intruder will have a greater chance of getting a private key. The results of their technique outperform the Integer factorization attack's limitations.

A standardized standard measurements analysis of the planned work is also highlighted in the study. The crucial recruitment process was only recently migrated to the cloud. The automated systems in charge of fill the entire recruitment of new employees are designed to make the recruiting process as efficient, precise, and cost-effective as possible. Furthermore, the internet exposed the weakness of such a standard business approach, resulting in applicants' privacy being violated and firms' reputations being harmed. So far, the most common cause of Online Admission fraud has been employment fraud. Unlike other forms of cybercrime, ORF has received relatively little attention and has been completely unnoticed until lately. The current project explains and defines the characteristics of this critical and important advanced cyber research topic in order to satisfy this demand. Additionally, it adds and evaluates the first publicly available dataset of 17,880 tagged job adverts acquired from a real-world system, to our knowledge.

III. MATHEMATICAL BACKGROUND

Two types of algorithms are compared to know the best accuracy for predicting the Real and Fake job

True Positive Rate(TPR) = $TP / (TP + FN)$; False Positive rate(FPR) = $FP / (FP + TN)$

Accuracy = $(TP + TN) / (TP + TN + FP + FN)$

Precision = $TP / (TP + FP)$

Recall = $TP / (TP + FN)$

Measure = $2TP / (2TP + FP + FN)$

F1 Score = $2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$

IV. ARCHITECTURAL DIAGRAM

The data set is initially gathered, and then the parameter is used to determine if the job descriptions are real or fake. The dataset contains the projected values for each job description. Data mining is used to pre-process and clean the values. The job description from the dataset is used as input for the Input Information. Supervised Machine Learning Technique is used to pre-process the data. The values of the parameters are acquired and saved so that the ML algorithms may compare them. The comparison is carried out in order to acquire the highest level of accuracy. Finally, the outcome is received once the model has been classified.

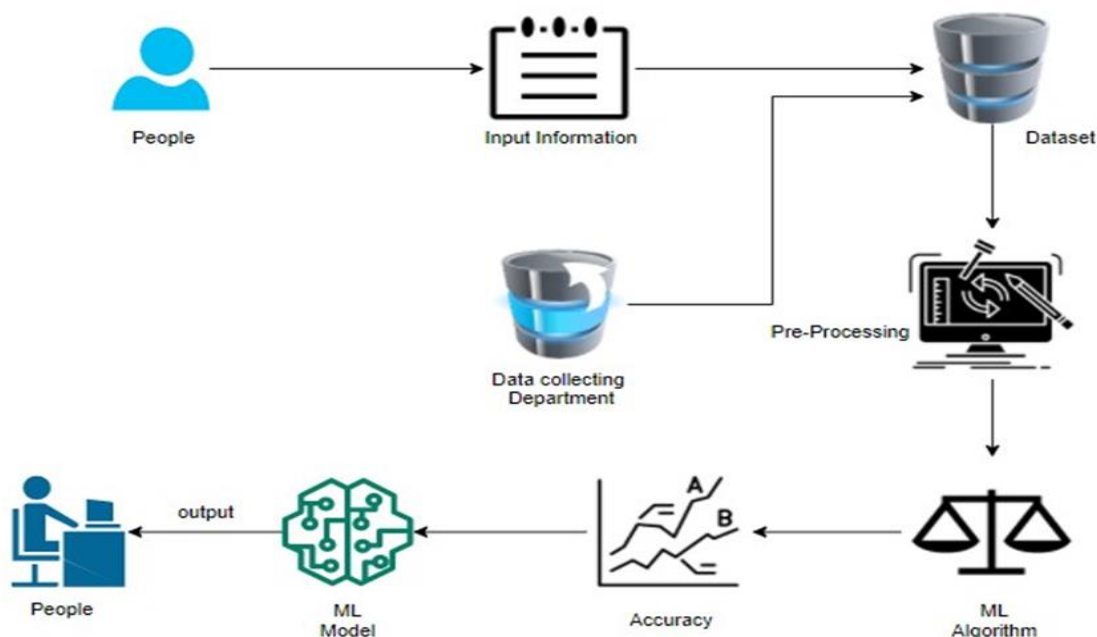


Fig. 1 System flow data

V. MODULES

Data Pre-processing Technique

Machine learning verification methodologies are used to determine the Machine Learning (ML) model's failure rate, which is as near to the dataset's true mistake rate as possible. If the data volume is huge to be an example of the overall public, validation methods may not be necessary. However, dealing with datasets that are not always representative of the public of a dataset is necessary in real-world situations. Identify the missing value by duplicating the value and the data type description, whether it's a float or an integer variable. While tuning model hyper parameters, a subset of the data is employed to give an honest perspective of a baseline model on the training dataset.

Data Analysis of Virtualization.

Visualization of data is one of the most significant skills in machine learning and statistics, as it focuses on analytical forms and estimations of data. Data visualization is a set of tools that can help you gain qualitative insight into your data. While investigating and getting to know a dataset, this could be valuable for finding patterns, erroneous data, anomalies.

Comparing Algorithm for high accuracy

It's vital to use scikit-learn to analyse the performance of several machine learning techniques in Python, as well as to build a check environment to compare different algorithms. This test harness could be used as a framework for your machine learning challenges, permitting you to experiment with more and alternative methods. Each model's performance characteristics will be unique. You might acquire an approximation of how correct each model is on unseen data using resampling approaches like cross-validation.

Deployment using flask framework

Flask is a Python-based microweb framework. Because it does not need any specialised devices or libraries, Flask is characterised as a micro-framework. It is missing an information base reflection layer, structure authorization, and other components that rely on third-party libraries to complete basic tasks. Plugins, on the other hand, may be used to improve the functionality of an application as if it were built into Flask. Object-relational mappers, form validation, upload handling, numerous open authentication protocols, and other framework-related capabilities all have extensions available. From a worldwide organisation of Python fetishists, Armin Ronacher of Pycoco built the flask framework in 2004. The Pycoco projects Werkzeug and Jinja were formed when Ronacher and Georg Brand constructed a bulletin board system in Python.

VI. RESULTS AND DISCUSSION

We developed a machine learning model for predicting whether a job is real or not, with the target of substituting updatable supervised machine learning classification models with forecasting outcomes in the form of maximum accuracy by contrasting supervised algorithms. The analysis technique included data processing and cleaning, missing value analysis, exploratory analysis, and construction and validation of model. A highly accurate score on a public test set represents the highest accuracy. This application can aid in determining the difference between real and phoney jobs.

REFERENCES

- [1] Rathi, Megha, and Vikas Pareek. "Spam mail detection through data mining-A comparative performance analysis." *International Journal of Modern Education and Computer Science* 5.12 (2013): 31.
- [2] Zhou, Kaimin, et al. "Early rumour detection." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 2019.
- [3] Shu, Kai, et al. "The role of user profiles for fake news detection." *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*. 2019.
- [4] Videos, Sokratis, et al. "Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset." *Future Internet* 9.1 (2017): 6.
- [5] Dutta, Shawni, and Samir Kumar Bandyopadhyay. "Fake Job Recruitment Detection Using Machine Learning Approach." *International Journal of Engineering Trends and Technology* 68.4 (2020): 48-53.
- [6] Lal, Sangeeta, et al. "ORFDetector: ensemble learning-based online recruitment fraud detection." *2019 Twelfth International Conference on Contemporary Computing (IC3)*. IEEE, 2019.
- [7] Heydari, Atefeh, et al. "Detection of review spam: A survey." *Expert Systems with Applications* 42.7 (2015): 3634-3642.
- [8] Androutsopoulos, Ion, et al. "An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal email messages." *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. 2000.
- [9] Li, Huayi, et al. "Spotting fake reviews via collective positive-unlabeled learning." *2014 IEEE international conference on data mining*. IEEE, 2014.
- [10] Akoglu, Leman, Rishi Chandy, and Christos Faloutsos. "Opinion fraud detection in online reviews by network effects." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 7. No. 1. 2013.
- [11] Raghunandhan, K. R., et al. "Enhanced RSA algorithm using fake modulus and fake public key exponent." *2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*. IEEE, 2018.
- [12] Nasser, Ibrahim, and Amjad H. Alzaanin. "Machine Learning and Job Posting Classification: A Comparative Study." *International Journal of Engineering and Information Systems (IJEAIS) ISSN* (2020): 6-14.
- [13] Liu, Chunhong, et al. "Predicting job failure in computing cloud based on online extreme learning machine: a comparative study." *IEEE Access* 5 (2017): 9359-9368.
- [14] Khattar, Dhruv, et al. "Mvae: Multimodal variational autoencoder for fake news detection." *The worldwide web conference*. 2019.
- [15] Potthast, Martin, et al. "A stylometric inquiry into hyperpartisan and fake news." *arXiv preprint arXiv:1702.05638* (2017).
- [16] Shu, Kai, et al. "The role of user profiles for fake news detection." *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*. 2019.

- [17] Li, Quanzhi, Qiong Zhang, and Luo Si. "Rumor detection by exploiting user credibility information, attention and multi-task learning." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.
- [18] Rathi, Megha, and Vikas Pareek. "Spam mail detection through data mining-A comparative performance analysis." International Journal of Modern Education and Computer Science 5.12 (2013): 31.
- [19] Lester, Candice, Jeromy Anglim, and Christie Fullarton. "Individual differences in intention to fake job interviews: Personality, self-monitoring, and the theory of planned behaviour." The Australasian Journal of Organizational Psychology 8 (2015).
- [20] Peterson, Mitchell H., et al. "Examining faking in real job applicants: A within-subjects investigation of score changes across applicant and research settings." RL Griffith & MH Peterson (Chairs), Examining faking using within-subjects designs and applicant data. Symposium conducted at the 23 rd Annual Conference for the Society for Industrial and Organizational Psychology: San Francisco, CA. 2008.
- [21] Rubin, Victoria L., et al. "Fake news or truth? using satirical cues to detect potentially misleading news." Proceedings of the second workshop on computational approaches to deception detection. 2016.
- [22] Chen, Yimin, Niall J. Conroy, and Victoria L. Rubin. "Misleading online content: recognizing clickbait as" false news"." Proceedings of the 2015 ACM on workshop on multimodal deception detection. 2015.
- [23] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19.1 (2017): 22-36.
- [24] Papat, Kashyap, et al. "Declare: Debunking fake news and false claims using evidence-aware deep learning." arXiv preprint arXiv:1809.06416 (2018).
- [25] De Sarkar, Sohan, Fan Yang, and Arjun Mukherjee. "Attending sentences to detect satirical fake news." Proceedings of the 27th international conference on computational linguistics. 2018.
- [26] Rubin, Victoria L., et al. "Fake news or truth? using satirical cues to detect potentially misleading news." Proceedings of the second workshop on computational approaches to deception detection. 2016.
- [27] Ma, Jing, Wei Gao, and Kam-Fai Wong. "Detect rumor and stance jointly by neural multi-task learning." Companion Proceedings of the web conference 2018. 2018.
- [28] Wu, Lianwei, and Yuan Rao. "Adaptive interaction fusion networks for fake news detection." arXiv preprint arXiv:2004.10009 (2020).
- [29] Ma, Jing, Wei Gao, and Kam-Fai Wong. "Detect rumor and stance jointly by neural multi-task learning." Companion Proceedings of the web conference 2018. 2018.
- [30] Ruchansky, Natali, SungyongSeo, and Yan Liu. "Csi: A hybrid deep model for fake news detection." Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017.

