Evaluating the Performance of Machine Learning Classifiers for Detecting Twitter Spam

Dipalee B. Borse
Research Scholar
North Maharashtra University, Jalgaon

Swati K. Borse, PhD
Associate Professor
SSVPS's P.R. Ghogrey College of Science, Dhule

Vijaya Ahire
Research Scholar
North Maharashtra University, Jalgaon

ABSTRACT
The usage of social networking sites is rising rapidly every day. The popularity of twitter as a microblogging site is huge in normal users as well as illegitimate users. The people with wrong intentions use twitter to spread spam posts which results in phishing, monetary loss, un-useful or noisy data on social media, stealing personal information etc. It becomes extremely important to stop spamming activities. In this paper six machine learning classifiers, which are Logistic regression & Support Vector machine (linear models) and Random forest, K-Nearest Neighbor, Decision tree and Naive Bayes, (nonlinear models), have been implemented on existing data and compared the performance using different parameters such as accuracy, F1-score, recall, precision, f-measure. Among the six classifiers random forest has shown better accuracy followed by K-nearest neighbor classifier for large continuous dataset than small or random dataset. The accuracy is increased from 3% to 13% for large continuous data. Also False positive ratio of random forest and K-nearest neighbor algorithm 0.001 and 0.005 respectively which is much lesser than other algorithms. With lowest accuracy and highest FPR Naive Bayes algorithm performed worst for large datasets.

General Terms
Information security, cyber security Spam detection, Microblog spam detection, low quality content

Keywords
Spam detection, Machine learning, twitter spam detection, information security.

1. INTRODUCTION
Twitter is the fastest growing microblogging site. There are several people who use twitter to express their views and feelings on different topics, share news, and communicate with other people around the world. They can connect with old friends as well as make new friends. People can follow others and can see everything posted by them. These attributes and convenience of twitter makes it well liked among the normal users as well as illegitimate users.

Illegitimate users outspread unwanted messages over Twitter. Such unwanted messages used for phishing, containing malicious links, dispersing wrong information, advertising, etc. are called spam messages [4]. Spam having malevolent links may direct users to different websites which are having malware downloads. Spam content reduces the quality of data on social media which may be used for research purposes and leads to weakening the quality of research outcomes too. Spammers are the users who spread spam. Researchers are continuously working on spam detection on twitter by learning spamming behavior. There are several challenges in this such as spammers can take over legitimate users' accounts and use it for spamming, using different message templates, many times users click on links which are posted by unknown

people without confirmation or accept connection requests sent by unknown people. spammers use different techniques and strategies to spread spam even spammers are constantly evolving their techniques to evade spam. Due to continuously changing behavior of spammers, there is a need to work on spam detection for latest spamming practices. The different existing machine learning techniques for twitter spam detection needs to be studied and evaluate the performance of them.

In this paper, existing supervised machine learning techniques are used on a sample dataset for spam detection and recorded the performance of each algorithm on different parameters.

The paper is structured as follows. Section II elaborates on the relevant literature. Section III, explained the machine learning classifiers used and the reasons for choosing these algorithms. This paper section IV discusses the dataset chosen. Section V explicates the experimental analysis with the performance attributes for each algorithm and the paper is concluded in section VI.

2. RELATED WORK
Spam classification has become a challenging phase for researchers as spammers continuously find different ways of spamming. There are various machine learning classifiers used to solve this problem. Twitter spam detection is achieved using different combinations of features, different aspects of spamming behavior. This section discusses literature description related to this research.

There are a lot of researchers working towards twitter spam detection based on different combinations of features, different aspects of spamming behaviors. The real time spam detection can be done on content and user based features which are easy and faster to compute than graph based features which take more time for collection itself [2], [4].

Chen et.al. [4] Proposed method with evaluation of impact of data related factors (spam and no spam ratio, training data size, data sampling) with 12 light weight features and six ML classifiers. It is found out that the feature of spam tweets varies with time and the performance decreases due to distribution of features changes after a few days of dataset. Sun et al. [2] developed a near real-time spam detection system using a parallel computing technique. They studied the behavior of nine machine learning techniques in terms of stability, scalability and accuracy. The C5.0 and random forest showed higher detection accuracy whereas the random forest proved to be more stable than other classifiers. Lin et al. [3] compared the performance of machine learning classifiers with regard to

detection accuracy, the TPR/FPR, and the F-measure to identify an algorithm that shows considerable detection performance and stability on a large dataset [1] under varying size of training data and different ratios of spam and non-spam. C5.0 and Random Forest have higher detection accuracy from 85% to 90%. Abu-Salih et.al. [6] Constructed a distinctive vector of features by topic based analysis of tweets and implemented machine learning algorithms on them. Also developed a model for spammer detection which overshadows the baseline models. Chinnaiah et.al. [6] Proposed a model that performs a heterogeneous feature analysis on the twitter data streams for classifying the unsolicited messages using binary and continuous feature extraction with sentiment analysis on social network datasets. The suggested method achieves an accuracy of 90.72% when compared with the other most recent methods. Tingmin et al. [7] proposed a deep learning model and evaluated the performance based on proportion of spam and non-spam ratio. The indicated algorithm remains stable with varying spam ratio as compared to existing text based methods and continuous dataset performed slightly better than random dataset. Anisha P Rodrigues et. al [8] focuses on real time spam detection and sentiment analysis on stored and streaming tweets. The multinomial naïve Bayes classifier accomplished 97.78 percent accuracy and LSTM in deep learning performed well with 98.74 percent validation accuracy for detecting spam. Zhu et. al [9] proposed tangrams to extract templates of spam and matching the message to those templates for faster spam detection. Although it is faster for spam detection, the attackers may use different templates each time. Wang et. al. [10] developed a drifted twitter spam categorization method by using MDDT on K-L divergence, where in case of drift KL divergence has steady change patterns between features. MDDT achieves accuracy 98.86 percent. Tajalizadeh et.al.[11] developed a novel stream clustering framework is introduced which enhances the functioning of every stream clustering procedure in which a set of incremental classifiers are replaced by the Euclidean distance function to assign incoming samples to most relative micro clusters with random distribution. DenStream was advanced to INB-DenStream. Proposed approach has shown considerable enhancement in comparison with the standard methods of clustering. Jain et. al. [12] developed a new architecture based on convolutional NN and LSTM with a new semantic layer just before the embedding layer to incorporate semantic knowledge. The proposed SSCL outperforms other ML models with 99.01% accuracy for SMS data and for dataset twitter its 95.48%. El-Mawass et.al.[13] developed a probabilistic graphical model to detect online abuse. The proposed model shows that traditional classifiers have elevated precision and lower recall. Tang et. al. [14] proposed ensemble learning method for imbalance problems in spammer detection. It has combined multiple base classifiers to improve learning performance. To handle imbalanced data during the training stage of base learners, fuzzy logic based oversampling and cost sensitive SVM are used for handling data imbalance. In this method recall raised by 6.5% while precision achieved is 87.53 percent, F-score is 88.7 percent. Ameen et. al. [17] developed a novel spam detection technique in which syntax of each tweet will be learned through WordVector and trained using deep learning & constructed binary classifiers for spam and non-spam classification. PCA is used to analyze the attributes and it shows that extracted attributes are more proficient than features drawn out by standard methods..

3. PROPOSED WORK

For this study ICC dataset by chao chen et.al.[1] has been used. First exploratory data analysis is performed to learn characteristics of data. As a part of data analysis the dataset and correlation between attributes, understanding of the data attributes, handling outliers is studied. After EDA the following classifiers are considered for this comparison analysis for twitter spam detection out of which Logistic regression & Support Vector machine are linear models and Random forest, K- Nearest Neighbor, Decision tree and Naive Bayes, are nonlinear models. Reasons to choose these algorithms:

- The algorithms selected are broadly used by both industry and researchers for spam detection [5]. The performance of these algorithms for tweets datasets with different context need to be explored.
- KNN is selected because it works well for data samples with small dimensions [2].

3.1 Algorithms

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.1.1. Random forest: RF is based on ensemble learning concept usually trained with the "bagging" method which increases overall result. RF builds numerous decision trees for different subsets of the specified dataset and calculates their average in order to increase the dataset's accuracy. It is quite accurate for more trees and also prevents overfitting. For a classification problem, random forest collects a class vote for each tree before classifying using majority vote. In classification problem, the default value of m is

$[\sqrt{p}]$the bare minimum of nodes should be one.
Forest Output Probability

$$(v) = \frac{1}{T}\sum_t^T \quad p_t(v) \qquad (1)$$

3.1.2. Decision tree: Decision tree works on labeled datasets and it is used for classification as well as regression. Decision Tree builds a training model that may be used to determine the target variable's class or value by mastering decision rules derived from the training dataset. In this tree based classifier the internal node acts as features of datasets and branches represent decision rules and the outcome is each leaf node 'i' and each leaf node. The attribute selection is the main challenge in the decision tree implementation where the attributes for each level need to be considered and that has to be identified.

3.1.3. Naive Bayes: Based on the Bayes theorem, Naive Bayes is a probabilistic classifier that predicts using the likelihood of the object. The Bayes theorem can be expressed as follows:

$$P(X) = \frac{P(y)p(y)}{P(X)} \qquad (2)$$

Where, y = class variable and X = {x_1, x_2,... x_n} a set of features, P(y|X) is Posterior probability, P(X|y) is Likelihood probability, P(y) is Prior Probability, P(X) is Marginal Probability. Substitute **X** and expand using the chain rule results,

$$P(x_1, \ldots., x_n) = \frac{P(y)p(y)\ldots P(y)P(y)}{P(x_1)P(x_2)\ldots P(x_n)} \qquad (3)$$

The denominator remains static for the whole dataset. Therefore, the proportionality can be established by removing the denominator.

$P(x_1, \ldots., x_n) \propto P(y) \prod_{i=1}^{n} \quad P(x_i|y)$ (4)

In case of multivariate classification, where class variable *y* may have more than 2 outcomes, the one with highest probability needs to be considered.

$y = argmax_y P(y) \prod_{i=1}^{n} \quad P(y)$ (5)

The above function is generally used for high dimensional data to do text classification.

3.1.4. Support Vector machine: SVM classifies the data into multidimensional space using a border line known as a hyper-plane. The extreme points to the hyper-plane are called support vectors. The gap connecting the support vectors and the hyper-plane has to be increased. The loss function that helps increase the margin is hinge loss.

$c(x, y, f(x)) = \{0 \quad if\ y * f(x) \geq 1\ 1 - y - f(x), \quad else$ (6)

If the projected and actual value have the same sign then the cost is zero, otherwise the loss value is computed. To balance the margin boost and loss, the regularization parameter can be added to the cost function as below [16]:

$min_w \lambda \|w\|^2 \sum_{i=1}^{n} \quad (1 - y_i < x_i, w >)_+$ (7)

3.1.5. K- Nearest Neighbor: KNN checks the resemblance between new and existing data and places the new data based on similarity. KNN does not learn from its available dataset.

In classification, for a given value of K, the KNN will choose the K nearest neighbors of the new data point and then the class having the highest number of data points from all classes of K neighbors; will be assigned to that data point. The distance between data points is calculated using Euclidean metric:

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + \cdots + (x_n - x_n')^2} \quad (8)$$

Finally, the class with the highest probability will be assigned to input x.

$$p(X = x) = \frac{1}{K} \sum_{i \in A} \quad I(y^{(i)} = j) \quad (9)$$

The K in KNN is a hyper-parameter. A smaller K gives the most adjustable fit, with low bias and higher variance. A larger K value is more suitable to deal with outliers. It results in a smooth decision boundary that means low variance but bigger bias.

3.1.6. Logistic Regression: LR also works on labeled datasets. It predicts categorical dependent variable such as Spam or not spam, Yes or No, True or False, 0 or 1, etc. using a given independent variable. It is a probability-based predictive analytic method. This algorithm applies sigmoid function as a cost function. This sigmoid function is used to model the data in logistic regression [15]. It can be written as:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (10)$$

Where $h_\theta(x)$ = value between 0 to 1, x is input and e is the natural log base. When the data is passed to the function, it produces the S-curve shown in the figure below:

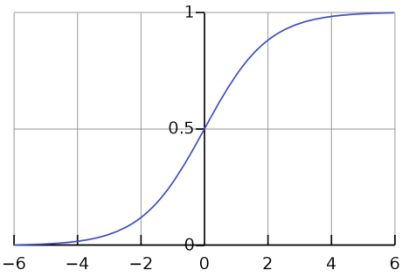


Figure 1: S-Curve

If y=1 we want $h_\theta(x) \approx 1$, $\theta^T x \gg 0$
If y=0 we want $h_\theta(x) \approx 0$, $at \ll 0$

The values over the threshold level are rounded up to 1, while values less than the threshold level to 0.

3.2 Dataset

The existing dataset collected by chao chen et.al. is used [1]. The dataset has 13 lightweight features. The dataset is divided into continuous (dataset 1 & dataset 3) and random (dataset 2 & dataset 4) data. All 4 datasets are used for analysis. Table 1 shows the spam and nonspam ratio of all these four datasets.

Table 1: Spam and non-spam ratio across all datasets

	Dataset1	Dataset2	Dataset3	Dataset 4
Size	10000	10000	100000	100000
Spam	5000	5000	5000	5000
Non Spam	5000	5000	95000	95000

As shown below, the 13 lightweight features of the dataset may be separated into two categories: user-based features and content-based characteristics.

Table 2: Features classification

User Based	Content based
follower count following count favourites count lists count tweets count account age	Url count, hashtag count, retweets count, user mentions count, tweet favourites count, char count, digits count

The user based characteristics shows the behavior of the account such as how long the account existed, followers count of the account, friends count, favourites count, list and tweets count by the user. These are also called profile based or account based characteristics. The Tweet related or content related features are depending upon the tweet text such as count of urls, count of hashtags, count of retweets, count of mentions, count of favorites, count of letters, count of digits in text. The user and content based features are effortless to fetch but easy to avoid as well by spammers. As these are statistical features it requires less processing time for training and testing the model.

4. EXPERIMENTAL ANALYSIS

To analyze the performance of classifiers the above six supervised machine learning classifiers are tested on ICC datasets with the help of metrics such as accuracy, precision,

recall, F-measure, FPR. In this experiment existing datasets are used with all thirteen lightweight user and content or tweet related features from the dataset. This section will discuss the evaluation metrics considered, and performance analysis based on the selected metrics.

4.1 Evaluation Metrics:

Evaluation of all 6 classifiers using accuracy, recall, precision, FPR, F-measure is done. Accuracy is computed as percentage total count of precisely classified spam and non-spam to total evaluated records (11). Precision the ratio of tweets volume categorized rightly as spam to total count of classified spam tweet posts (12). Recall (also called sensitivity/TPR) is the percentage of tweets categorized accurately as spam to the count of real spam messages remaining as spam (13). The FPR can be calculated as the percentage of non-spam tweets wrongly classified as spam tweets in the total number of actual non-spam tweets. And F- measure is calculated as recall and precision's harmonic mean (15).

$$Accuracy = \frac{(TN+TP)}{(TN+TP+FN+FP)} \quad (11)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (12)$$

$$Recall/TPR = \frac{TP)}{(TP+FN)} \quad (13)$$

$$FPR = \frac{FP}{(TN+FP)} \quad (14)$$

$$F-measure = \frac{2*(Recall*Precision)}{Recall+Precision} \quad (15)$$

4.2 Performance analysis:

The dataset 1 & 2 contains 10K records and dataset 3 &4 contains 100K records. The training and testing datasets are divided as a ratio 70:30 and 80:20. The algorithms are tested for different values for hyper-parameters such as random state, k value in K nearest neighbor algorithm, Regularization parameter (c) for support vector machine.

As the results show, the classifiers work better on large continuous datasets than random data.

Random forest has shown Accuracy of 99.23% for dataset 3 followed by decision tree algorithm with accuracy 98.63% for dataset 3 whereas naive bayes worked worst with accuracy 21.55% for dataset 3. The following figure represents the accuracy of each classifier for all 4 datasets. It shows improved results for more training data. All the 6 classifiers had shown improved accuracy for dataset 3 & 4. It clearly shows that all algorithms worked well for dataset 3 having large amounts of continuous data except the naive bayes algorithm.

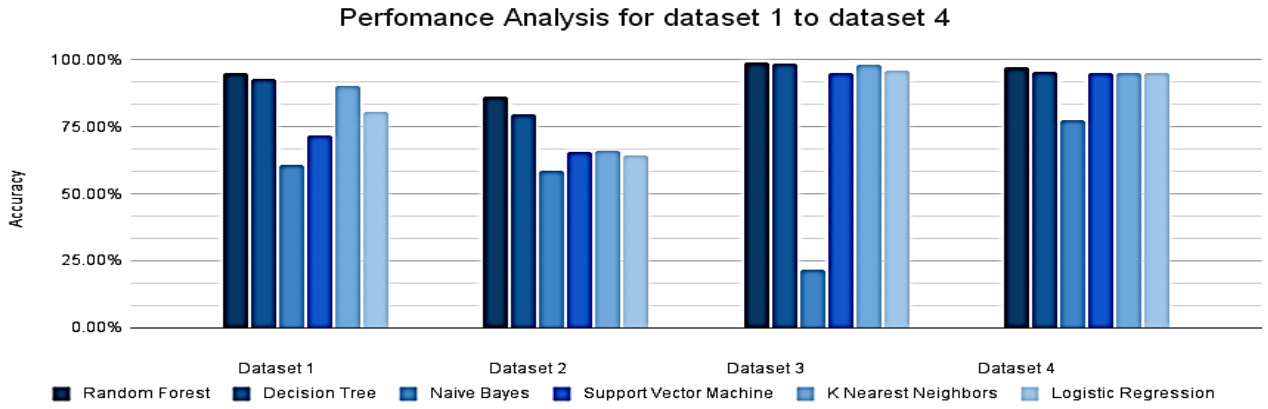
The figure 2 represents the accuracy of each classifier for all 4 datasets. It shows improved results for more training data.

The following table 3, demonstrates the confusion matrices of random forest for all dataset1 to dataset 4

Table 3: Confusion matrices of random forest for dataset 1 to dataset 4

Actual	Prediction							
	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Spam	NonSpam	Spam	NonSpam	Spam	NonSpam	Spam	NonSpam
Spam	965	65	872	158	876	133	470	539
Non Spam	29	941	113	857	21	18970	31	18960

Figure 2: Performance analysis of classifiers for Dataset 1 to Dataset 4



The table 4 shows the performance comparison for dataset 1 to dataset 4 with evaluation parameters accuracy, recall, precision FPR and F1-score.

Random forest has shown Accuracy of 99.23% for dataset 3 followed by 97.15% for dataset 4. Both the datasets 3 & 4 are having 100K records whereas random forest showed 95.30% accuracy for dataset 1 having 10K records. It shows that the accuracy is decreased by 2% approximately for 100K random data and 9% for 10K random data as compared to continuous data. Similar is the case with other classifiers, accuracy of the decision tree is dropped by 3% for 100K random data whereas 13% for 10K random data in comparison with continuous data. K nearest neighbor showed reduced accuracy by 25% for dataset2 in comparison with dataset1 and 3% for dataset 4 than dataset 3. Naive bayes worked worst with accuracy 58.50% for dataset2 which was reduced by 2% from the dataset1.naive bayes particularly works well on smaller datasets hence the accuracy for dataset 3 is 21.55%. All the 6 classifiers have shown improved accuracy for dataset 3 & 4. It clearly shows that all algorithms worked well for dataset 3 having large amounts of continuous data except the naive bayes algorithm. False positive means prediction is positive for negative truth. That is the post is not spam but the test inaccurately shows that it is spam. It is also known as a Type I error in statistics. So in the spam detection case the lower value for FPR indicates higher performance. As shown in

Table 4 Random forest shows lowest FPR value i.e.0.001 for dataset 3 followed by 0.002 for dataset 4. K-Nearest Neighbor having FPR0.005 for large dataset 3 & 4.

Naive Bayes has the highest FPR value of all four datasets. Random forest outperforms other classifiers followed by decision tree and K- nearest neighbor classifier in relation to accuracy, precision, recall and f measure. All algorithms except Naïve Bayes performed well for dataset 3 having large continuous data.

Table 4: Accuracy, FPR, precision, recall and f measure of all classifiers for dataset1 to dataset 4

	Dataset	Accuracy	Recall	Precision	False Positive Rate	F1-Score
Random Forest	**Dataset 1**	95.30%	93.69	97.08	0.030	95.36
	Dataset 2	86.45%	84.66	88.53	0.116	86.55
	Dataset 3	**99.23%**	86.82	97.66	**0.001**	91.92
	Dataset 4	97.15%	46.58	93.81	**0.002**	62.25
Decision Tree	**Dataset 1**	92.85%	93.34	92.7	0.077	93.02
	Dataset 2	79.85%	80.2	79.8	0.205	80
	Dataset 3	**98.63%**	88.29	84.62	0.008	86.42
	Dataset 4	95.45%	56.61	53.84	0.025	55.19
Naive Bayes	**Dataset 1**	60.60%	97.96	56.81	0.791	71.92
	Dataset 2	58.50%	96.25	55.15	0.802	70.12
	Dataset 3	21.55%	97.77	5.81	0.824	10.96
	Dataset 4	77.36%	43.51	9.98	0.208	16.24
Support Vector Machine	**Dataset 1**	71.85%	80.39	69.64	0.372	74.63
	Dataset 2	65.60%	78.16	62.88	0.473	69.69
	Dataset 3	96.231%	3.6	76.4	0.246	0.48
	Dataset 4	**95.12%**	0.31	75	0.012	0.43
K-Nearest Neighbor	**Dataset 1**	90.20%	87.48	93.08	0.069	90.19
	Dataset 2	65.85%	67.09	65.99	0.354	66.54
	Dataset 3	**98.39%**	77.4	88.67	**0.005**	82.65
	Dataset 4	95.18%	11.81	56.52	**0.005**	19.53
Logistic Regression	**Dataset 1**	80.50%	83.3	79.74	0.225	81.48
	Dataset 2	64.35%	48.24	70.65	0.198	57.33
	Dataset 3	**95.85%**	2.3	5.45	0.0313	11.35
	Dataset 4	94.99%	0.1	3.85	0.0013	0.2

5. CONCLUSION

The popularity of twitter is rising every day. With genuine user there are many spammers as well on twitter. Spammers are using twitter to spread spam posts which includes unwanted messages, hijacking trending topics, hijacking normal users accounts to spread spam posts etc. Researchers are working on spam post detections and prevention continuously. This paper discusses the implementation and performance evaluation of six machine learning techniques on existing datasets containing random and continuous data. Here the performance of the algorithms according to Accuracy, Recall, Precision, False positive rate, & F-measure is checked. Among all six classifiers random forest has shown accuracy upto 99.23% and reduced FPR upto 0.001 for large continuous dataset followed by Decision tree with accuracy 98.63% and KNN with accuracy up to 98.20% and FPR upto 0.005 for large continuous data. The accuracy is decreased slightly for large random dataset whereas major reduction in accuracy is recorded for smaller datasets. Among all the classifiers Naive bayes classifier performed worst for large continuous dataset. This research helps to understand the attributes and parameters to improve performance of the classifier. In future the aim is to collect balanced twitter dataset and to check the performance of each of the classifier also designs a new framework for spam classification.

6. REFERENCES

[1] Chen, Chao, Jun Zhang, Xiao Chen, Yang Xiang, and Wanlei Zhou. "6 million spam tweets: A large ground truth for timely Twitter spam detection." In 2015 IEEE international conference on communications (ICC), pp. 7065-7070. IEEE, 2015.

[2] Sun, Nan, Guanjun Lin, Junyang Qiu, and Paul Rimba. "Near real-time twitter spam detection with machine learning techniques." International Journal of Computers and Applications 44, no. 4 (2022): 338-348

[3] Lin, Guanjun, et al. "Statistical twitter spam detection demystified: performance, stability and scalability." IEEE access 5 (2017): 11142-11154.

[4] Chen, Chao, Jun Zhang, Yi Xie, Yang Xiang, Wanlei Zhou, Mohammad Mehedi Hassan, AbdulhameedAlElaiwi, and MajedAlrubaian. "A performance evaluation of machine learning-based streaming spam tweets detection." IEEE Transactions on Computational Social systems 2, no. 3 (2015): 65-76.

[5] Borse, D., Borse, S. (2022). State of the Art on Twitter Spam Detection. In: Iyer, B., Crick, T., Peng, SL. (eds) Applied Computational Technologies. ICCET 2022. Smart Innovation, Systems and Technologies, vol303. Springer, Singapore. https://doi.org/10.1007/978-981-19-2719-5_46

[6] Abu-Salih, Bilal, Dana Al Qudah, Malak Al-Hassan, Seyed Mohssen Ghafari, Tomayess Issa, Ibrahim Aljarah, Amin Beheshti, and Sulaiman Alqahtan. "An Intelligent System for Multi-Topic Social Spam Detection in Microblogging." arXiv preprint arXiv:2201.05203 (2022).

[7] Wu, Tingmin, Shigang Liu, Jun Zhang, and Yang Xiang. "Twitter spam detection based on deep learning." In Proceedings of the australasian computer science week multiconference, pp. 1-8. 2017.

[8] Rodrigues, Anisha P., Roshan Fernandes, Adarsh Shetty, Kuruva Lakshmanna, and R. Mahammad Shafi. "Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques." Computational Intelligence and Neuroscience 2022 (2022).

[9] Zhu, Tiantian, Hongyu Gao, Yi Yang, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, and Alok N. Choudhary. "Beating the artificial chaos: Fighting OSN spam using its own templates." IEEE/ACM Transactions on Networking 24, no. 6 (2016): 3856-3869.

[10] Wang, Xuesong, Qi Kang, Jing An, and Mengchu Zhou. "Drifted Twitter spam classification using multiscale detection test on KL divergence." IEEE Access 7 (2019): 108384-108394.

[11] Tajalizadeh, Hadi, and Reza Boostani. "A novel stream clustering framework for spam detection in Twitter." IEEE Transactions on Computational Social Systems 6, no. 3 (2019): 525-534.

[12] Jain, Gauri, Manisha Sharma, and Basant Agarwal. "Spam detection in social media using convolutional and long short term memory neural network." Annals of Mathematics and Artificial Intelligence 85.1 (2019): 21-44.

[13] El-Mawass, Nour, Paul Honeine, and Laurent Vercouter. "SimilCatch: Enhanced social spammers detection on twitter using Markov random fields." Information Processing & Management 57, no. 6 (2020): 102317.

[14] Tang, Wenbing, Zuohua Ding, and Mengchu Zhou. "A spammer identification method for class imbalanced weibo datasets." IEEE Access 7 (2019): 29193-29201.

[15] https://vkosuri.github.io/CourseraMachineLearning/

[16] https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

[17] Ameen, Aso Khaleel, and Buket Kaya. "Spam detection in online social networks by deep learning." 2018 International Conference on Artificial Intelligence and Data Processing (IDAP). IEEE, 2018.