IJDI-ERET

INTERNATIONAL JOURNAL OF DARSHAN INSTITUTE ON ENGINEERING RESEARCH AND EMERGING TECHNOLOGIES
Vol. 13, No. 1, 2024
www.ijdieret.in

DOI 10.32692/IJDI-ERET/13.1.2024.2403

(Research Article)

Fake Job Prediction Using Machine Learning

V Anbarasu[1*], Dr. S. Selvakani[2], Mrs. K. Vasumathi[3]

[1,2,3]PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, INDIA*

Abstract

In order to mitigate the proliferation of deceptive employment solicitations on the internet, a sophisticated automated tool employing machine learning-based classification methodologies is posited within the confines of this scholarly work. Various classifiers are deployed to scrutinize online postings for fraudulent employment opportunities, and the outcomes of these classifiers are systematically juxtaposed to ascertain the most efficacious model for detecting spurious job listings. This approach facilitates the identification and subsequent elimination of counterfeit job posts from an extensive array of online submissions. The investigation encompasses two principal categories of classifiers: individual classifiers and ensemble classifiers, both instrumental in discerning deceitful job postings. Nonetheless, empirical findings unequivocally affirm that ensemble classifiers exhibit superior efficacy in discerning scams when compared to their singular counterparts. The technological landscape has ascended to a heightened echelon, ushering in a paradigm wherein corporations engage in the recruitment of personnel through the conduit of online methodologies. This not only expedites the acquisition of requisite personnel for businesses but also augurs well in terms of cost-effectiveness. The virtual expanse facilitates individuals in procuring employment commensurate with their qualifications and desired occupational spheres. However, the veracity of these posted job opportunities remains shrouded, posing an inherent challenge for job seekers. In response to this predicament, we proffer a pioneering software meticulously crafted to prognosticate the authenticity of job posts, discerning between genuine and spurious listings. Embarking upon the realm of machine learning, our innovative system, aptly named "Fake Job Post Prediction," leverages the formidable Random Forest classifier. This sophisticated algorithm boasts a commendable efficiency in generating precise outcomes, with a remarkable 98% accuracy vis-à-vis its predecessors. Recognizing the perils faced by students or job seekers navigating the labyrinth of online employment opportunities, our system becomes a beacon of protection against unwittingly submitting personal information to fraudulent job posts. Instances of potential deception, such as solicitation of application fees or promises of employment contingent upon monetary transactions, are thus preemptively averted through the discerning capabilities of our framework, thereby safeguarding users from falling prey to scams.

Keywords: Fake Job Post, Random Forest Classifier, Machine Learning, Legitimate Job, Decision Tree, Online Recruitment, Ensemble Approach.

1. Introduction

In the contemporary milieu, securing gainful employment has become a formidable challenge. Antecedent to partaking in any interview, the prospective candidate must navigate the intricate process of job application and registration. The pivotal initial stride necessitates aligning one's job application with the stipulated requisites of a company, tailored to the aspirant's chosen professional domain. The vast expanse of the internet unfurls myriad job postings, yet discerning between spurious and authentic opportunities proves a daunting task. The labyrinthine nature of this landscape renders it arduous for users to ascertain the veracity of posted positions, prompting the imperative need for a discerning software capable of differentiating between legitimate and counterfeit job listings. Business enterprises, cognizant of the imperative to streamline and expedite the hiring process, disseminate information about job vacancies. Employing diverse data mining techniques, we address the endemic issue of counterfeit job postings. The application of the Random Forest Classifier, a discerning choice, yields superlative results, surpassing the efficacy of previously employed methodologies. This advanced classifier serves as a bulwark against potential financial losses, preemptively identifying deceptive practices such as

[*]Corresponding Author: e-mail: anbarasuvengatesan@gmail.com, Tel-+91-6382453814
ISSN 2320-7590 (Print) 2583-3863 (Online)

solicitation of application fees or other pecuniary demands masquerading as legitimate recruitment processes.

The ubiquity of online recruitment processes, wherein companies articulate job details and applicants vie for positions based on congruence with specified criteria, has become susceptible to exploitation. Job seekers, driven by the imperative need for employment, may inadvertently place unwarranted trust in dubious job postings, risking the compromise of sensitive information such as bank details. It behooves individuals to exercise prudence and circumspection during the application process to preclude falling victim to malevolent entities disseminating deceptive job opportunities. The chosen Random Forest classifier, integral to our discernment software, stands as a paragon of efficacy, outperforming antecedent algorithms in terms of accuracy, efficiency, and temporal expediency. The trajectory of online employment procedures has been marred by the scourge of frauds and scams, jeopardizing personal information and tarnishing the reputation of companies. Our developed project, with its multifaceted enhancements, serves as a resolute corrective measure against such pernicious malfeasance, fostering a more secure and credible employment landscape.

In the wake of economic adversity compounded by the ramifications of the coronavirus, there has been a conspicuous contraction in occupational opportunities, leading to substantial unemployment. Opportunistic malefactors are quick to exploit such circumstances, preying upon the vulnerable populace ensnared in the throes of an unprecedented crisis [16]. A considerable number of individuals find themselves ensnared by these deceptive manipulators, who capitalize on the prevailing desperation induced by extraordinary events. The stratagem employed by the majority of fraudsters revolves around the surreptitious acquisition of personal information from their intended victims. This encompasses details such as addresses, bank account numbers, and social security identifiers. Perfidious scammers lure their targets with ostensibly attractive employment propositions, only to subsequently coerce pecuniary contributions as a prerequisite for securing said opportunities. Alternatively, they may solicit a financial investment from prospective job seekers, ostensibly in exchange for the assurance of gainful employment. The contemporary landscape, dominated by pervasive unemployment, has spawned a proliferation of nefarious schemes, colloquially referred to as job scams [17]. A recruiter possesses the capability to identify a suitably qualified candidate by leveraging diverse online platforms. Regrettably, disingenuous recruiters occasionally exploit job portals, strategically posting positions with the sole aim of pecuniary gain. This predicament afflicts numerous employment boards, thereby impinging upon the experience of those earnestly seeking bona fide opportunities for professional advancement. Individuals, driven by a genuine pursuit of legitimate employment, migrate to newly established job portals, unwittingly carrying with them the risk of encountering spurious recruiters who have similarly infiltrated these platforms. Consequently, it becomes imperative to adeptly discern between authentic and spurious prospects for employment. The realm of Online Recruitment Frauds (ORF) has, in recent years, grappled with the formidable challenge of employment fraud, emerging as a paramount concern within this domain [18]. Modern organizations, cognizant of the digital age, frequently opt to publicize their job vacancies through online channels, streamlining the job-seeking process for aspirants. Nevertheless, this laudable practice may inadvertently serve as a conduit for unscrupulous individuals, posing as legitimate entities, who exploit unsuspecting job seekers by proffering ostensibly enticing employment opportunities in exchange for financial remuneration.

2. Related Work

A compendium of scholarly investigations delves into the nuanced landscape of fraud identification in online processes. Vidros et al. [1] have made a seminal contribution by adeptly employing the Random Forest Classifier methodology to discern and pinpoint fraudulent activities within the realm of online hiring scams. Distinct from this, electronic scams have been meticulously scrutinized, with SVM being harnessed for discerning feature selection, and the Random Forest Classifier deployed for the precise tasks of detection and classification.

Alghamdi and Alharby [2], leveraging the openly accessible EMSCAD dataset, have achieved an impressive culmination with a 97.41% success rate. The focal points of their scrutiny encompass not only corporate logos but also other pivotal attributes.

In a model posited by Tin Van Huynh et al. [3], a judicious consideration of an individual's knowledge and abilities is underscored as imperative for effective employee hiring. The utilization of diverse neural networks, including Text CNN and BI-GRU-LSTM, complemented by pre trained data, yields an efficacious output, manifesting in a 72.71% F1-score. Jiawei Zhang et al. [4] Elucidate the burgeoning growth of online social networking, both politically and economically. Their focus centers on discerning the veracity of news stories through the application of machine learning algorithms, unraveling the origins and subjects of news propagated across online social networks. The overarching aim is to ensure the dissemination of high-quality, authentic news. Thin Van Dang et al. [5] illuminate the intricate workings of Deep Neural Networks (DNN) wherein virtual neurons, initialized with random weights, undergo a process of adjustment during training. The resultant outcomes, confined within the range of 0 to 1, are stratified into diverse groups. Prudent considerations to address over fitting issues involve the strategic deployment of dense layers during data training. The model's architecture is further fine-tuned by judiciously curtailing layers for specific parameters. Activation functions, notably the rectified linear unit (ReLU), and the Adam optimizer are integral components in optimizing the learning rate during the training

regimen. P. Wang et al. [6] posit that the tenets of neural networks mimic the foundational principles governing the human brain's functionality. Neural networks, configured with multiple layers of interconnected nodes, facilitate pattern recognition.

Jihadists [7] expound on the hierarchical arrangement of perceptrons, strategically interconnected in layers to diminish error rates through weight adjustments in input layers. This nuanced orchestration holds the potential to enhance the overall efficacy of neural network models. FHA. Shibly, Uzzal Sharma, and HMM [8] conducted a discerning inquiry into the classification of data, employing the two-class decision boosted tree and two-class decision forest algorithms. The crux of their findings lies in the efficacious performance evaluation of these algorithms, ultimately discerning the superiority of the two-class boosted decision tree over its counterpart, the two-class decision forest algorithm. A salient drawback of their exploration, however, manifests in the protracted training duration and the plethora of hyper parameters, rendering the model susceptible to overfitting. Prof. R. S. Shishupal, Varsha, Supriya Mane, Vinita Singh, Damini Wasekar [9], in their scholarly pursuit, embarked upon the development of an Android-based application for the comparative analysis of diverse classifiers in predicting counterfeit job profiles. Employing Multinomial Naive Bayes, Android, Flask API, Blender, and Natural Language Processing (NLP), their model transforms textual input into speech-driven results. Regrettably, the model's capability is confined solely to textual data, precluding its proficiency in handling numeric inputs.

Shawni Dutta and Prof. Samir Kumar Bandyopadhyay [10] posited a resolute conclusion that ensemble classification outperforms individual classifiers. Employing an array of classifiers such as Naive Bayes, Decision Tree, Multi-Layer Perceptron Classifier, K-nearest Neighbor, AdaBoost, Gradient Boost, and Random Forest Classifiers, their exploration sought to identify the preeminent model among these classifiers. Despite its computational exigency and diminished interpretability, this ensemble approach aspires to elevate classification efficacy. Ibrahim M. Nasser and Amjad H. Alzaanin [11], in their discerning pursuit, navigated the terrain of classification using Multinomial Naive Bayes, Support Vector Machine, Decision Tree, K-Nearest Neighbors, and Random Forest algorithms. Integral to their approach is the utilization of the TF-IDF vectorizer for feature extraction. While demonstrating commendable simplicity and effectiveness, their model grapples with imbalanced data, thereby potentially incurring underfitting and data loss during the classification process. Bandar Alghamdi, Fahad Alharby [12], anchored their exploration in the empirical domain, utilizing Support Vector Machine, Random Forest Classifier, and data mining tools. Employing the Weka tool, their empirical study aspired to augment model efficiency through data mining techniques, albeit falling short in analyzing company profiles, logos, and requisite key attributes. Okti Nindyati, I Gusti Bagus, Baskara Nugraha [13], embarked on an exploration into employment scam detection, introducing the IESD dataset and proposing behavioral context-based features. Leveraging a diverse set of classifiers including Naive Bayes, K-nearest Neighbor, Logistic Regression, Decision Tree, Neural Networks, and Support Vector Machine, their model achieved a noteworthy 90% accuracy, demonstrating efficacy in the identification of fraudulent job vacancies. However, their model's efficacy is constrained by the use of a relatively small dataset encompassing diverse fields.

Sangeeta Lal, Rishabh Jaiswal, Neethu Sardana, Ayushi Verma, Amanpreeth Kaur, Rahul Mourya [14], presented an ensemble-based model, ORF Detector, for Online Fraud Detection (ORF). Utilizing baseline classifiers such as J48, Logistic Regression (LR), and Random Forest (RF), and applying ensemble techniques, their model exhibited a commendable average f1-score and accuracy of 94% and 95.4%, respectively. Nonetheless, the model grapples with interpretability issues and computational expense. Elsevier B V [15 B] delineated the prowess of diverse machine learning techniques in financial fraud detection, encompassing Classification and Regression Tree, Naïve Bayes, and K-Nearest Neighbor methods. The findings underscore the prevalence of hybrid fraud detection techniques, amalgamating the strengths of traditional methods. The model's capacity to handle voluminous transactions at lightning speed with unfaltering accuracy positions it as a robust contender, albeit necessitating substantial investments in data storage and management due to the enormity of the involved datasets.

Sudhakar et al. [19] have advanced an innovative algorithm designed to discriminate between spurious information and authentic news. This investigation delves into logistic regression, Support Vector Machine (SVM), and a pioneering ensemble methodology rooted in machine learning algorithms. The study partitions the data into sample sizes of 620 per group. Employing a dataset comprising 10,000 records featuring binary classifications (fabricated news, genuine news), the outcomes showcase that the proposed avant-garde ensemble technique achieves a superior accuracy rate of 95% and a negligible loss rate of 0.5% in contrast to alternative algorithms. Thus, these findings substantiate that the proposed algorithm epitomizes an ensemble strategy amalgamating decision tree methodologies with AdaBoost, thereby yielding notably enhanced accuracy levels.

Amaar et al. [20] employed six sophisticated machine learning models to scrutinize the authenticity of job advertisements. Subsequently, we juxtaposed all models utilizing both Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) features to assess the classifier's holistic efficacy. A notable challenge encountered in this investigation pertained to the nature of our utilized dataset. The disproportionate ratio between genuine and fraudulent job postings led to the phenomenon of over-fitting on the majority

class data within the model. To circumvent this limitation, we adopted the adaptive synthetic sampling approach (ADASYN), which rectifies the class imbalance predicament by synthetically augmenting the number of samples for the minority class. We conducted two distinct experiments, one employing a balanced dataset and the other utilizing the imbalanced data. Through meticulous experimental analysis, the Extremely Randomized Trees (ETC) model achieved an outstanding 99.9% accuracy rate, leveraging ADASYN as an over-sampling technique alongside TF-IDF as the feature extraction method. Furthermore, this study delves into a comprehensive comparative scrutiny of our proposed methodology against cutting-edge deep learning models and alternative re-sampling techniques.

3. Methodology

We have employed an array of data mining methodologies to discern the veracity of job postings. Following meticulous pre-processing, the EMSCAD dataset was utilized to train our classifiers. The resultant classifier now serves as a discerning mechanism for identifying spurious job advertisements disseminated across the internet.

3.1 Neural network: Harnessing the foundational principles mirroring human cognitive processes, neural networks facilitate the computational assessment of pattern similarity or dissimilarity. Within this paradigm, neurons function as mathematical operators extracting features and categorizing distinct patterns. Comprising multiple layers of interconnected nodes, each perceptron node executes a form of multiple linear regression. The outcome of this regression undergoes transformation via non-linear activation functions within perceptrons. These layers, intertwined, iteratively adjust the weights of input layers to mitigate error rates, thereby functioning as a supervised learning classifier.

3.2 Deep neural network: Deep neural networks (DNNs) represent an extension of Artificial Neural Networks (ANNs), incorporating numerous intermediary layers between input and output layers. Driven by the feedforward algorithm, data traversal commences from the input layer to culminate at the output layer. DNNs instantiate a multitude of virtual neurons initialized with random connection weights, where the input is multiplied by these weights, yielding an output ranging between 0 and 1. Training processes adeptly adjust these weights to optimize categorization efficacy. However, the model may exhibit overfitting, wherein aberrant patterns are assimilated from additional layers. Mitigating this risk, dropout layers facilitate model generalization by curbing the volume of trainable parameters.

3.3 Alternative classifiers: Within our analytical framework, we explore a spectrum of classifiers, including the Decision Tree, Naive Bayes Classifier, K Nearest Neighbor, Random Forest Classifier, and Support Vector Machine. Notably, our research dataset is primed for training across the Multilayer Perceptron (MLP) architecture, affording comprehensive exploration of diverse modeling techniques as shown in Figure 1.

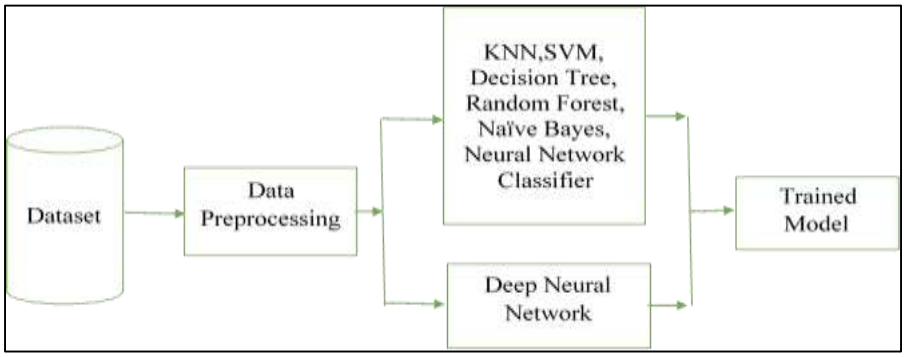

Figure 1. Proposed methodology

3.4 Data compilation: The EMSCAD dataset has been meticulously harnessed for discerning counterfeit job listings. Each entry within this dataset encapsulates 18 distinct attributes, inclusive of the categorical label, culminating in a comprehensive assembly of 18,000 samples. Among these attributes are delineations encompassing employment classification, requisite professional experience, educational prerequisites, sectoral categorizations, functional domains, salary brackets, corporate profiles, job descriptions, stipulated qualifications, employment perks, telecommuting options, company insignia visibility, inquiry provisions, job identifiers, designations, geographic placements, departmental affiliations, and indicators of fraudulent intent. However, for analytical expedience, a refined subset of these attributes has been selectively employed, amounting to a total of seven features. Notably, attributes such as telecommuting feasibility, presence of company insignia, inquiry provisions, employment classification, educational requisites, requisite professional experience, and indicators of fraudulent activity have been converted from textual representations to categorical values. For instance, the values delineating "employment type" have undergone transformation, with "none" corresponding to 0, "full-time" denoted by 1, "part-time" represented by 2, "others" signified as 3, "contract" encapsulated within 4, and "temporary" encapsulated as 5, thereby enhancing computational efficacy and interpretability.

3.5 Machine learning: Machine Learning encompasses a collection of computational algorithms that possess the capacity to autonomously assimilate knowledge from exemplars, progressively refining their performance sans the need for explicit programming by a developer. The formulation of recommendations stands as a ubiquitous challenge within the domain of machine learning. Furthermore, machine learning finds application across a diverse spectrum of tasks.

3.6 Cognitive process in machine learning: The cognitive processes in machine learning are akin to the cerebral activities inherent in human learning. Analogous to the human cognitive mechanism, machine learning involves the acquisition of knowledge primarily through experiential exposure. Human learning is significantly influenced by exposure and experience, a parallel observed in machine learning where training data plays a pivotal role in shaping the system's cognitive capabilities.

3.7 Uncertainty in encountering novel scenarios: The prospects of success are notably diminished when faced with unfamiliar circumstances, mirroring the human predilection for challenges encountered in unknown domains. Similar to human cognition, machines undergo training to enhance their adaptability to novel situations. This training facilitates more precise predictions through pattern recognition, effectively equipping the system to anticipate outcomes based on analogous cases.

3.8 Data-driven discovery: The genesis of machine learning lies in data-driven revelations, where the discerning eye of a data scientist assumes paramount significance. The meticulous curation of input data represents a cardinal skill, allowing the data scientist to selectively furnish the machine with information. The feature vector, a compendium of attributes meticulously chosen to address a problem, serves as the foundational unit of data employed in this cognitive process.

3.9 Model construction through algorithmic sophistication: The machine, through the employment of sophisticated algorithms, distills and encapsulates discovered patterns into a comprehensible model during the learning phase. This model, a refined representation of reality, emerges as a result of the machine's ability to discern underlying structures in the input data.

3.10 Dichotomy of machine learning: Machine learning manifests in two primary forms—Supervised Learning and Unsupervised Learning. In Supervised Learning, the machine is trained with input data, transforming it into a predictive tool through various classifiers and algorithms. Conversely, Unsupervised Learning involves algorithmic exploration of input data without predefined output variables, allowing the machine to discern trends in unclassified data, thereby offering insights into latent patterns.

3.11 Ensemble learning through random forest classifier: The amalgamation of decision tree classifiers collectively referred to as the random forest classifier epitomizes a robust approach to predictive modeling. Predicated upon a voting mechanism, outcomes are determined by the preeminence of predictions within the ensemble. The procedural sequence unfolds as follows:

- Commencing with the provided dataset, a random subsample is systematically selected.
- For each sampled subset, a distinct decision tree is meticulously crafted, furnishing individual predictions for every data point within that subset.
- Aggregating the ensemble's predictive output, each distinct outcome undergoes a democratic voting process.
- The ultimate prediction is adjudicated in favor of the outcome attaining the pinnacle of the voting hierarchy, signifying the most widely endorsed result.

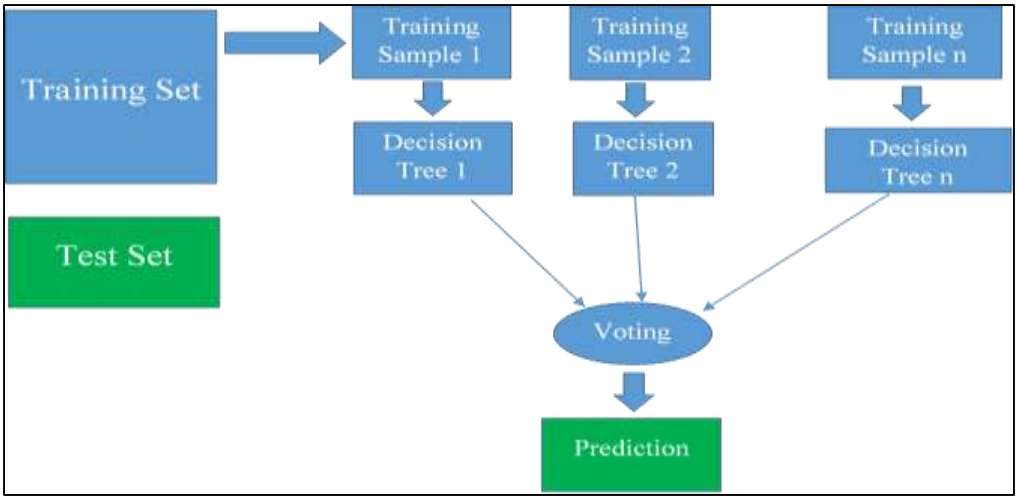

Figure 2. Random forest classifier

4. Experiments and Results

All aforementioned classifiers have undergone training and evaluation processes to discern spurious employment solicitations within a designated dataset encompassing both deceptive and authentic job postings. Table 1 encapsulates a comprehensive examination of these classifiers, elucidating their performance through various evaluative metrics. Meanwhile, Table 2 furnishes outcomes specific to classifiers founded upon ensemble methodologies.

Table 1. Performance comparison chart for single classifier based prediction

Performance measure metric	Naïve bayes classifier	Multi-layer perceptron classifier	K-Nearest neighbor classifier	Decision tree classifier
Accuracy	72.10%	96.12%	95.90%	96.95%
F1-Score	0.75	0.95	0.96	0.98
CohenKappa	0.13	0.2	0.38	0.66
MSE	0.52	0.06	0.03	0.04

Table 2. Performance comparison chart for ensemble classifier based prediction

Performance measure metric	Random forest classifier	AdaBoost classifier	Gradient boosting classifier
Accuracy	98.27%	97.45%	97.65%
F1-Score	0.97	0.98	0.98
Cohen-Kappa score	0.74	0.64	0.64
MSE	0.02	0.03	0.03

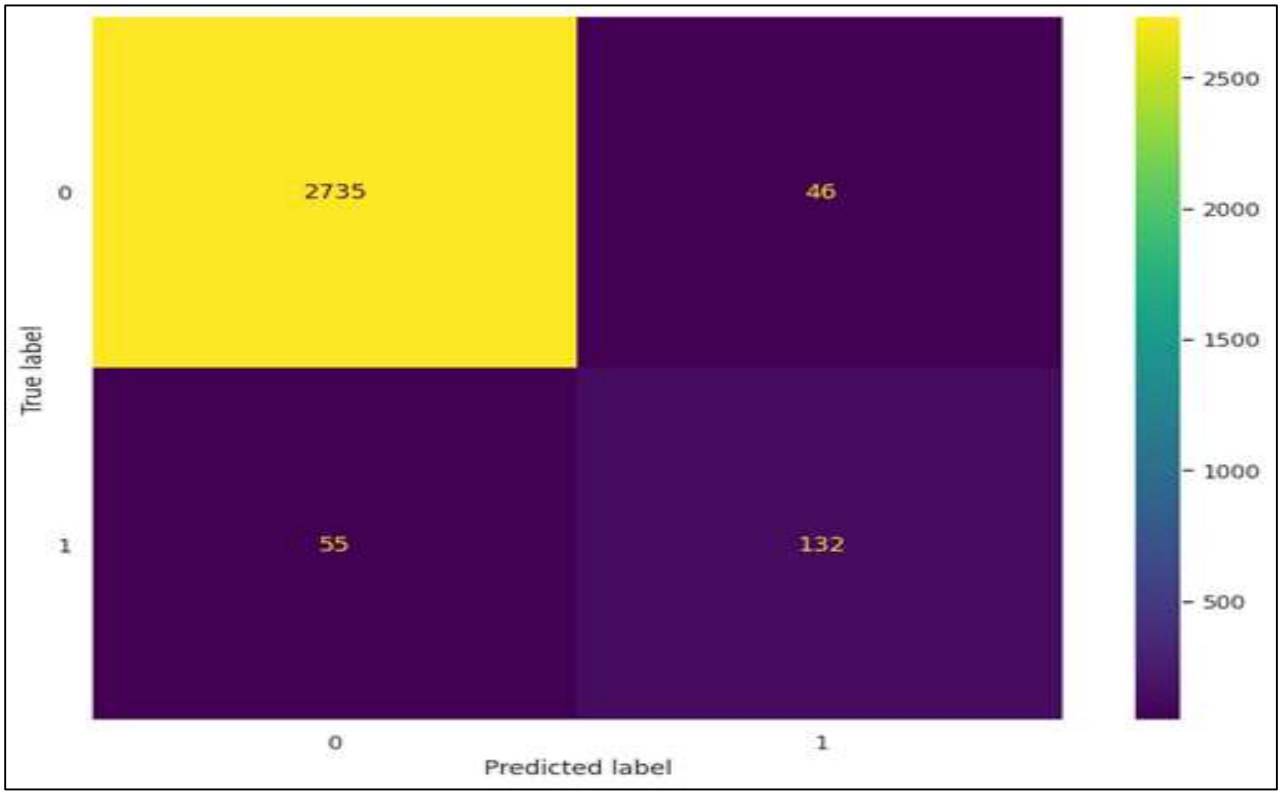

Figure 3. Comparison chart

Upon comparison, it is evident that the SGD performs much better than the random forest classifier on our data. Using the SGD classifier in the pipeline, we are able to predict whether a job is real or fake with adequate precision and recall.

4.1 Final results:

Accuracy = 0.986
F1 Score = 0.89

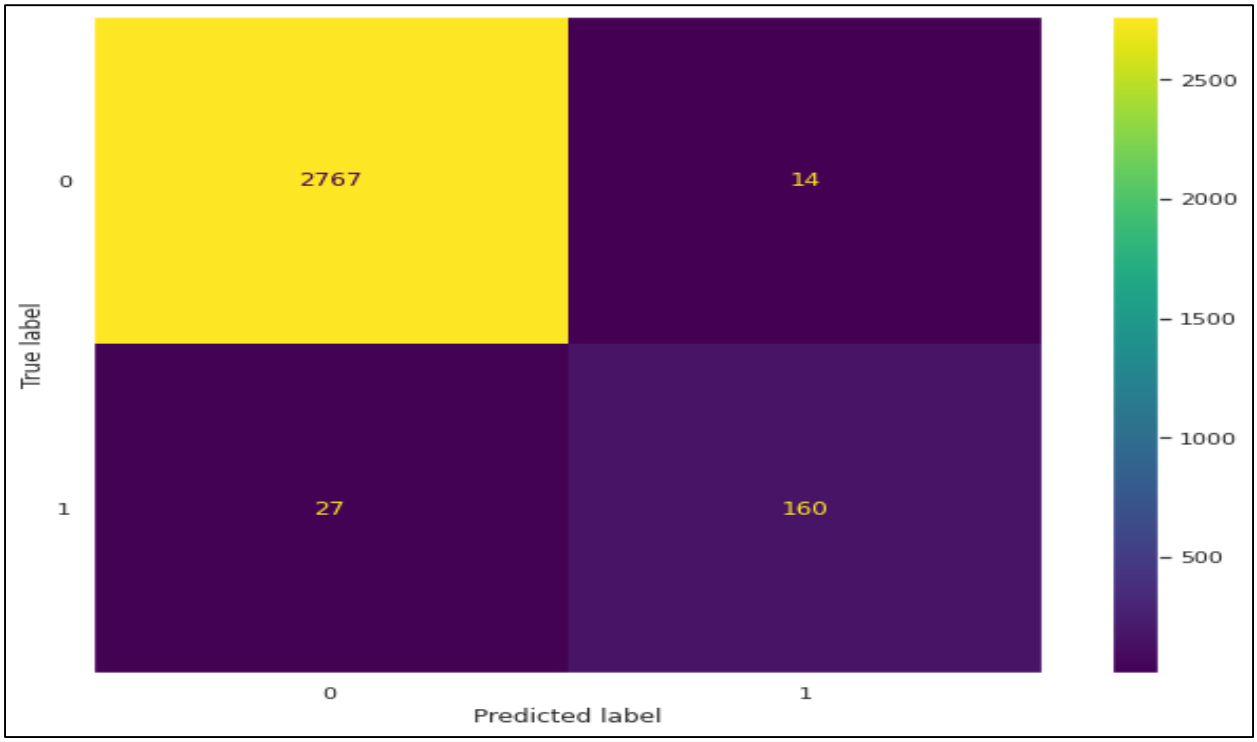

Figure 4. Prediction vs true label

There were a total of 2781 real and 187 fake jobs in the validation set. Our model correctly classified 2767 of 2781 real jobs as real (accuracy = 99.50 %) and 160 of 187 fake jobs as fake (accuracy = 85.56 %)

5. Future Work

Prospective endeavors in this domain may direct their attention towards the investigation of diverse methodologies for assimilating contextual and semantic information from employment solicitations. Furthermore, the integration of additional sophisticated natural language processing techniques, such as Latent Semantic Analysis, the ELECTRA transformer, or GloVe and fast Text for word embeddings, could be considered. Paramount to these considerations, the efficacy of this research and the broader research continuum would be significantly enhanced by the establishment of a publicly accessible database housing contemporaneous job advertisements. Such a repository would afford researchers the opportunity to formulate recommendations that are not only more pertinent but also temporally aligned in the ongoing efforts to mitigate the scourge of online recruitment fraud. Regarding prospective initiatives, we unequivocally commit to the refinement of our technological infrastructure to align with the latest advancements, thereby ensuring satisfaction for our esteemed applicants. Subsequently, our endeavors will encompass the expansion into additional epochs of professional engagement, accompanied by the augmentation of job categories within our application. Such modifications will be meticulously orchestrated in response to forthcoming requisites, exemplified by the prospective inclusion of posts pertaining to academic scholarships, as one illustrative facet of these anticipated enhancements.

6. Conclusions

The contemporary global landscape has witnessed an escalating predicament pertaining to the detection of fraudulent employment opportunities. In the course of our investigation, we have delved into the ramifications of such employment scams, recognizing them as a profoundly lucrative subject for scholarly exploration. The challenge lies in discerning counterfeit job postings, a task rendered formidable by their increasingly sophisticated nature. To address this issue, we conducted a comprehensive analysis employing the EMSCAD dataset, encompassing real-time job advertisements. Our utilization of the random forest classifier yielded a remarkable accuracy rate of 98%, surpassing the efficacy of previously employed algorithms such as SVM and Decision Tree Classifier, which registered accuracies of 90%. This empirical advancement fortifies our commitment to refining the hiring process in the online domain, mitigating the risks associated with fraud and deception within the realm of employment. Consequently, we advocate for the pursuit of online job applications, assuring a safer and more secure hiring procedure. This strategic approach not only safeguards individuals from potential financial losses but also ensures the protection of personal information, thereby fostering a more resilient and trustworthy online employment ecosystem. The identification of employment scams serves as a compass for discerning job-seekers, ensuring that they exclusively encounter bona fide offers from reputable companies. In the

quest to address the challenge of employment scam detection, this scholarly discourse proffers various machine learning algorithms as strategic countermeasures. Employing a supervised mechanism, this study exemplifies the application of multiple classifiers dedicated to the purpose of employment scam detection.

Empirical findings substantiate the superiority of the Random Forest classifier over its counterparts within the classification domain. The proposed methodology attains a commendable accuracy rate of 98.27%, thereby eclipsing the efficacy of prevailing methodologies. This remarkable achievement underscores the innovative strides taken in fortifying the reliability of employment scam detection methodologies.

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*, vol. 9, no. 1, p. 6, Mar. 2017, doi: 10.3390/fi9010006.

[2] B. Alghamdi and F. Alharby, "An Intelligent Model for Online Recruitment Fraud Detection," *Journal of Information Security*, vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.

[3] T. Van Huynh, K. Van Nguyen, N. L.-T. Nguyen, and A. G.-T. Nguyen, "Job Prediction: From Deep Neural Network Models to Applications," 2019, doi: 10.48550/ARXIV.1912.12214.

[4] J. Zhang, B. Dong, and P. S. Yu, "FakeDetector: Effective Fake News Detection with Deep Diffusive Neural Network," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, Dallas, TX, USA: IEEE, Apr. 2020, pp. 1826–1829. doi: 10.1109/ICDE48307.2020.00180.

[5] 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, Dec. 21, 2019. [Online]. Available: http://arxiv.org/abs/1911.03644

[6] D. V. Thin, V. D. Nguye, K. V. Nguyen, and N. L.-T. Nguyen, "Deep Learning for Aspect Detection on Vietnamese Reviews," in *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City: IEEE, Nov. 2018, pp. 104–109. doi: 10.1109/NICS.2018.8606857.

[7] 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, Jan. 2016, doi: 10.1016/j.neucom.2015.09.096.

[8] J. R. Scanlon and M. S. Gerber, "Automatic detection of cyber-recruitment by violent extremists," *Security Informatics*, vol. 3, no. 1, p. 5, Dec. 2014, doi: 10.1186/s13388-014-0005-5.

[9] S. Dutta and S. K. Bandyopadhyay, "Fake Job Recruitment Detection Using Machine Learning Approach," *International Journal of Engineering Trends and Technology*, vol. 68, no. 4, pp. 48–53, Apr. 2020, doi: 10.14445/22315381/IJETT-V68I4P209S.

[10] I. M. Nasser and A. H. Alzaanin, "Machine Learning and Job Posting Classification: A Comparative Study," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 4, no. 9, pp. 06–14, 2020.

[11] F. Shibly, U. Sharma, and H. Naleer, "Performance Comparison of Two Class Boosted Decision Tree snd Two Class Decision Forest Algorithms in Predicting Fake Job Postings," *Annals of the Romanian Society for Cell Biology*, pp. 2462–2472, Apr. 2021.

[12] B. Alghamdi and F. Alharby, "An Intelligent Model for Online Recruitment Fraud Detection," *JIS*, vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.

[13] O. Nindyati and I. G. Bagus Baskara Nugraha, "Detecting Scam in Online Job Vacancy Using Behavioral Features Extraction," in *2019 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia: IEEE, Nov. 2019, pp. 1–4. doi: 10.1109/ICISS48059.2019.8969842.

[14] R. S. Shishupal, Varsha, S. Mane, V. Singh, and D. Wasekar, "Efficient Implementation using Multinomial Naive Bayes for Prediction of Fake Job Profile," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 286–291, May 2021, doi: 10.48175/IJARSCT-1241.

[15] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised Learning for Fake News Detection," *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 76–81, Mar. 2019, doi: 10.1109/MIS.2019.2899143.

[16] A. Das, S. A. Desale, and MCA Department, Mumbai University ASM's Institute of Management and Computer Studies Thane, Maharashtra, India, "Techniques to Analyse, Identify and Verify the Online Job Offers by Fake Companies World Wide," *International Journal of Trend in Scientific Research and Development*, vol. Volume-2, no. Issue-4, pp. 2660–2663, Jun. 2018, doi: 10.31142/ijtsrd15684.

[17] L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, and C. Clavel, "HireNet: A Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 573–581, Jul. 2019, doi: 10.1609/aaai.v33i01.3301573.

[18] S. Lal, R. Jiaswal, N. Sardana, A. Verma, A. Kaur, and R. Mourya, "ORFDetector: Ensemble Learning Based Online Recruitment Fraud Detection," in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, Noida, India: IEEE, Aug. 2019, pp. 1–5. doi: 10.1109/IC3.2019.8844879.

[19] M. Sudhakar and K. P. Kaliyamurthie, "Efficient Prediction of Fake News Using Novel Ensemble Technique Based on Machine Learning Algorithm," in *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*, vol. 401, M. S.

Kaiser, J. Xie, and V. S. Rathore, Eds., Singapore: Springer Nature Singapore, 2023, pp. 1–8. doi: 10.1007/978-981-19-0098-3_1.

[20] A. Amaar, W. Aljedaani, F. Rustam, S. Ullah, V. Rupapara, and S. Ludi, "Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches," *Neural Processing Letters*, vol. 54, no. 3, pp. 2219–2247, Jun. 2022, doi: 10.1007/s11063-021-10727-z.

Biographical notes

V. Anbarasu has received his BCA Degree from Government Arts and Science College, Arakkonam. He is doing his M.Sc. Computer Science at Government Arts and Science College Arakkonam. He has presented more than 2 papers in the International Conference. His research interests are Big Data, Deepa Learning, Machine Learning and Network Security.

Dr.S.Selvakani Working as an Assistant Professor and Head in the Department of Computer Science / Computer Applications at Government Arts and Science College, Arakkonam (Formerly Thiruvalluvar University College of Arts and Science). She Completed Ph.D (Computer Science), M.Tech (Computer Science and Engg), M.Phil (Computer Science), MCA (university III Rank) at Manonmanium Sundaranar University, Tirunelveli, Tamilnadu, India. She is a dynamic hardworking professional with rich experience of 21 years in Teaching, Administration/ Training & Development. She is having hands on experience in general administrative activities, Research contributions and Student Counseling. She has published more than 69 papers in the peer reviewed International Journals and more than 6 papers in National Journals. Besides presented more than 120 papers in the National and International Conferences and guiding several Ph.D. Scholars in Computer Science. She has published 8 books and 6 Patents. Her interest areas are Data Science, Big Data, Machine Learning and Network Security.

Mrs.K Vasumathi Working as an Assistant Professor Department of Computer Science / Computer Applications at Government Arts and Science College, Arakkonam (Formerly Thiruvalluvar University College of Arts and Science). She Completed S.E.T (Computer Science), M.Phil. (Computer Science and Engg), M.Phil (Computer Science), M.Sc. (university VI Rank) at Thiruvalluvar University, Vellore, Tamilnadu, India. She is a dynamic hardworking professional with rich experience of 16 years in Teaching. She is having hands on experience in Research contributions and Student Counseling. She has published more than 20 papers in the peer reviewed International Journals. Besides presented more than 42 papers in the National and International Conferences. She has published 2 books and 2 Patents. Her interest areas are Operating System, Big Data, and Machine Learning.