International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 4th Sept 2022. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 09

DOI: 10.48047/IJIEMR/V11/ISSUE 09/05

Title **Fake Job AD Prediction Using Machine Learning**

Volume 11, ISSUE 09, Pages: 37-45

Paper Authors
Rekapalli Harshavardan, K.Rajesh Kumar, Dr.P.Srinivasulu, TulasiRaju Nethala

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Fake Job AD Prediction Using Machine Learning

[1]Rekapalli Harshavardan, [2]K.Rajesh Kumar, [3]Dr.P.Srinivasulu, [4]TulasiRaju Nethala

[1]MTech, Department of Computer science and Engineering,
Swarnandhra College of Engineering & Technology, Affiliated to JNTUK
[2.4] Assoc.Professor, Department of Computer science and Engineering,
Swarnandhra College of Engineering & Technology, Affiliated to JNTUK
[3]HOD & Professor, Department of Computer science and Engineering,
Swarnandhra College of Engineering & Technology, Affiliated to JNTUK

harsharekapalli@gmail.com , kumarkaligithi@gmail.com, csehod@swarnandhra.ac.in,
raju.tulasi81@gmail.com

ABSTRACT

Because of advances in modern technology and social communication in recent years, advertising new job ads has become a very common issue in today's world. As a result, the task of predicting fake job postings will be of great concern to all. Fake job posing prediction, like many other classification tasks, presents a number of challenges. As a result of advances in present day innovation and social correspondence lately, promoting new position promotions has turned into an extremely normal issue in this day and age. Subsequently, the errand of anticipating counterfeit work postings will be of extraordinary worry to all. Counterfeit work presenting expectation, in the same way as other order errands, presents various difficulties. This paper proposed utilizing different information mining strategies and characterization calculations, for example, KNN, choice tree, support vector machine, guileless bayes classifier, irregular backwoods classifier, multi-facet perceptron, and profound brain organization to foresee whether a task ad is certifiable or fake. We explored different avenues regarding the Business Trick Aegean Dataset (EMSCAD), which contains 18000 examples. Profound brain networks perform honorably as classifiers in this grouping task. This profound brain network classifier has three thick layers. The prepared classifier predicts a false work promotion with 98% grouping precision (DNN).

Introduction

In this day and age, headways in industry and innovation have made a plenty of new and various open positions for work searchers. Work searchers can limit their choices in view of their time, capability, experience, appropriateness, and different elements by involving the notices for these employment opportunities. The force of the web and online entertainment has now impacted the enlistment cycle. Since the outcome of an enlistment interaction is reliant upon its ad, the effect of virtual entertainment on this is huge. Virtual entertainment and electronic media promotions have extended the quantity of ways of sharing position data.

All things considered, the fast development of chances to share work postings has expanded the level of false work postings, making provocation work searchers. Thus, individuals are more averse to communicate interest in new position postings to keep up with the security and consistency of their own, scholarly, and proficient data. Hence, the genuine objective of substantial work postings by means of social and electronic media faces a very troublesome test in acquiring individuals' trust and dependability.

Innovations are surrounding us to make our lives simpler and more grew, yet not to establish a risky climate for our expert lives. On the off chance that work postings can be appropriately sifted to foresee bogus work postings, this will be a tremendous step in the right direction in enlisting new representatives. Counterfeit work postings make it hard for work searchers to track down their optimal positions, bringing about a critical misuse of their time. A mechanized framework that predicts bogus work postings makes the way for new difficulties in the field of Human Asset The board.

Fake Job Posting: Job Scam

Work tricks are online work promotions that are bogus and principally try to take individual and expert data from work searchers instead of furnish them with reasonable business valuable open doors. Once in a while scalawags endeavor to gather cash from work searchers illicitly. As per a new overview directed by ActionFraud in the Unified Realm, over 67% of individuals who search for occupations on the web however know nothing about counterfeit work postings or occupation tricks are at high gamble [2]. In the Unified Realm, almost 700,000 work searchers revealed losing more than $500000 because of a task trick. As per the report, the UK has seen an almost 300% expansion over the most recent two years [2].

Understudies and late alumni are much of the time focused on by scalawags since they are searching for a steady occupation for which they will pay a premium. Since fraudsters change their strategies for work misleading so much of the time, cybercrime aversion or insurance procedures neglect to diminish this offense.

Common types of Job Scam

Fraudsters make counterfeit work ads to acquire others' very own data, for example, protection subtleties, bank subtleties, annual duty subtleties, date of birth, and public id. Advance charge tricks happen when fraudsters demand cash because of reasons, for example, administrator expenses, data security checking costs, the board costs, etc. At times fraudsters act like businesses and request data, for example, visa numbers, bank proclamations, driving permit numbers, etc as a pre-work check. Innovations are surrounding us to make our lives simpler and more grew, yet not to establish a risky climate for our expert lives. On the off chance that work postings can be appropriately sifted to foresee bogus work postings, this will be a tremendous step in the right direction in enlisting new representatives. Counterfeit work postings make it hard for work searchers to track down their optimal positions, bringing about a critical misuse of their time. A mechanized framework that predicts bogus work postings makes the way for new difficulties in the field of Human Asset The board.

Unlawful cash thinking about tricks happen when they convince understudies to store cash into their records and afterward pull out it [2]. This 'cash close by' procedure permits you to work cash close by and try not to settle charges. To get work searchers, tricksters for the most part make counterfeit organization sites, clone bank sites, clone official-looking reports, etc. Most work tricksters attempt to bait individuals in by means of email as opposed to eye to eye correspondence. They normally utilize online entertainment stages, for example, LinkedIn to lay down a good foundation for themselves as enrollment offices or talent scouts. They normally attempt to give the work searcher as precise a portrayal of their organization profile or site as could be expected. Anything sort of occupation trick they use, they generally target work searchers and gather data from them.

Related Works

Many examinations have been led to decide if a task posting is real or fake. A significant measure of exploration is being directed to research online work extortion. Work con artists were distinguished as phony internet based work promoter by Vidros et al. They found measurements about numerous genuine and notable organizations and endeavors that made fake work promotions or opportunity posts with ulterior thought processes. They investigated the EMSCAD dataset with different order calculations like gullible bayes classifier, irregular woodland classifier, Zero R, One R, etc. The Arbitrary Woods Classifier played out the best on the dataset, with 89.5% arrangement precision. They found that calculated relapse performed very inadequately on the dataset. At the point when they adjusted the dataset and investigated it, one R classifier performed well. In their work, they endeavored to distinguish issues in the ORF model (Online Enrollment Extortion) and to take care of those issues utilizing different prevailing classifiers. Through convolution and pooling layers, this model prepared information. The prepared loads were then leveled prior to being passed to the completely associated layer. For arrangement, this model utilized the softmax capability.

Alghamdi et al. proposed a model to distinguish possible misrepresentation in a web-based enrollment framework. They utilized an AI calculation to probe the EMSCAD dataset. This dataset was dealt with in three phases: information pre-handling, highlight choice, and misrepresentation identification utilizing a classifier. They eliminated commotion and html labels from the information during the pre-handling move toward protect the overall text design.

They utilized the element determination method to actually and effectively diminish the quantity of qualities. Support Vector Machine was utilized to choose highlights, and a gathering classifier in light of arbitrary backwoods was utilized to identify counterfeit work postings in the test information. The irregular backwoods classifier gave off an impression of being a tree-organized classifier that worked as a troupe classifier utilizing the larger part casting a ballot procedure. This classifier distinguished counterfeit work postings with 97.4% exactness.

Huynh et al. proposed utilizing profound brain network models pre-prepared with text datasets like Text CNN, Bi-GRU-LSTM CNN, and Bi-GRU CNN. They dealt with ordering an IT work dataset.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

They prepared the TextCNN model on an IT work dataset, which incorporated a convolution layer, a pooling layer, and a completely associated layer.

Through convolution and pooling layers, this model prepared information. The prepared loads were then leveled prior to being passed to the completely associated layer. For arrangement, this model utilized the softmax capability.

To further develop characterization exactness, they likewise utilized a troupe classifier (Bi-GRU CNN, Bi-GRU-LSTM CNN) with a greater part casting a ballot method. They found that TextCNN had a grouping precision of 66% and Bi-GRU-LSTM CNN had a characterization exactness of 70%. The gathering classifier performed best in this arrangement task, with an exactness of 72.4%. Zhang and partners [4] proposed a programmed counterfeit indicator model that utilizes text handling to recognize valid and counterfeit news (counting articles, makers, and subjects). They had utilized a custom dataset of information or articles posted on Twitter by the PolitiFact site's twitter account. The proposed GDU diffusive unit model was prepared utilizing this dataset. This prepared model performed well as a programmed counterfeit locator model when contribution from numerous sources was gotten simultaneously.

To accomplish great execution in the field of phony work post grouping, specialists tried countless classifiers and component choice methods. Text handling with a profound learning model, highlight choice with a help vector machine, information pre-handling, and different methodologies were mentioned[8], [9], [10], [11], and [12]. To foresee work tricks, we proposed utilizing a profound brain organization. Rather than utilizing text information, we utilized the preparation strategy on the EMSCAD dataset's all out credits as it were. This technique actually diminishes the quantity of teachable traits while requiring less handling time. We led an examination concentrate on utilizing K Closest Neighbor, Guileless Bayes classifier, fluffy KNN, choice tree, support vector machine, irregular timberland classifier, and brain network on similar elements of the EMSCAD dataset. Through convolution and pooling layers, this model prepared information. The prepared loads were then leveled prior to being passed to the completely associated layer. For arrangement, this model utilized the softmax capability.

Methodology

We used various data mining techniques to predict whether or not a job posting is genuine.

Following a pre-processing step, we trained EMSCAD data in the classifiers. The trained classifier can detect fake job postings online.

Neural Network

The neural network operates on the fundamental principle of human brain function. It enables a computer to compare one pattern to another and determine how similar or dissimilar the two are. A neuron is a mathematical function that extracts features and classifies specific patterns. A neural network is made up of many layers of interconnected nodes. Each perceptron node functions as a multiple linear regression. The output of multiple linear regression is passed through this perceptron to produce a non-linear activation function. Perceptrons are arranged in layers that are interconnected. The hidden layers optimise the error rate by adjusting the weights of the input layers. A neural network functions as a classifier for supervised learning.

artificial neuron has a specific threshold and weight. If your input data is above the threshold, the neuron passes the data to the next layer. As you increase the amount of input (training) data, the accuracy of the neural network will increase with each calculation. As you continue to use neural networks for active trading and investing, the AI will continue to optimize its performance and adjust its weighting.

Use of neural networks in investingNeural networks help you develop strategies based on your overall investment strategy: high-risk but growth-focused (short-term trades) or a conservative approach for long-term investment. Let us clarify one thing: neural networks do not make stock forecasts - they help investors evaluate new opportunities. Let's understand this with an example:

Assume you are looking for listed companies that match the high-growth performance of one of the existing companies in your portfolio. In such a situation, the neural network may highlight a handful of other companies with similar fundamentals that will make your work easy.

The above example has a single data pointer as input. You can have N number of data pointers to find new investment opportunities.

How are companies using neural networks?Companies are using a neural network to input their investment and trading ideas that are time-tested. After implementation, you gain data that will help you figure out how effective your idea is likely to be - traditional AI models cannot do it.

A neural network allows you to modify your idea by changing the parameters for the data inputs you want to consider. We have used this feature to modify JARVIS time and again. It helps us improve our platform.

Let's discuss the use of neural networks in a real-world investing problem: Suppose, we have a diversified portfolio and want to manage the portfolio according to a predetermined set of rules, for example, based on risk profile. We can build a neural network to act on investors' portfolios by training it on specific instructions. How can it be implemented? Assume a diversified portfolio where the requirement is to mitigate systematic risk. To accomplish this, we can train a neural network to liquidate a portfolio (or any stock) if it declines by X% on any given day - it acts like a personal circuit breaker.

The best way to implement this strategy is not to follow a discrete set of rules but rather a continuous set of rules that are updated regularly by a combination of reinforcement, unsupervised, and online supervised learning algorithms. To conclude Every investor needs to know about market trends. However, due to the dynamic nature of the market, it is a big challenge to predict the stock price.

FANN

This is an undeniable level open source library written in C. It's restricted to completely associated and inadequately

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

associated brain organizations. Be that as it may, it's been well known throughout the long term, and has even been remembered for Linux appropriations. It's as of late appeared here on Hackaday in a robot that figured out how to walk utilizing support learning, an AI method that frequently utilizes brain organizations.

Deep Neural Network

DeepNeural Network is an Artificial Neural Network (ANN) with multiple layers between the input and output layers. DNN is based on the feed forward algorithm. The data flow is directed from the input layer to the output layer [13]. DNN generates a set of virtual neurons, each with a random numerical value assigned as a connection weight. This weight is multiplied by the input to yield a value between 0 and 1. The weights are adjusted during the training process to efficiently classify the output. Adding layers causes the model to learn rare patterns, which leads to overfitting. Dropout layers reduce the number of trainable parameters in order to generalise the model. For training the data in this paper, we used a sequential model of dense layers with relu as the activation function.

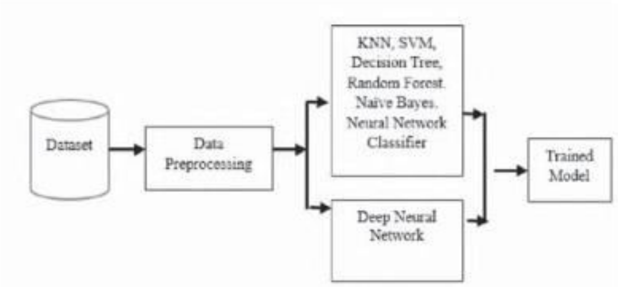
Other classifiers

Our work dataset is trained in the classifiers K Nearest Neighbor, Random Forest Classifier, Decision Tree, Naive Bayes Classifier, Support Vector Machine (RBF kernel), and Multilayer Perceptron (MLP).

Dataset

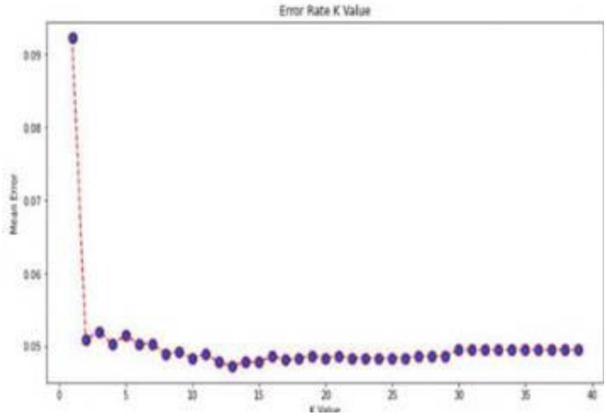
We utilized EMSCAD to recognize counterfeit work postings. This dataset contains 18000 examples, with each column containing 18 credits, including the class name. Work id, title, area, division, compensation range, organization profile, depiction, necessities, benefits, telecom, has organization logo, has questions, business type, required insight, required schooling, industry, capability, and fake are the traits (class name). We utilized just seven of the 18 credits that were changed over into all out ascribes. Working from home, has organization logo, has questions, business type, required insight, required training, and false are changed from text values over completely to downright qualities. For instance, the qualities for "work type" are supplanted as follows: 0 for "none," 1 for "full-time," 2 for "parttime," 3 for "other people," 4 for "agreement," and 5 for "brief." The primary justification behind changing these characteristics over completely to unmitigated structure is to characterize them.

The primary goal of converting these attributes into categorical form is to classify fraudulent job advertisements without using text or natural language processing. We only used categorical attributes in this work.



International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Experimental Result Analysis

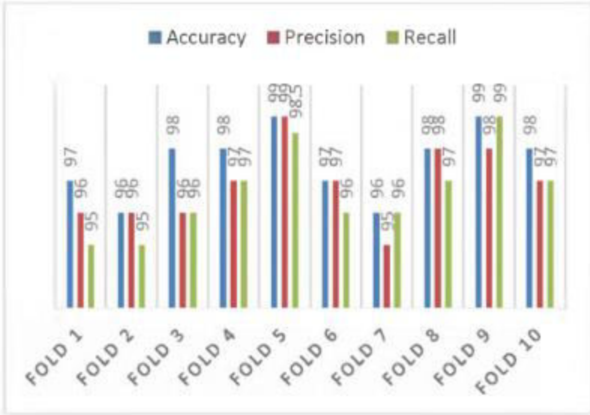


Relation between mean error and K value in KNN

Model	Accuracy	Precision	Recall	F1 Score
K Nearest Neighbor	95.2	93	95	93
Random Forest Classifier	**96.5**	93	95	93
Decision Tree	96.2	93	95	93
Support Vector Machine	95	90	95	92
Naïve Bayes Classifier	91.35	95	96	95
Multilayer perceptron	96	94	95	93

Table comparison among the classifier

Table I displays the classification accuracy, precision, recall, and f1 score of each of these classifiers. We achieved 97% classification accuracy (highest) with the Random Forest classifier. We also looked at the f1 score to see if the model works well with both false positive and false negative samples.



Accuracy Precision and Recall for 10 folds in DNN Model

matrix for DNN Model

CONCLUSION

Identifying position tricks has as of late turned into a central issue from one side of the planet to the other. In this paper, we analyzed the impacts of occupation tricks, which can be an entirely productive area of exploration, making many difficulties to identify fake work promotions. We explored different avenues regarding the EMSCAD dataset, which contains genuine phony work ads. In this paper we have tested both AI calculations (SVM, KNN, Gullible Bayes, Irregular Backwoods and MLP) and profound learning model (Profound Brain Organization) (Profound Brain Organization). This paper presents a near investigation of customary AI and profound learning-based classifiers. We found that the Irregular Woodland Classifier has the most elevated characterization exactness among customary AI calculations, with almost 100% precision for DNN (overlay 9) and 97.7% exactness for DNN (overlap 10).

REFERENCES

[1] S. Vidros, C. Kolias , G. Kambourakis ,and L. Akoglu, “Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset”, Future Internet 2017, 9, 6; doi:10.3390/fi9010006.

[2] B. Alghamdi, F. Alharby, “An Intelligent Model for Online Recruitment Fraud Detection”, Journal of Information Security, 2019, Vol 10, pp. 155 176, https://doi.org/10.4236/iis.2019.103009 .

[3] Tin Van Huynh1, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen1, and Anh Gia-Tuan Nguyen, “Job Prediction: From Deep Neural Network Models to Applications”, RIVF International Conference on Computing and Communication Technologies (RIVF), 2020.

[4] Jiawei Zhang, Bowen Dong, Philip S. Yu, “FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network”, IEEE 36th International Conference on Data Engineering (ICDE), 2020.

[5] Scanlon, J.R. and Gerber, M.S., “Automatic Detection of Cyber Recruitment by Violent Extremists”, Security Informatics, 3, 5, 2014, https://doi.org/10.1186/s13388-014-0005-5

[6] Y. Kim, “Convolutional neural networks for sentence classification,” arXiv
Prepr. arXiv1408.5882, 2014.

[7] T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A.G.- T. Nguyen, “Hate Speech Detection on Vietnamese Social Media Text using the Bi-GRU-LSTM-CNN Model,” arXiv Prepr. arXiv1911.03644, 2019.

[8] P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao, “Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification,” Neurocomputing, vol. 174, pp. 806 814, 2016.

[9] C. Li, G. Zhan, and Z. Li, “News Text Classification Based on Improved BiLSTM-CNN,” in 2018 9th International Conference on Information Technology in Medicine and Education (ITME), 2018, pp. 890-893.

[10] K. R. Remya and J. S. Ramya, “Using weighted majority voting classifier combination for relation classification in biomedical texts,” International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014, pp. 1205-1209.

[11] Yasin, A. and Abuhasan, A. (2016) An Intelligent Classification Model for Phishing Email Detection. International Journal of Network Security& Its Applications, 8, 55-72. https://doi.org/10.5121/imsa.2016.8405

[12] Vong Anh Ho, Duong Huynh-Cong Nguyen, Danh Hoang Nguyen, Linh Thi-Van Pham, Duc-Vu Nguyen, Kiet Van Nguyen, and Ngan Luu- Thuy Nguyen."Emotion Recognition for Vietnamese Social Media Text", arXiv Prepr. arXiv:1911.09339, 2019.

[13] Thin Van Dang, Vu Duc Nguyen, Kiet Van Nguyen and Ngan Luu- Thuy Nguyen, “Deep learning for aspect detection on vietnamese reviews” in In Proceeding of the 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), 2018, pp. 104-109.

[14] Li, H.; Chen, Z.; Liu, B.; Wei, X.; Shao, J. Spotting fake reviews via collective positive-unlabeled learning. In Proceedings of the 2014 IEEE International Conference on Data Mining

(ICDM), Shenzhen, China, 14-17 December 2014; pp. 899-904.
[15] Ott, M.; Cardie, C.; Hancock, J. Estimating the prevalence of deception in online review communities. InProceedings of the 21st international conference on World Wide Web, Lyon, France, 16-20 April 2012; ACM: New York, NY, USA, 2012; pp. 201-210.
[16] Nizamani, S., Memon, N., Glasdam, M. and Nguyen, D.D. (2014) Detection of Fraudulent Emails by Employing Advanced Feature Abundance. Egyptian Informatics Journal, Vol.15, pp.169-174. https://doi.org/10.1016/j.eij.2014.07.002