

MACHINE LEARNING BASED FAKE JOB PREDICTION

Viragoni Venkatesh(20641A6765), Btech Student, CSE (Data Science), Vaagdevi College of Engineering

Mohammed Nawaz(20641A6744), Btech Student, CSE (Data Science), Vaagdevi College of Engineering

Nallabeli Roshan(21645A6713), Btech Student, CSE (Data Science), Vaagdevi College of Engineering

Nerra Kalyan(21645A6715), Btech Student, CSE (Data Science), Vaagdevi College of Engineering

Mrs. Zareena Begum, Assistant Professor, CSD, Vaagdevi College of Engineering

ABSTRACT

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

1. INTRODUCTION

Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF). In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them. Fraudulent job advertisements [1] can be posted against a reputed company for violating their credibility. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs.

For this purpose, machine learning approach is applied which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially. A classifier maps input variable to target classes by considering training data. Classifiers addressed in the paper for identifying fake job posts from the others are described briefly. These classifiers based prediction may be broadly categorized into -Single Classifier based Prediction and Ensemble Classifiers based Prediction.

A. Single Classifier based Prediction

Classifiers are trained for predicting the unknown test cases. The following classifiers are used while detecting fake job posts

Naive Bayes Classifier-

The Naive Bayes classifier is a supervised classification tool that exploits the concept of Bayes Theorem of Conditional Probability. The decision made by this classifier is quite effective in practice even if its probability estimates are inaccurate. This classifier obtains a very promising result in the following scenario- when the features are independent or features are completely functionally dependent. The accuracy of this classifier is not related to feature dependencies rather than it is the amount of information [2] loss of the class due to the independence assumption is needed to predict the accuracy.

Multi-Layer Perceptron Classifier -

Multi-layer perceptron can be used as supervised classification tool by incorporating optimized training parameters. For a given problem, the number of hidden layers in a multilayer perceptron and the number of nodes in each layer can differ. The decision of choosing the parameters depends on the training data and the network architecture.

K-nearest Neighbor Classifier-

K-Nearest Neighbour Classifiers, often known as lazy learners, identifies objects based on closest proximity of training examples in the feature space. The classifier considers k number of objects as the nearest object while determining the class. The main challenge of this classification technique relies on choosing the appropriate value of k.

Decision Tree Classifier-

A Decision Tree (DT) is a classifier that exemplifies the use of tree-like structure. It gains knowledge on classification. Each target class is denoted as a leaf node of DT and non-leaf nodes ofDT are used as a decision node that indicates certain test. The outcomes of those tests are identified by either of the branches of that decision node. Starting from the beginning at the root this tree are going through it until a leaf node is reached. It is the way of obtaining classification result from a decision tree. Decision tree learning is an approach that has been applied to spam filtering. This can be useful for forecasting the goal based on some criterion by implementing and training this model.

2. LITERATURE SURVEY

This study research attempts to prohibit privacy and loss of money for individuals and organization by creating a reliable model which can detect the fraud exposure in the online recruitment environments. This research presents a major contribution represented in a reliable detection model using ensemble approach based on Random forest classifier to detect Online Recruitment Fraud (ORF) [3]. The detection of Online Recruitment Fraud is characterized by other types of electronic fraud detection by its modern and the scarcity of studies on this concept. The researcher proposed the detection model to achieve the objectives of this study. For feature selection, support vector machine method is used and for classification and detection, ensemble classifier using Random Forest is employed. A freely available dataset called Employment Scam Aegean Dataset (EMSCAD) is used to apply the model. Pre-processing step had been applied before the selection and classification adoptions. The results showed an obtained accuracy of 97.41%. Further, the findings presented the main features and important factors in detection purpose include having a company profile feature, having a company logo feature and an industry feature.

The naive Bayes classifier greatly simplify learn-ing by assuming that features are independent given class. Although independence is generally a poor assumption, in practice naive Bayes often competes well with more sophisticated classifiers. Our broad goal is to understand the data character-istics which affect the performance of naive Bayes. Our approach uses Monte Carlo simulations that al-low a systematic study of classification accuracy for several classes of randomly generated prob-lems. We analyze the impact of the distribution entropy on the classification error, showing that low-entropy feature distributions yield good performance of naive Bayes. We also demonstrate that naive Bayes works well for certain nearly-functional feature dependencies, thus reaching its best performance in two opposite cases: completely independent features [4] (as expected) and function-ally dependent features (which is surprising). An-other surprising result is that the accuracy of naive Bayes is not directly correlated with the degree of feature dependencies measured as the class-conditional mutual information between

the fea-tures. Instead, a better predictor of naive Bayes ac-curacy is the amount of information about the class that is lost because of the independence assumption.

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

We review the theory and practice of the multilayer perceptron. We aim at addressing a range of issues which are important from the point of view of applying this approach to practical problems. A number of examples are given, illustrating how the multilayer perceptron compares to alternative, conventional approaches [5]. The application fields of classification and regression are especially considered. Questions of implementation, i.e. of multilayer perceptron architecture, dynamics, and related aspects, are discussed. Recent studies, which are particularly relevant to the areas of discriminant analysis, and function mapping, are cited.

We analyze a Relational Neighbor (RN) classifier, a simple relational predictive model that predicts only based on class labels of related neighbors, using no learning and no inherent attributes. We show that it performs surprisingly well by comparing it to more complex models such as Probabilistic Relational Models and Relational Probability Trees on three data sets from published work.

As the computer technology and computer network technology are developing, the amount of data in information industry is getting higher and higher. It is necessary to analyze this large amount of data and extract useful knowledge from it. Process of extracting the useful knowledge from huge set of incomplete, noisy, fuzzy [6] and random data is called data mining. Decision tree classification technique is one of the most popular data mining techniques. In decision tree divide and conquer technique is used as basic learning strategy. A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. This paper focus on the various algorithms of Decision tree (ID3, C4.5, CART), their characteristic, challenges, advantage and disadvantage.

The upsurge in the volume of unwanted emails called spam has created an intense need for the development of more dependable and robust antispam filters. Machine learning methods of recent are being used to successfully detect and filter spam emails. We present a systematic review of some of the popular machine learning based email spam filtering approaches. Our review covers survey of the important concepts, attempts, efficiency, and the research trend in spam filtering. The preliminary discussion in the study background examines the applications of machine learning techniques to the email spam filtering process of the leading internet service providers (ISPs) like Gmail, Yahoo and Outlook [7] emails spam filters. Discussion on general email spam filtering process, and the various efforts by different researchers in combating spam through the use machine learning techniques was done. Our review compares the strengths and drawbacks of existing machine learning approaches and the open research problems in spam filtering. We recommended deep leaning and deep adversarial learning as the future techniques that can effectively handle the menace of spam emails.

Several machine-learning algorithms have been proposed for remote sensing image classification during the past two decades. Among these machine learning algorithms, Random Forest (RF) and Support Vector Machines (SVM) have drawn attention to image classification in several remote sensing applications. This paper reviews RF and SVM [8] concepts relevant to remote sensing image classification and applies a meta-analysis of 251 peer-reviewed journal papers. A database with more than 40 quantitative and qualitative fields was constructed from these reviewed papers. The meta-analysis mainly focuses on: (1) the analysis regarding the general characteristics of the studies, such as geographical distribution, frequency of the papers considering time, journals, application domains, and remote sensing software packages used in the case studies, and (2) a comparative analysis regarding the performances of RF and SVM classification against various parameters, such as data type, RS applications, spatial resolution, and the number of extracted features in the feature engineering step. The challenges, recommendations, and potential directions for future research are also discussed in detail. Moreover, a summary of the results is provided to aid researchers to customize their efforts in order to achieve the most accurate results based on their thematic applications.

Pattern recognition systems have been widely used in *adversarial classification* tasks like spam filtering and intrusion detection in computer networks. In these applications a malicious

adversary may successfully mislead a classifier by "poisoning" its training data with carefully designed attacks. Bagging is a well-known ensemble construction method, where each classifier in the ensemble is trained on a different bootstrap replicate of the training set. Recent work has shown that bagging can reduce the influence of outliers in training data, especially if the most outlying observations are resampled [9] with a lower probability. In this work we argue that poisoning attacks can be viewed as a particular category of outliers, and, thus, bagging ensembles may be effectively exploited against them. We experimentally assess the effectiveness of bagging on a real, widely used spam filter, and on a web-based intrusion detection system. Our preliminary results suggest that bagging ensembles can be a very promising defence strategy against poisoning attacks, and give us valuable insights for future research work.

Gradient boosting machines are a family of powerful machine-learning techniques that have shown considerable success in a wide range of practical applications. They are highly customizable to the particular needs of the application,[14],[15] like being learned with respect to different loss functions. This article gives a tutorial introduction into the methodology of gradient boosting methods with a strong focus on machine learning aspects of modeling. A theoretical information is complemented with descriptive examples [10] and illustrations which cover all the stages of the gradient boosting model design. Considerations on handling the model complexity are discussed. Three practical examples of gradient boosting applications are presented and comprehensively analyzed.

3. PROBLEM STATEMENT

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

Review Spam Detection

People often post their reviews online forum regarding the products they purchase. It may guide other purchaser while choosing their products. In this context, spammers can manipulate reviews for gaining profit and hence it is required to develop techniques that detects these spam reviews. This can be implemented by extracting features from the reviews by extracting features using Natural Language Processing (NLP) [11]. Next, machine learning techniques are applied on these features. Lexicon based approaches may be one alternative to machine learning techniques that uses dictionary or corpus to eliminate spam reviews.

Email Spam Detection

Unwanted bulk mails, belong to the category of spam emails, often arrive to user mailbox. This may lead to unavoidable storage crisis as well as bandwidth consumption. To eradicate this problem, Gmail, Yahoo mail and Outlook [12] service providers incorporate spam filters using Neural Networks. While addressing the problem of email spam detection, content based filtering, case based filtering, heuristic based filtering, memory or instance based filtering, adaptive spam filtering approaches are taken into consideration.

Fake News Detection

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

4. PROPOSED SYSTEM

The target of this study is to detect whether a job post is fraudulent or not. Identifying and eliminating these fake job advertisements will help the job seekers to concentrate on legitimate job posts only. In this context, a dataset from Kaggleis [13] employed that provides information regarding a job that may or may not be suspicious.

Implementation of Classifiers

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

Performance Evaluation Metrics

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

5. IMPLEMENTATION

There are 2 modules:

- 1. Admin
- 2. User or Candidate

Admin:-

- > Login
- User Management
 - Pending Users
 - All User
- ➤ Fake job
 - Upload Dataset
 - View Dataset
- > Algorithm
 - SVM Algorithm
 - Decision Tree Algorithm
 - Naïve Bayes Algorithm
 - K-NN Bayes Algorithm
 - Random Forest Algorithm
- ➢ Graph Analysis
 - Comparision Graph

User:-

- ➢ Register
- > Login
- > Predict

6. EXPECTED RESULTS







😩 Sign in 🕼 🗖 🗗	out × +			- o ×
← C ① 127.0.0.1:8000				0 4 6 8 📀
MACHINE	E LEARNING BASE	D FAKE JOB PREDIC	LIION Home User Admin	About Contact
				I
	ABOUT US			I
	#1 Digital solution	with 10		
	years of experienc	e		
	Diam dolor diam ipsum et tempor sit. Aliqu diam labore. Clita erat ipsum et lorem et sit, sed stet n	amet diam et eos o labore lorem sit		
	clita duo justo eirmod magna dolore erat amet	9.5%		
		80%		
	Candidata satisfaction	00%		î
				•
Mostly sunny	Q Search		2 🔯 💆 🔷 e	■ ENG 奈 (小) ID 29-04-2024 単
 Sign in) Image: Sign in) Image: Si	xout × +		A* ☆	- ° × □ ¢ @ ⊗ … Ø
MACHINE	E LEARNING BASE	D FAKE JOB PREDI	CTION Home User Admin	About Contact
		——————————————————————————————————————		
		Our Team Members		
	John Doe CEO & Founder	Jessica Brown Web Designer	Tony Johnson SEO Expert	
	f 🎔 🎯 in	f 🎔 💿 in	f 🎔 💿 in	
				\bigcirc
ee 40°C Mostly sunny	Q Search	See □ □ □ □ □ □	2 💆 💆 💆 🗠 🤇	S ENG (⇒ Φ) ■ 16:46 IN (⇒ Φ) ■ 29-04-2024





D	Admin ashboard							(💽 John Doe 🦴
Ø	Dashboard					All Us	ers		
	Managarillagua		ID	Name	DOB	Email	Contact	Occupation	Action
-	Manage Users	•	1	viswa	Feb. 1, 2023	sri961@gmail.com	9949226084	developer	accepted
?	Fake Job	~							
	Algorithms	~							
Lui	Analysis	~							
	,								
			© Co	debook.in,	All Rights Reserved.			Desig	ned By Codebook.

D	Admin ashboard		•		🧕 John Doe 🗸
Ø	Dashboard				
⊒	Manage Users	*		Upload Dataset	
?	Fake Job	•	Upload your da	taset file (only .csv and .xlsx files accepted)	
	Algorithms	•	Choose File	No file chosen	
	Analysis	v		Upload Dataset	
4			© Codebook.in, All Rights Reserved.		Designed By Codebook.in

C	127.0.0.1:8000	0/admin-view		🗸 🗠 🤹 👌 📬 🏽
				John Doe 🗸
Da	Admin ashboard			
			fraud	ulent
		0	0	Marketing Intern US, NY, New York We're Food52, and we've created a groundbreaking and award-winning cooking maddened by typos and broken links, delighted by finding and fixing themCheerful under pressureExcellent comm
	Dashboard	1	0	Customer Service - Cloud Video Production NZ, , Auckland 90 Seconds, the worlds Cloud Video Production Servic entering the next growth stage of our business and growing quickly internationally. Therefore, the position is burs
1	Manage Users	↓ 2	0	show initiative and think on your feet as this is a constantly evolving atmosphereFlexible - fast turnaround work a Commissioning Machinery Assistant (CMA) US, IA, Wever Valor Services provides Workforce Solutions that meet
	Fake Job	~		authorization by site manager to receive or issue a Permit To Work according to project Permit To Work procedure Account Executive - Washington DC US, DC, Washington Our passion for improving quality of life through geograp
),	Algorithms	✓	0	maintain a healthy pipeline of opportunities for business growthDemonstrate a thoughtful understanding of insigi real estate industry fiscal year, budgeting, and procurement cycleHighly motivated team player with a mature, pos
<u>,</u>	Analysis	~ ⁴	0	Bill Review Manager US, FL, Fort Worth SpotSource Solutions LLC is a Global Human Capital Management Consul AND RESPONSIBILITIES:Oversee company's Itemization Review department in its operationsResponsible for enco
		5	0	Accounting Clerk US, MD, Job OverviewApex is an environmental consulting firm that offers stable leadership and experience in a wide variety of industries. Our corporate professional resume includes proven capabilities in the a
		6	0	Head of Content (m/f) DE, BE, Berlin Founded in 2009, the Fonpit AG rose with its international web portal ANDRO booming industryFast decision-making thanks to flat hierarchies and clear structuresFreedom to unfold your own
		7	0	Lead Guest Service Specialist US, CA, San Francisco Airenvy's mission is to provide lucrative yet hassle free full s SolverAirenvy is growing faster than we can handle, which is why we're looking for someone to help us scale! We

Admin							🧕 John Doe 🗸			
D	ashboard									
æ	Dashboard		Decision Tree Algorithm Run algorithm							
⊒	Manage Users	*								
?	Fake Job	~	Algorithm Details							
	Algorithms	*	Algorithm Used	Accuracy	Precision Score	Recall Score	F1 Score			
1.00	Analysis	*	DecisionTreeClassifier	0.9781879194630873	0.8917677658586449	0.8546277665995976	0.8721626			
			4				•			
			© Codebook.in, All Rights	s Reserved.		D	esigned By Codebook.i			

D	Admin						🧿 John Doe 🗸			
	Dashboard		Naive Bayes Algorithm Run algorithm							
□ ?	Manage Users Fake Job	* *								
	Algorithms	• •	Algorithm Deta	ed Accuracy	Precision Score	Recall Score	F1 Score			
-	.,		Naive-Dayes	0.93700003319701	0.0	0.77004041709000000	0.407191000070090			
4			© Codebook.in, All	Rights Reserved.			Designed By Codebook.in			
		(;	≡)				🎯 John Doe 🗸			
D	Admin ashboard									
2	Dashboard			I	ogistic Regro	ession				
₽	Manage Users	•								
?	Fake Job	*	Algorithm Deta	ils						
	Algorithms	*	Algorithm Used	Accuracy	Precision Score	Recall Score	F1 Score			
<u>[.11</u>	Analysis	~	Logistic Regression	0.9750186428038777	0.7132670553700844	0.9678476492908649	0.7892893996247655			
			© Codebook.in, All	Rights Reserved.			Designed By Codebook.in			

	Admin		=			(John Doe 🗸					
D	ashboard											
2	Dashboard		Random Forest Algorithm									
	Manage Users	•										
?	Fake Job	•	Algorithm Details									
	Algorithms	~	Algorithm Used	Accuracy	Precision Score	Recall Score	F1 Score					
1.00	Analysis	~	RandomForestClassifier	0.9837807606263982	0.8167912844449743	0.9785301981429282	0.879798					
			4				,					
	© Codebook.in, All Rights Reserved.											



Date	Used	Accuracy	Precision Score	Recall Score	F1 Score
01 Jan 2023	Logistic Regression	97.50186428038778	71.32670553700844<	96.78476492908649<	78.92893996247655<
01 Jan 2023	Decision Tree	97.81879194630872	89.17677658586449<	85.46277665995976<	87.21626897354007<
01 Jan 2023	Naive Bayes	95.76808351976138	50.0<	47.88404175988069<	48.9191505570898<
01 Jan 2023	Random Forest	98.37807606263982	81.67912844449742<	97.85301981429282<	87.97983193277311<



MACHINE LEARN	ING BASED FAK	E JOB PREDICTION	Dashboard	Predict M Job Pr	y ofile	Log	
	Job Title	Job Location					
	Job Title	Job Location					
	Department	Salary Range					
	Department	Salary Range					
	Description	company profile					
	Description	Company_Profile					
	Requirements	Required Education					
	Requirements	Required Experience					
	Required Experience	Benefits					
	Required Experience	Benefits					
	Employment Type	Industry					
	Employment Type	Industry					
	Function						
	Function						-
							Ŷ
	f a start	Predict Job		Show hidden ico	ns		

Signin Signin Transform Prediction Results Frediction Results	x + 2782 ARNING BASED FAKE I	OB PREDICTION Dashboard	A ^N ☆ Predict	D s My Profile	¢ @	– Se	• ×
	This job p	osting is	100	FIGHE		out	_
	Job Details	GENUINE HIMAN SHUMB					
	Job Title	Job Location					
	Head of Content (m/f)	DE, BE, Berlin					
	Department	Salary					
	ANDROIDPIT	20000-28000					
	Description	Requirements				1	
	Your Responsibilities: Â Manage the English-speaking editorial team and build a team of best-in-class editorsSet	Your Know- How:Â Â Â Â Â Â Â Â Â Â Â Â Â Â Â Â Â Â Â	Â.				
eee 40°C Mostly sunny	Q Search 🏫	🖬 🔉 💿 📮 💼 💽 🛤	^	C ENG	(î) (î)	29-0	17:55 14-2024



7. CONCLUSIONS

Employment scam detection will guide job-seekers to get only legitimate offers from companies. For tackling employment scam detection, several machine learning algorithms are proposed as countermeasures in this paper. Supervised mechanism is used to exemplify the use of several classifiers for employment scam detection. Experimental results indicate that Random Forest classifier outperforms over its peer classification tool. The proposed approach achieved accuracy 98.27% which is much higher than the existing methods.

8. FUTURE SCOPE

The future scope for fake job prediction using machine learning lies in enhancing algorithmic sophistication to detect increasingly complex fraudulent patterns. Integration with natural language processing (NLP) and deep learning can refine models to recognize nuanced deception in job postings and applicant behavior. Collaborations with job platforms and recruitment agencies can facilitate real-time monitoring and prevention of fake job listings, bolstering trust and efficiency in the hiring process. Continuous adaptation to evolving tactics of scammers will be crucial, necessitating a dynamic and agile approach to algorithm development and data analysis. Ultimately, leveraging machine learning in fake job prediction can contribute to a more secure and transparent job market, benefiting both employers and job seekers alike.

9. REFERENCES

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

[2] I. Rish, —An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive Bayes classifier, I no. January 2001, pp. 41–46, 2014.

[3] D. E. Walters, —Bayes's Theorem and the Analysis of Binomial Random Variables, Biometrical J., vol. 30, no. 7, pp. 817–825, 1988, doi: 10.1002/bimj.4710300710. [4] F. Murtagh, —Multilayer perceptrons for classification and regression, Neurocomputing, vol.
2, no. 5–6, pp. 183–197, 1991, doi: 10.1016/0925-2312(91)90023-5.

[5] P. Cunningham and S. J. Delany, —K -Nearest Neighbour Classifiers, Mult. Classif. Syst., no. May, pp. 1–17, 2007, doi: 10.1016/S0031-3203(00)00099-6.

[6] H. Sharma and S. Kumar, —A Survey on Decision Tree Algorithms of Classification in Data Mining, Int. J. Sci. Res., vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.

[7] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems, Heliyon, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.

[8] L. Breiman, —ST4_Method_Random_Forest, Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001,
 doi: 10.1017/CBO9781107415324.004.

[9] B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, —Bagging classifiers for fighting poisoning attacks in adversarial classification tasks," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6713 LNCS, pp. 350–359, 2011, doi: 10.1007/978-3-642-21557-5_37.

[10] A. Natekin and A. Knoll, —Gradient boosting machines, a tutorial, Front. Neurorobot., vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.

[11] N. Hussain, H. T. Mirza, G. Rasool, I. Hussain, and M. Kaleem, —Spam review detection techniques: A systematic literature review, Appl. Sci., vol. 9, no. 5, pp. 1–26, 2019, doi: 10.3390/app9050987.

[12] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, —Fake News Detection on Social Media, ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017, doi: 10.1145/3137597.3137600.

[13] Shivam Bansal (2020, February). [Real or Fake] Fake JobPostingPrediction,Version 1.Retrieved March 29,2020 from <u>https://www.kaggle.com/shivamb/real-or-fake-fakejobposting-prediction</u>

[14] H. M and S. M.N, —A Review on Evaluation Metrics for Data Classification Evaluations,
Int. J. Data Min. Knowl. Manag. Process, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.

[15] S. M. Vieira, U. Kaymak, and J. M. C. Sousa, —Cohen's kappa coefficient as a performance measure for feature selection," 2010 IEEE World Congr.Comput.Intell. WCCI 2010, no. May 2016, 2010, doi: 10.1109/FUZZY.2010.5584447