

Detection of Spam Tweets by Using Machine Learning Methods

Shakeeb Ahmed Khan*

Dept. of Computer Science, University of Southern Punjab, Multan, Pakistan.
Corresponding Author Email: shakeebahmedkhan.sherwani@gmail.com

Shoaib Ali Qureshi

Dept. of Computer Science, Hameeda Rasheed Institute Of Science And Technology, Multan, Pakistan.. shoaib.ali357@hotmail.com

Qandeel Asghar

Dept. of Computer Science, University of Southern Punjab, Multan, Pakistan.
qandeel@usp.edu.pk

Rana Muhammad Nadeem

Department of Computer Science Govt. Graduate College
Burewala.rananadim@hotmail.com

Talha Farooq Khan

Dept. of Computer Science, University of Southern Punjab, Multan, Pakistan.
talhafarooqkhan@gmail.com

ABSTRACT

Nowadays, we live in a digital world where we can connect and communicate with anybody all over the world with the help of social networking sites like Twitter and Facebook. Among many other social networking sites, Twitter is a rapidly growing platform. Due to its popularity, it has to face many spammers, which spread forged, bogus, and misleading information. Due to this spamming, it is difficult for researchers to get un-spammed and authentic information from Twitter. In this paper, we introduce an autonomous model that classifies between spammer and legitimate account tweets using machine learning algorithms. We are using API provided by Twitter for the dataset and apply four machine learning classifiers: Decision tree, Support Vector Machine, Naïve Bayes, and Logistic Regression. These machine learning classifiers evaluate on the basis of accuracy, precision, recall and F1-score. The result shows that Naïve Bayes classifier achieve the best result with 96.83 percent accuracy among four classifiers.

Keywords: Spam, Twitter, Machine Learning, Classifiers, Social Networking Site

Introduction

Social networking platforms like Twitter and Facebook turn this world into a social village. Researchers estimate that more than 50% of the world's population connects using different social networking platforms [1]. Popularity and number of followers play a vital role on these platforms. In recent years, social media platforms, particularly Twitter, have become prominent spaces for global communication, enabling users to broadcast ideas and share messages with a large audience [2]. Twitter is one of the most famous online social networking sites with a vast user base. Twitter is a small weblog site where users type messages that contain up to 280 characters. It is called **Tweet**. Twitter also empowers its users to include four photos, a gif, and a video in their tweets. According to the latest statistics, Twitter has more than 330 million active users per month [3]. This exponential increase in Twitter users allows a number of users to connect and share information. With over 42 million new accounts created each month, Twitter's openness has unfortunately attracted spamming operations, including phishing attacks, malicious code injections, and malware distribution [4] [5]. These activities pose significant privacy and economic risks as attackers seek access to personal information that is valuable on the black market [4] [6] [7]. Advanced persistent threat (APT) reports from 2013 indicate that social media usage increases the probability of successful attacks, with 92% of experts agreeing [8]. Malicious profiles constitute around 8%–10% of all social media accounts [9]. Fake or spam accounts generate and spread fabricated, inappropriate, and bogus information, which leads to misguiding and astray their followers.

Spam means sending irrelevant or unrequested messages over the internet to an enormous number of users. Moreover, Spam tweets are irrelevant or useless tweets in the form of text or the form of another webpage link. There are different intentions behind spam messages, such as advertising, spreading viruses, etc. One thing to keep in mind about spam is that it may be possible for one user to be considered spam, and for another user, it is totally helpful. In recent years, there has been an exponential increase in spamming, which makes it difficult for researchers to get un-spammed and authentic information from social networks. However, the

popularity of Twitter also attracts many spammers, which leads to the spread of false and bogus information. Twitter spams usually refer to tweets that contain promotions or commercial URLs, fake news, abusive content, or messages that redirect users to harmful URLs. Researchers have observed that an average automated spam posting account posts at least 12 tweets per day within well-defined activity periods. One out of every 200 messages on social networking websites and one out of 21 tweets is considered spam [10].

Numerous approaches have been proposed to detect spam on social media platforms such as Twitter. Traditional methods include the use of blocklists, which prevent access to malicious URLs or IP addresses. However, generating and maintaining blacklists is resource-intensive in terms of both time and effort. Furthermore, with the increasing use of shortened URLs in recent years, attackers have found ways to bypass blocklists, making such methods less effective [11] [12]. Another technique involves the deployment of honeypots strategically placed within social media environments to attract and monitor malicious accounts [13] [14]. While honeypots are commonly utilized for spam and attack detection, they face significant limitations, including poor adaptability, scalability issues, the need to process large volumes of data, and limited portability.

Moreover, honeypot systems are often complex and require expert intervention for implementation, making them time-consuming to deploy [11] [15]. Recent research proposes several solutions to address these challenges, focused on the development of machine learning (ML) models to detect spam and harmful behavior on social media more efficiently. In this proposed methodology, we learn the difference between tweets published by spammers and tweets published by legitimate users. The main goal of this proposed methodology is to fetch useful features from data, which can be used in the training of machine learning algorithms. Despite the efforts in using machine learning algorithms for spam detection, evolving attack strategies and data imbalance make it challenging to maintain detection accuracy. As attackers adapt, automated systems may fail to recognize new spam types, reducing the effectiveness of spam detection systems [16]. After training, these machine learning algorithms are able to differentiate between

spamming and legitimate tweets. Our study important contributions are following:

- Twitter spam detection in real-time
- We meticulously compare the performance of four machine learning classifiers: Support Vector Machine, Decision tree, Naïve Bayes, and Logistic Regression. These algorithms are thoroughly analyzed for their detection performance and efficiency.
- We design a framework to evaluate the detection scheme.

Literature Review

Recently, the concept of Artificial Intelligence (AI) has been adopted in academia to detect spam information on social networking sites like Facebook, Twitter, and so on. For instance, in [17], authors proposed a methodology of a real-time spam tweet detection system that provides online detection results and features of specific Twitter accounts. This system detects spam based on account and feature content extraction. This sentence seems written in the passive voice. Rewrite it in the active voice. In [18], authors proposed a methodology to detect spam tweets with the help of the Deep Learning Model (CNN) and information on spam tweet user accounts like age, followers, etc. This work also compares the Deep learning model accuracy with multiple machine learning models, i.e. (SVM, Decision Tree, Random Forest, and Logistic Regression).

Moreover, in [19], the authors presented a system with practical implications for detecting doubtful or spammer user accounts that spread spam tweets on Twitter. The system is designed in such a way that it starts with a semi-supervised learning model and is updated on a supervised learning model at training time. This system not only detects spammer accounts, but it also detects and removes matching tweets. Likewise, in [20], this paper proposes a system that extracts tweet and user features and stops spammers from making spam tweets from new accounts created, offering a practical solution to a common problem. In this paper, we apply machine learning algorithms and achieve 91% accuracy with the neural network model. In [21], the authors discussed a framework that elaborates tweet text into clusters and finds similar tweets with the help of the Siamese recurrent network, providing a practical method for identifying spam. As a result,

this framework classifies them as spam or ham with recall and precision. In [22], this paper proposed a methodology that quickly detects and classifies spam accounts on Twitter, offering a practical tool for social media managers. It applies a sentimental analysis algorithm to the tweets of spammer accounts and analyzes their tweet patterns. Once the analysis is complete, spam accounts will be removed from legitimate accounts lists. In [23], a system finds out the bogus news and bogus users on the Twitter network, and it has also provided the credibility of tweets with the help of the system, offering a practical solution to the problem of fake news.

However, in [24], a supervised learning algorithm was applied to detect spamming accounts, collecting data from Sina Weibo nearly 30000 then building a labeled dataset of users and manually categorizing users into spammer or non-spammer, extracting features from user messages and social behavior and applying into SVM. It provided the success rate of spammer and non-spammer gains of 99.1% and 99.9%. In [25], they proposed a methodology that forecasts the mobility of humans through the Twitter dataset. It identified the following user location on the basis of its historic pattern through the Twitter social network. In [26], the authors describe a framework that detects spam accounts or spammers and classifies them as legitimate users on Twitter. This proposed methodology relies on content-based and user-based features. Content-based features include the average length of a tweet, the Number of replies/mentions to friends, the average tweet length, and the average use of hashtags. On the other hand, user base features depend upon the Number of friends, Number of followers, and the repute of a user. After extraction of features, four machine learning algorithms (SVM, K-NN Nearest Neighbor, and Naïve Bayes) are trained and tested on these user base and content base features dataset, and in the end, compare their performance. In [27], the authors proposed a bottom-up approach using a dynamic relational network, which they used to examine the credibility of news based on the content of the nodes. In [28], the author proposed a methodology based on three steps: data preprocessing, features selection, and applying three machine learning algorithms i.e. (Support Vector Machine, Logistic Regression, and Decision Tree).

Data preprocessing includes:

- Dividing the data into training and testing sets,
- looking for missing or null values in the dataset, and
- Etching out features.

After applying preprocessing, extract features from the dataset. There are some benefits of extracting features before applying ML, like better accuracy, minimal over fitting, and Minimal training time. After the features are extracted, three ML models are applied, and their performance is compared.

In [29], the author proposed a method for detecting malicious profiles by first identifying fake accounts and their interactions as legitimate users. It utilizes collaborative and content-based filtering from recommender systems to classify accounts similar to harmful profiles, achieving superior accuracy and reduced loss compared to traditional machine learning techniques. In [30], this research proposes a deep learning-based approach to identify spam tweets and harmful profiles. We initially offer automated honeypot-based false profile detection so that the profiles can communicate with them as authentic profiles. After that, we identify individuals who engage in harmful profiles. Lastly, we use a convolution neural network technique to gather their shared material in order to identify spam tweets. We propose to leverage recommender system collaborative filtering and content filtering algorithms to identify spam and accounts that resemble malicious profiles and spam that convolution neural networks have detected. Compared to regular learning algorithms, we obtain a highly compelling and intriguing conclusion with a lower loss and a higher accuracy of 99.23%. In [31], researchers used a study-based approach to detect malicious profiles and spam tweets. It employs an automated honeypot to identify fake profiles and uses a convolutional neural network algorithm to detect spam content shared by malicious profiles.

Additionally, we utilize collaborative and content filtering algorithms to identify accounts similar to harmful profiles, achieving high accuracy (99.23%) and reduced loss compared to traditional methods. In [32], the author proposed a model consisting of two components that analyze the input dataset, including tweets and meta-data like follower count and actions. The first component focuses on tweet text, using the GLoVe model to extract features and detect spam with an LSTM model.

The second component analyzes meta-data and heuristic features, such as tweet length and punctuation, classifying them with a CNN. The system makes the final decision by combining the outputs of both models.

The proposed research [33] acknowledges that due to the immense popularity of Twitter, it has become a target for spam which ever lacks. In an effort to resolve the issues presented above, a semi-supervised learning method was applied using YATSI and live Twitter data. YATSI aimed to improve the accuracy of spam detection on social networks through the Twitter API. While performing the experiments with KNN, SVM, and NB classifiers, SVM showed the most efficient performance. In this research [34], the study focused on enhancing spam detection in Natural Language Processing (NLP) by comparing traditional machine learning algorithms, such as SVM and RF, with deep learning models like LSTM, BiLSTM, and ARABET, particularly for Arabic dialects. The goal is to improve both the accuracy and efficiency of spam detection, addressing limitations of traditional methods that often require extensive data and time. In [35], this article reviews publicly accessible evaluation datasets and analyzes their limitations in identifying deceptive opinion spam. This study systematically outlines the key features and methodologies proposed to address the challenge of detecting misleading opinion spam. This paper [36] systematically reviews key features and methodologies introduced for the detection of misleading opinion spam. Conducted a comprehensive systematic literature review (SLR) aimed at identifying, categorizing, and comparing current Twitter spam detection techniques. This SLR examines 70 relevant publications from 2010 to October 2022, presenting an in-depth analysis of approaches for detecting Twitter spam. The review highlights that most methods focus on analyzing textual content (tweets) and utilizing Machine Learning (ML)-based algorithms. These ML techniques differ primarily in their use of various clustering and classification algorithms, each relying on distinct feature selection strategies. The classification approaches rely on several types of feature analyses, including content-based, user-based, tweet-based, network-based, and hybrid analyses.

In [37], the author proposed an unsupervised method that does not rely on labeled

datasets. The authors introduce a fully unsupervised approach to identify spammers by analyzing a user's peer acceptability within a social network. Peer acceptability is measured based on shared interests in various topics. The key innovation is the development of an utterly unsupervised spammer detection technique centered on user acceptability. However, this method provides a valuable alternative to traditional classifiers for spam detection.

Proposed Methodology

The proposed spam detection framework contains four main modules: fetching data from Twitter API, Data Pre-processing, applying Latent Semantic Analysis (LSA) to basic spam keywords, applying a machine learning algorithm, and comparing their performance. The flow diagram of the proposed system is shown in Figure1.

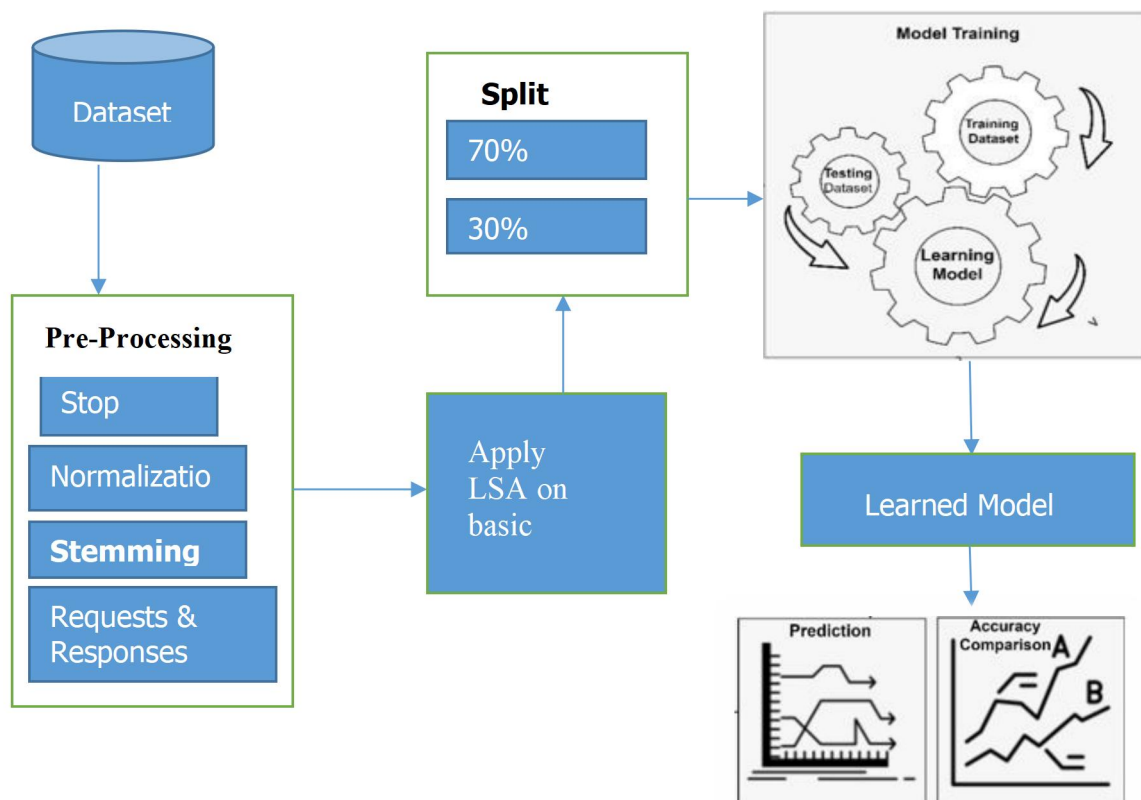


Figure. 1 System Architecture

We used the social honeypot dataset as sample data, which contained more than 5 million tweets from the Twitter API, as shown in Table 1. We have not been able to access all the tweets of a single user, so we collected random individual tweets using

the honeypot. This dataset contains tweets from both spam and legitimate user accounts. Multiple features, such as date, time, etc., have been identified in the dataset.

Table 1 Sample Dataset

Content Pollute Tweet	2333690
Legitimate User Tweet	3246376

Real-world data is often in raw form, which generally contains errors such as data instability, null values, irregularity, and unessential data.

Preprocessing

The intermediate steps of data preprocessing applied to the thematic corpus of collected tweets involve several techniques to refine the raw data, making it more comprehensible and ready for analysis. This process includes:

- Remove invalid data (e.g., poorly structured or incomplete entries).
- Removal of duplicate data.
- Removal of special characters.
- Removal of stop words.
- Deletion of numbers, emoticons, and emojis.
- Convert words into root words by stemming.
- Erasing of those words which are like URLs,

For this data preprocessing, we use a natural language tool kit (NLTK) of natural language programming (NLP) for the removal of stop words and stemming. Stop words are those words that do not add much meaning to a sentence. We take out the stop words from the sentences without loss of the meaning of sentences. Stemming is also removed from the sentences with the help of the NLTK library. When two or more words have the same root, like age or age, the search includes these types of words that should be considered as exact words, which are the root words. After preprocessing, we labeled the spam tweet with “0” and legitimate user tweets with “1,” as shown in Figure 2 and Figure 3.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

	Tweet	Time	spam
	THE BURLESQUE BOOTCAMP SYDNEY - Open Date tick...	2009-11-10 15:46:05	0
	THE BURLESQUE BOOTCAMP SYDNEY - Open Date tick...	2009-11-10 15:46:40	0
	THE BURLESQUE BOOTCAMP SYDNEY - Open Date tick...	2009-11-10 15:47:03	0
	Come to "The Burlesque Bootcamp - Sydney" Satu...	2009-11-10 15:56:03	0
	21st Century Pinups write about our girls perf...	2009-11-10 16:05:05	0
	The Burlesque Bootcamp – Coming to Sydney! ...	2009-11-10 17:09:32	0
	ATTN MELBOURNE: We have some group specials av...	2009-11-10 21:22:14	0
	Story on our lovely Vivi Valentine, on contrac...	2009-11-10 21:29:31	0
	and also Black Flamingo in Berlin with Catheri...	2009-11-10 21:34:10	0
	I posted 69 photos on Facebook in the album "F...	2009-11-10 23:23:12	0
	Tweet	Time	legitimate
	Taxing — at SLC Salt Lake City International...	2009-11-20 20:42:48	1
	Almost home! — at PDX Portland International ...	2009-11-20 18:11:01	1
	Lunch! — at Ten01 http://gowal.la/s/Awa	2009-11-20 15:04:42	1
	Mm ... books — at @Powells http://gowal.la/s/6fe	2009-11-20 15:03:06	1
	@AltF4LJ Drama ... biz trip. All booked up. WILL ...	2009-11-20 14:37:54	1
	... streetcarring in Portland!	2009-11-20 10:00:50	1
	In bed. Wishing he could stay another night —...	2009-11-20 00:47:54	1
	@nathanrawlins ... will do! Having a lot of fun!	2009-11-20 00:44:10	1
	@flpatriot ... :)	2009-11-20 00:42:58	1
	I wish I had more free time. I'd LOVE to see you!	2009-11-19 18:16:40	1

Figure.2 Spam Tweets

Figure.3 Legitimate tweets

After data preprocessing, we can't give the text data to a machine learning model for training. We have to convert it into numerical format before training. For the purpose of converting the dataset into numerical format, we use the Bag of Words (Bow) approach, in which we embed words on the basis of their frequencies. The bow model learns a vocabulary list from a given corpus and represents each document based on some counting methods of words. This model provides the score of each word in a sentence and also defines the sentiment score of encoded emojis. It also understands acronyms like LOL, etc. In this part, we will explore the model performance of using Bow with supervised learning algorithms. Here is the workflow in this part.

Step 1: Preprocess raw reviews to cleaned reviews

Step 2: Generate a Bag of Words (Bow) by using Count Vectorize, Term Frequency (TF), and inverse document frequency (IDF vectorize) in the sklearn library.

Step 3: Convert the review text into a numerical view with the help of feature

vectors.

Step 4: Fit feature vectors to supervised learning classifiers, such as Naive Bayes, Support Vector Machine, Decision Tree, or Logistic regression.

Step 5: Improve the model performance by Grid Search.

Now, we have cleaned the dataset in numerical format for machine learning classifier training and testing. We extract features from this dataset and then give them as input to the machine learning classifiers. Table 2 explains the top 10 features with the smallest and largest coefficients.

Table 2. Features with Largest and smallest coefficient

Sr. No	Features with Largest Co-efficient	Features with Smallest Co-efficient
1	Gowal	Ly
2	La	Bit
3	Di	Http
4	It	To
5	Niet	About
6	Rt	This
7	Via	You
8	That	With
9	Een	From
10	My	New

These machine-learning classifiers are trained and tested on the dataset. Different classification algorithms, i.e., Logistic Regression, Support Vector Machine, Naïve Bayes, and Decision tree Classifier, are used for training and testing. We split the dataset into training and testing parts. We distribute the dataset at a ratio of 70:30, using 70 percent for training and 30 percent for testing. After splitting the dataset, we train the model by using the training dataset. The researchers then test the trained model on the testing dataset. The team also compares all the machine learning classifiers, as shown in Figure 4 and Table 3.

Experiment and Results

In this section, we describe the results and effectiveness of the experiment we conducted. Through this proposed methodology, we attempt to answer the following two questions.

- How functional is the suggested methodology as compared to the existing machine learning approaches?
- How is the practical architecture of this proposed methodology compared to the architecture of the existing approaches?

To answer these questions, we meticulously present the dataset on which our experiment is based. We then conduct a comprehensive comparison of the performance of supervised learning classifiers in terms of Accuracy, Precision, Recall, F1-score, and provide a detailed analysis and discussion of the results.

Dataset

We used the Twitter honeypot dataset in our methodology. We have not been able to access all the tweets of a single user, so we collect random individual tweets by using the honeypot dataset. This honeypot dataset contains five files: spammer and legitimate Twitter account following dataset, spammer and legitimate Twitter account tweets dataset. The primary purpose of this research is to distinguish between spammers' and legitimate users' tweets. So we just the spammer and legitimate user account tweet dataset consider. The dataset of tweets was collected from 2 April 2017 to 10 July 2017. Each tweet has particular attributes like tweet text, retweets, Time, and date. Some tweets also contain URLs and hashtags. We consider tweets with correct links or URLs and neglect tweets with invalid links or URLs. We check these tweet links or URLs by using the Python library 'Requests and Responses. Table 2 provides the general statistics of the dataset.

Table 2 Characteristics Of Used Dataset

Sr. No	Property	Number
1	Number of Tweet	5580066
2	Number of Spam User Tweet	2333690
3	Number of Non-spam User Tweet	3246376

4	Number of Twitter User	5580066
5	Number of Tweets with URL or Links	797152

Experimental Setup

In this experiment, we used the Anaconda Navigator tool for back-end coding. We designed our proposed system for training and testing machine learning classifiers in Python. Moreover, we used Twitter API to fetch a tweet dataset. Finally, we ran the experiments on Windows OS (i.e., Windows 10) with a 3.00 GHz Intel Core i5 processor and 6 GB of RAM.

Numerical Metrics

We considered four metrics to analyze the performance of machine learning classification models. These four assumed metrics are Accuracy, Precision, Recall, and f-1 score. We utilize accuracy to calculate the detection rate, while we use precision, recall, and F1 scores to evaluate the performance. Accuracy is the ratio of the total number of correctly classified data of both classes over the total number of all data in the dataset. Precision is the ratio of the number of correctly classified data to the total number of data. Recall defines the ratio of the number of instances correctly classified to the total number of predicted instances. Below are the formulas for these metrics:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F-1score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where,

TP= True Positive TN=True Negative FP= False Positive FN=False Negative

Evaluation

In this section, we evaluate our proposed approach. We have applied multiple machine learning classifiers, such as Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression (LR) approaches, to our

Online ISSN

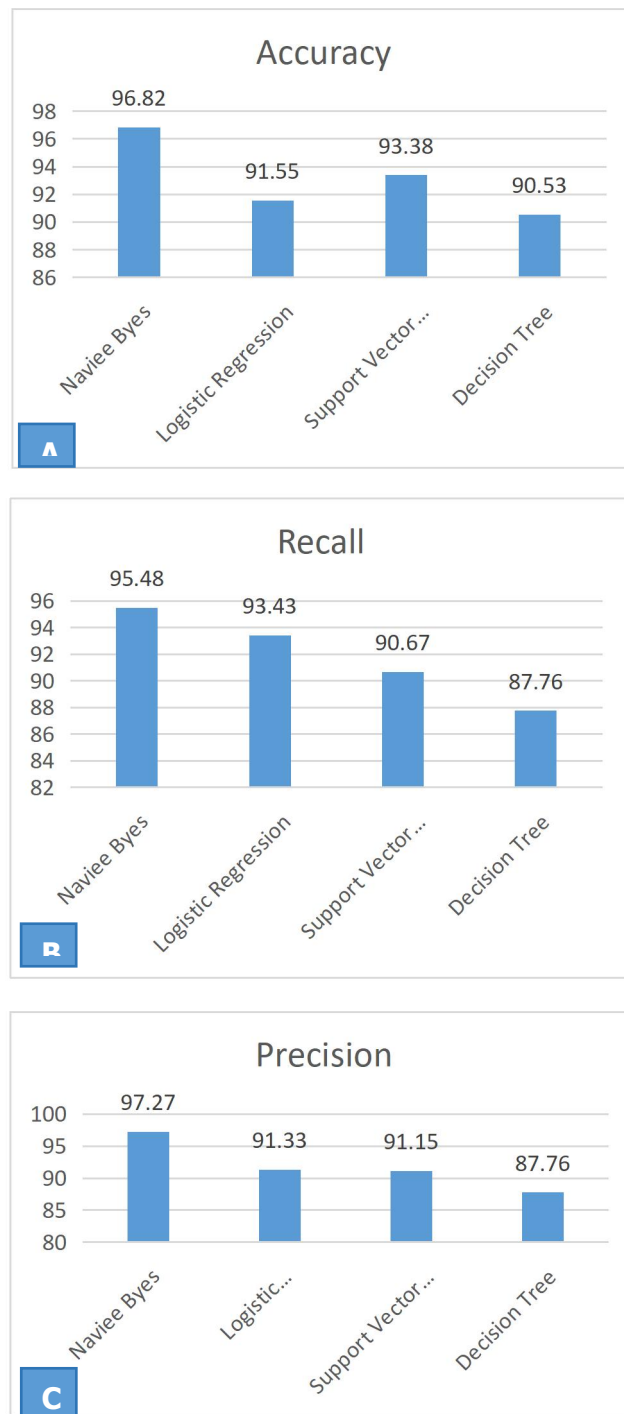
3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

proposed scheme. The outcomes of our approaches and comparison of these algorithms in terms of accuracy, precision, Recall, and F1-score are shown in Figure 4 (a), (b), (c), and (d).



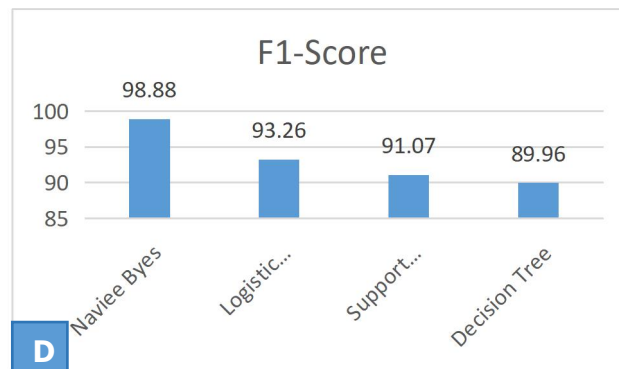


Figure. 4. Performance Comparison result (Accuracy, Precision, Recall and F1-Score) over dataset

Table 3 shows the comparison of accuracy, Recall, precision, and F1-score of the previous machine learning approach vs. our approach. We show the best outcomes in bold. You can notice that Naïve Bayes (NB) achieves 96.82% accuracy, Logistic Regression (LR) gives 91.55% accuracy, Support Vector Machine (SVM) gives 93.38% accuracy, and Decision Tree (DT) gives 90.53% accuracy. Our proposed methodology gives better performance than all considered approaches. The only difference between our approach and previous approaches is that we use a Latent semantic analysis (LSA) approach in our methodology. First, we designed a dictionary of LSA keywords. Free, lots, win, trade, good, great, make, create, twitter, followers, check, gain, buy, account, get, making, online, want.

Table 3. Comparison Of Our Approach Vs Traditional Approach

		Our Approach	Traditional Approach
Accuracy	Naïve Bayes	96.82%	33%
	Logistic Regression	91.55%	11%
	Support Vector Machine	93.38%	37%
	Decision Tree	90.53%	39%
Precision	Naïve Bayes	97.27%	91.50%
	Logistic Regression	91.33%	84.11%
	Support Vector Machine	91.15%	83.50%
	Decision Tree	87.76%	76.36%

Recall	Naïve Bayes	95.48%	91.13%
	Logistic Regression	93.43%	83.50%
	Support Vector Machine	90.67%	83.50%
	Decision Tree	87.76%	72.70%
F1-Score	Naïve Bayes	98.88%	91.05%
	Logistic Regression	93.26%	83.43%
	Support Vector Machine	91.07%	82.69%
	Decision Tree	89.96%	71.09%

After the dataset's preprocessing, it will pass through the LSA analyzer. If four keywords from this list are shown in any tweet, it will be tagged as a spam tweet, and less than four keywords will be tagged as a Ham tweet. The figure 5 shows the comparison of all four classifiers with used evaluating measures. The graphical representation of ML classifiers Logistic Regression, Naïve Bayes, and Support vector Machine after testing on the dataset is shown in Figure 6 (a), (b), and (c).

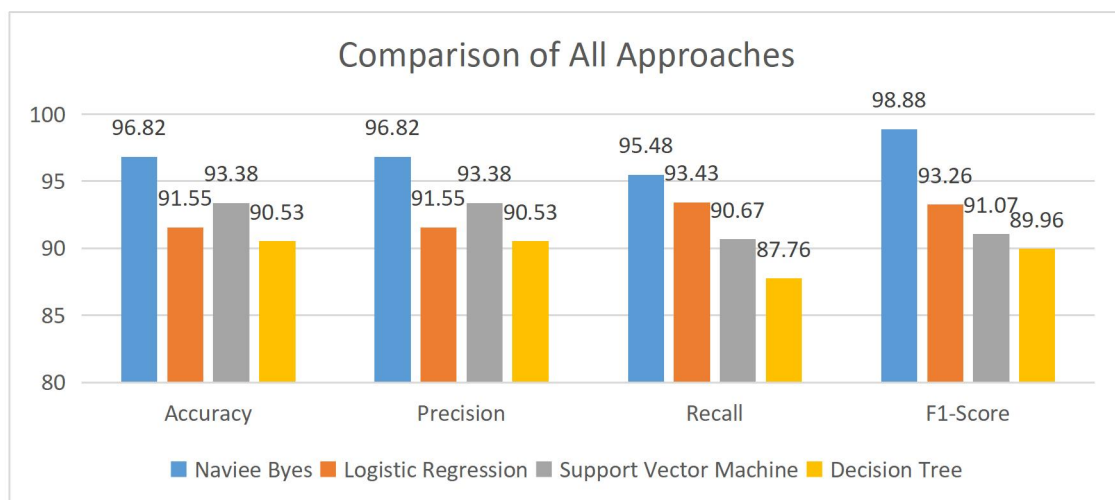


Figure 5: Represents comparison between all four machine learning classifiers with all four evaluating measures.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

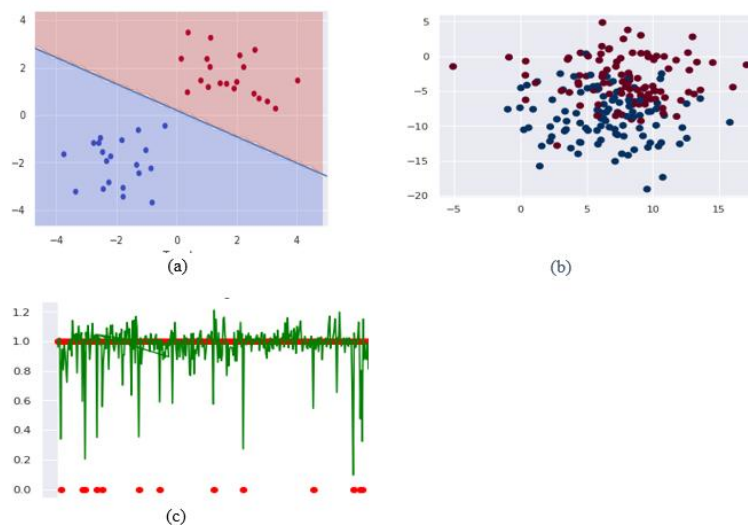


Figure.6. Graphical representation of ML classifiers

Conclusion

In conclusion, we apply different classification algorithms (Decision tree, Support Vector Machine, Naïve Bayes and Logistic Regression classifiers) under different scenarios to evaluate their performance in detecting Twitter spam in terms of detection accuracy, the True Positive Rate TPR, the False Positive Rate FPR, and the F-measure and study their performance in terms of stability, and scalability using tweets datasets we have collected as shown in figure 4 and table 3. Our results indicate that the Naïve Bayes classifier gives the most outstanding performance. By using the Naïve Bayes classifier, we acquired 96.8% accuracy, as shown in Figure 4 and Table 3.

References

- [1] "Social media statistics and facts," [Online]. Available: <http://www.statista.com/topics/1164/social-networks>, [Accessed 02 03 2021].
- [2] T. Wu, S. Wen, S. Liu, J. Zhang, Y. Xiang, M. Alrubaian and M. M. Hassan, "Detecting Spamming Activities in Twitter Based on Deep-Learning Technique,," *concurrency and computation practice and experience*, 2017.
- [3] "Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019,," [Online]. Available:

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>. [Accessed 02 03 2021].
- [4] B. Eshete, A. Villafiorita and K. Weldemariam , "BINSPECT: Holistic Analysis and Detection of Malicious Web Pages," in *Security and Privacy in Communication Networks*, 2012.
- [5] A. Aggarwal, A. Rajadesingan and P. Kumaraguru, "PhishAri: Automatic realtime phishing detection on twitter," *IEEE*, 2012.
- [6] M. S. Rahman, T.-K. Huang, H. V. Madhyastha and M. Faloutsos, "Efficient and Scalable Socware Detection in Online Social Networks," in *Security'12: Proceedings of the 21st USENIX conference on Security symposium*.
- [7] B. Eshete, A. Villafiorita, K. Weldemariam and M. Zulkernine, "EINSPECT: Evolution-guided analysis and detection of malicious web pages," in *Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference*, 2013.
- [8] "ISACA. Advanced Persistent Threat Awareness," [Online]. Available: http://www.trendmicro.com/cloudcontent/us/pdfs/business/datasheets/wp_appt-survey-report.pdf 8. Ahmad. How Many Internet and #SocialMedia. [Accessed 2013].
- [9] "digital information world," [Online]. Available: <http://www.digitalinformationworld.com/2015/04/infographic-how-many-internets-users-are-fake.html> . [Accessed 2 4 2015].
- [10] I. Inuwa-Dutse, M. Liptrott and I. Korkontzelos, "Detection of spam-posting accounts on Twitter," *Neurocomputing*, 2018.
- [11] S. Rao, A. K. Verma and T. Bhatia, "A review on social spam detection: Challenges, open issues, and future directions," *Expert Systems with Applications - Elsevier*, 2021.
- [12] C. Grier, K. Thomas, V. Paxson and M. Zhang, "@spam: the underground on 140 characters or less," in *CCS '10: Proceedings of the 17th ACM conference on Computer and communications security*, 2010.

- [13] K. Lee, B. Eoff and J. Caverlee, "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [14] G. Stringhini, C. Kruegel and G. Vigna, "Detecting spammers on social networks," in *Proceedings of the 26th Annual Computer Security Applications Confere*, Austin, TX, USA, 2010.
- [15] J. Cho and C. Jeon, "Cost-Based Heterogeneous Learning Framework for Real-Time Spam Detection in Social Networks With Expert Decisions," *IEEE Access*, 2021.
- [16] N. M and J. Prakash, "Detecting Malicious Posts in Social Networks Using Text Analysis," *International Journal of Science and Research (IJSR)*, 2016.
- [17] N. Sun, G. Lin, J. Qiu and P. Rimba, "Near real-time twitter spam detection with machine learning techniques," *International Journal of Computers and Applications*, 2020.
- [18] Z. Alom, B. Carminati and E. Ferrari , "A deep learning model for Twitter spam detection," *Online Social Networks and Media*, 2020.
- [19] R. J. R. Raj, S. Srinivasulu and A. Ashutosh, "A Multi-classifier Framework for Detecting Spam and Fake Spam Messages in Twitter," in *Conference: 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*.
- [20] H. Gupta, M. S. Jamal and S. Madisetty, "A framework for real-time spam detection in Twitter," in *Conference: 2018 10th International Conference on Communication Systems & Networks (COMSNETS)*, Bengaluru, India, 2018.
- [21] M. Mostafa, A. Abdelwahab and H. M. Sayed, "Detecting spam campaign in twitter with semantic similarity," in *Journal of Physics Conference Series*, 2020.
- [22] G. Shetty, A. Nair, P. Vishwanath and A. Stuti, "Sentiment Analysis and Classification on Twitter Spam Account Dataset," in *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, 2020.

- [23] C.-S. Atodiresei, A. Tănăsescu and A. Iftene, "Identifying Fake News and Fake Users on Twitter," in *International Conference on Knowledge Based and Intelligent Information and Engineering*, 2018.
- [24] X. Zheng , Z. Zeng, Z. Chen, Y. Yu and C. Rong, "Detecting spammers on social networks," *Neurocomputing*, 2015.
- [25] C. Comito, "Human Mobility Prediction Through Twitter," *Procedia Computer Science*, 2018.
- [26] M. McCord and M. Chuah , "Spam detection on twitter using traditional classifiers," in *International Conference on Autonomic and Trusted Computing*, 2011.
- [27] Y. Ishida and S. Kuraya, "Fake News and its Credibility Evaluation by Dynamic Relational Networks: A Bottom up Approach," *Procedia Computer Science*, 2018.
- [28] Y. V. Biyani and D. R. A. Khan, "Spam Detection in Social Media using Machine Learning Algorithm," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2020.
- [29] W. H. Hantom and A. Rahman, "Arabic Spam Tweets Classification: A Comprehensive Machine Learning Approach," *MDPI Open Access Journals*, 2024.
- [30] S. Saffarzadeh and M. S. Haghighi, "Identifying Spam Tweets in Social Networks with Combined Approaches of Feature Selection and Deep Learning," in *IEEE Xplore*, 2024.
- [31] F. E. Mendili, M. Fattah, N. Berros, Y. Filaly and Y. E. B. El Idrissi, "Enhancing detection of malicious profiles and spam tweets with an automated honeypot framework powered by deep learning," *SpringerLink*, 2024.
- [32] P. Manasa, A. Malik and I. Batra, "Detection of Twitter Spam Using GLoVe Vocabulary Features, Bidirectional LSTM and Convolution Neural Network," *SpringerLink*, 2024.
- [33] P. Sivakumar, . M. Balasubramani, . R. Sowndharya, B. S. Deepa Priya, W. D.

- Priya and . M. , "Twitter spam drift detection by semi supervised learning approach using YATSI algorithm," *SpringerLink*, 2024.
- [34] M. Errami, M. A. Ouassil, R. Rachidi, M. Jebbari, B. Cherradi and A. Raihani, "Spam Detection in Arabic Tweets Using Artificial Intelligence Techniques," in *International Conference on Circuit, Systems and Communication (ICCSC)*, 2024.
- [35] S. K. Maurya, D. Singh and A. K. Maurya , "Deceptive opinion spam detection approaches: a literature survey," *SpringerLink*, 2023.
- [36] S. B. Abkenar, M. H. Kashani, M. Akbari and E. Mahdipour, "Learning textual features for Twitter spam detection: A systematic literature review," *Expert Systems with Applications-Elsevier*, 2023.
- [37] D. Koggalahewa, Y. Xu and E. Foo , "An unsupervised method for social network spammer detection based on user information interests," *Springer Nature Link*, 2022.