A Survey on Spam Detection Techniques In Social Media Using Machine Learning Algorithms

Ayushi[1], Anshul Kalia[2], Sumesh Sood[3]
[1]Dept of Computer Science
[2]Assistant Professor, Dept of Computer Science
[3]Associate Professor, Dept of Computer Science
[1, 2, 3] Himachal Pradesh University, Shimla. India.

Abstract- *The popularity of internet users, social media platform is growing rapidly among the internet. The social media platform has various advantageous like communicating, information exchanging, knowledge and another concept. However, these social media platforms attract some criminals users. These users disseminate risky information to average users. This information might send some harmful links, misleading emails, or otherwise annoy regular users. So, a major issue in social media sites today is spam detection. To identify spam in social media, machine learning is crucial. In this paper, the various machine learning algorithms to detect the spam of social media have been studied. Many researchers are conducting experiments to detect the spam using machine learning techniques and various datasets are used to detect the spam. The study presents the detailed review and the comparative analysis of existing spam detection techniques. The analysis compares various techniques of spam detection and accuracy of that techniques on different datasets. This comparative analysis is used to decide which technique is more suitable for spam detection.*

Keywords- Social networks, microblogging twitter, spam detection, machine learning.

I. INTRODUCTION

The usage of social networks for online communication and interaction has grown significantly. Users spend the majority of their time reading news, debating events, and publishing messages on well-known social networking platforms (such as WhatsApp, Facebook, Gmail, Twitter, etc.). Unfortunately, this popularity also draws number of spammers who persistently display different behaviour that causes significant misunderstandings[1]. Internet messages that are pointless or unsolicited are known as spams. These are usually sent to a large number of users for variety of use cases such as advertisement, phishing, spreading malware etc. Additionally, spammers frequently go unnoticed on social networking sites by setting up false accounts and stealing legitimate users' accounts for their own benefit. Spammers use different strategies for getting into user's network trust. The problem of spam is not only the annoyance, but also becoming a security threat. In this research the various machine learning techniqueshave been used for spam detection of social media.

Because of the following reasons, spam detection is a major difficulty for service providers:-

- Spam lowers the quality of search results and eliminates revenue for trustworthy websites.
- Spam has a negative economic impact because its increases online traffic because a high ranking results in a lot of free advertising.
- Because there is no cost involved in switching from one search provider to another, it erodes users' trust in search engine providers, which is a particularly noticeable problem.
- Spam websites is a distribution channel for malware, adult content, and fishing attacks.
- Spam makes a search engine provider waste a potent computing and storage resource.
- Finding the best tags for a piece of material while removing the spam tag is a significant challenge in tagging.[2].

II. TYPES OF SPAM

Nowadays various types of spams exist, but some types of spams are discussed below:-

a). Bulk Messaging :-Sending a lot of identical or similar messages to a group of recipients is known as bulk messaging. Sometime these messages are spam messages.

b). Fraudulent Reviews :-False reviews are those written by customers who haven't actually utilised the product, therefore misleading.

c). Malicious links :-using links in posts or tweets that are intended for a specific cruel purpose. This can involve a virus or the theft of user data.

d). Fake accounts :-Spammers often createfake accounts for stealing user's information. These accounts activity is quite different as they post high volume of content.

e). Phishers:-Phishers are people who appear as regular users in order to obtain personal information from other legitimate users.

f). Email spoofing:- In an email spoofing attack, a hacker sends an email that has been altered to appear as though it came from a reliable source.

III. SPAM DETECTION TOOL USED

a). WEKA:- WEKA is open source software that provides tools for data preprocessing, the implementation of various machine learning algorithms, and visualisation tools in order to construct machine learning approaches and utilise them to address real-world issues.

b). WordVector:- The wordvectorprogramme extracts word associations from a huge text corpus using a machine learning model. Once trained, a model like this one can suggest new words to finish a sentence or recognise concepts that are related to existing ones.

c).MapReducer:- Without the use of strict and explicit programming, it can help in the development of systems that learn from data. It is utilised in distributed searching, distributed sorting, and document clustering in addition to machine learning.

IV. LITERATURE REVIEW

This section conducts a survey of several spam detection methods. Based on it, the comparison of various techniques is shown. It gives way to research direction in which further work can be done.

In 2010, Alex Hai Wang [3]a method for detecting spam bots on social networking sites was introduced. The machine learning method is employed in the research to identify the twitter spams. In the research, the spam is detected using both graph-based features and content-based features. To find the spam on Twitter, the author employed a variety of classification techniques. Using the Twitter API, he gathers datasets from web crawlers. The author discovered that the Bayesian classifier performs better overall.

In 2011, De Wang et al. [4]presented a framework to detect the spams on multiple social network. To demonstrate the adaptability and viability of the system, the author employed real-time datasets from social networks. In their research, the author said that the framework provide some feature which are :-

a). The framework quickly detect spams on social network.

b). Accuracy of spam detection is improved with large datasets

c).A new social network can easily be plugged into the system.

In 2012, M. Soiraya et al. [5]presents a social networks spam detection application based on the texts. The application particularly detect the spams of Facebook. The quantity of keywords, average word count, text length, and number of links are among the indicators used to identify spam. The Weka tool is used in the study to develop the data mining model that uses the decision tree (J480).

In 2013, Xia Hu et al. [6]developed a methodology to use machine learning to identify spam on twitter. The authors provide a variety of user-based and content-based features that can be applied to the study to separate Twitter spammers from non-spammers. The authors chose 1,000 randomly chosen Twitter user accounts as their dataset, and the author employed classification techniques such Random Forest (RF), Support Vector Machine (SVM), Naive Bayesian (NB), and K-NN Neighbors. Following a performance analysis, the author discovered that Random Forest (RF) provides greater accuracy. The features reach 95.7% F-measure precision using RF classifier.

In 2014, Zachary Miller et al. [7]detect the spammer of twitter using the data stream clustering method. The study introduces the 95 one-gram features from tweet text for the goal of Twitter spam identification. Two stream-based clustering algorithms, DenStream and StreamKM++, are used in the study by the authors together with a stream of real-time tweets and user profile data. The accuracy rates for the DenStream technique were 97.1%, 84.2%, and 74.8% F-Measure, respectively, while they were 94% and 74.8% for the StreamKM++ approach. From the research, it is analysed that addition of one-gram feature can increase the accuracy of spam detection. Also the combination of these two algorithm with one-gram feature are used in future.

In 2014, Yu Liu et al. [8]provides a hybrid spam detection model for Weibo. The authors used Weibo user statistics and postings from both regular and spam users that were gathered by web crawlers using pre-established rules for their investigation. this hybrid model, also known as SDHM, includes OSN and behavioural features to identify spam. Using a user behaviour model, features, and OSN attribute, SDHM improves F-Measure by 17.95%.

Arushi Gupta and Rishabh Kaushal in 2015 [9]a total of three machine learning algorithms were applied to improve the ability to detect spam on social networking sites. Naive Bayes (NB), clustering, and decision tree (DT) are the three algorithms. A novel integrated technique that incorporates the benefits of the three learning algorithms defined above is given in order to improve the detection of spammers. Their integrated, novel, and combined approach produces better results. With 87.9% accuracy, the suggested algorithm was able to classify an account as spam or not.

In 2016, Xianghan Zheng et al. [10]suggested a supervised machine for spam identification that is based on an extreme learning machine (ELM). In order to conduct the research, firstly generated a labelled dataset by browsing Sina Weibo data and manually categorised the relevant users into spammer and non-spammer groups. A few characteristics are taken from user behaviour and message content and applied to the ELM-based spammer categorization algorithm. The authors approach produces greater dependability and viability. According to the experiment and evaluation, the suggested approach performs better at detecting spammers and non-spammers, with detection rates of 99% and 99.95%, respectively.

In 2016, Saumya Goyal et al. [11]gives new proposal to detect spams on social network by using decision tree and KNN algorithms. The methods were used on actual Twitter datasets to find spam messages. Weka tool is used in the analysis of the proposed mechanism. With the comparison of these two algorithm authors found that the KNN algorithms gives better accuracy then the decision tree algorithm. .

In 2017, Malik Mateen et al. [12]developed a hybrid approach to identifying spammers on the twitter platform that makes use of both content-based and graph-based features. On a genuine Twitter dataset with 11,000 users and more than 400 thousand tweets, the authors analyse the suggested technique. Their tests reveal a 97.6% categorization accuracy rate.

In 2018, Himank Gupta et al [13]provides a methodology for identifying live spam on Twitter. In order to categorise tweets using the tweet text feature, the authors construct a framework that analyses user- and tweet-based attributes. Using the tweet text function has the advantage of allowing users to recognise spam tweets even when the spammer uses a new account. Support Vector Machine, Neural Network, Random Forest, and Gradient Boosting are four different machine learning algorithms that were used to evaluate the response. It is able to attain 91.65% accuracy using neural networks.

In 2018, N. Senthil Murugan and G. Usha Devi [14]Review several machine learning techniques for spam detection on social networking sites. The research focuses on the detection rate and false positive rate of ML algorithms across various datasets. The variation is still present in the study's use of many datasets from other researchers, and random forest is providing high accuracy, or 99.94%.

In 2019, Alok Kumar et al. [15]proposes a method for detecting and removing spam from social networks that is unsupervised, distributed, and decentralised. The authors offer a novel approach that can identify spam from a single message stream and is based on fuzzy logic. The authors used a technology to operate on the MapReduce platform to manage massive amounts of data in networks. The twitter API dataset is utilised in their research to identify spam. This method has an accuracy rate of 94.3%.

In 2020, Zulfikar Alom et al. [16] proposed a new novel approach for spam detection using machine learning. In the research, the authors used two social honeypot dataset. They created a text-based classifier that just takes into account the text of users' tweets and a combined classifier that takes into account both users' tweet text and meta-data. The experiments demonstrate that the suggested technique produces superior outcomes. Their method yields the maximum accuracy for the datasets, at 99.68% and 93.12%, respectively.

In 2021,Poria Pirozmand et al. [17]Growing the SVM based on a combination of the GA and GELS to determine the most potent spam feature gives a new technique for spam detection. The suggested approach consists of two distinct algorithms. The first is a traditional genetic algorithm that is effective in exploring the solution space. The GELS algorithm is incorporated into the suggested strategy to improve the GA. A strong local search method that can increase exploitation efficiency is GELS. In order to reach the ideal answer, these two algorithms working together can be quite effective. This approach yields a greater accuracy of 97.22%.

V. COMPARATIVE AND ANALYSIS

This study shows the comparison between different techniques of machine learning on the basis of problem identified, dataset, tool and technique used. By analysing various machine learning techniques based on their performance the best spam detection is concluded.

Table 1 presents comparative analysis of different machine learning technique for spam detection

Year	Problem Identified	Dataset	Tool	Technique used	Result
2010	Find the best classification algorithm to detect spams in twitter	Twitter API	WEKA	Various classification algorithms	NB show better accuracy
2011	Presents Framework to detect spam in twitter	Real-Time Dataset	WEKA	Machine learning	Framework showsbest accuracy
2011	Build model to detect spam in twitter	Random dataset	WEKA	NB, SVM, RF, K-NN	RF (95.7%)
2012	Detect spams by using some feature	Real-time dataset	Orange	Data Mining model using Decision Tree algorithms	Better accuracy
2013	Built SSDM framework to detect spam in twitter	Real-time	WEKA	Content-based and Network-Based feature	Content based show better accuracy
2013	Present statistical approach to detect spams of facebook and twitter	Real-time	WEKA	NB, DT	DT ((97.6 %)
2014	Find best method to detect spam	Real-time	WEKA	DenStream, StreamKM++	DenStream (97.3%)
2014	Find best algorithm to detect spam	Real-time	WEKA	NB, DT	DT (96.4)
2014	Build hybrid model to detect spam in weibo	Web-crawler	WEKA	SDHM use content-base, Behavior-base and OSN feature	SDHM (17.95 % improvement)
2015	Find best algorithm to detect spam	Twitter-API	WEKA	NB, DT and clustering	Integrated approach show batter accuracy
2015	Propose framework to detect spam	UDI-Twitter	WEKA	Semi-supervised framework	Framework (97.2%)
2016	Create effective model to detect spams	Sinaweibo dataset	WEKA	ELM use content-base and behavior-base feature	ELM (99.9%)
2016	Develop novel technique to detect the spam	UDI-twitter	WEKA	Semi-supervised techniqu e which use K-Mediods and content-base and behavior-base feature	Novel techniq ue (94.72 %)
2016	Propose a framework to detect best method for spam detection	Real-time	WEKA	DT, KNN	DT (94.9%)
2017	Gives novel approach to detect	Real-time	Word Vector	Deep learning	Better accuracy

	spam				
2017	Propose hybrid technique to detect spams	Real-time	WEKA	Classification algorithms which use content-base and graph-base feature	Classification algorithm (97.6%)
2018	Create effective framework to detect spam	Real-time	WEKA	SVM,NN,RF, GB	NN (91.6%)
2018	Find best algorithm to detect spam	Real-time	WEKA	NB, K-NN, SVM, RF	RF (99.94 %)
2019	Presents new method to detect spam	Twitter-API	Map Reducer	Unsupervised technique	94.3%
2020	Develop novel approach to detect spam	Honeypot	WEKA	Text-based and combined feature use	99.6%
2021	Develop new method to detect spam	Real-time	WEKA	SVM, GA, GELS	97.2%
2021	Propose a new method to detect spam	Real-time	WEKA	Machine learning technique which use content-base and user-base feature	98.9%

The above table shows the comparison of various spam detection techniques used by researchers. In this table we mention the problem identified by the researchers, dataset used, tools which is used for implementation, the features and results produced by the various techniques.

VI. CONCLUSION

Spam is the one of the major problem in social networking site. It harms the user devices and also steal the user personal information. It has been observed that selection of algorithm improves the performance of the model. And the accuracy of any model is dependent on the datasets and algorithms. This research show that the WEKA tool is mostly used for implementation. Many researchers use the real time datasets and UDI-Twitter dataset. Some researchers use single algorithms and some use combination of algorithms with use of content-base, link-base and graph-base feature. And compare the result with another algorithms. The research presents a detailed review and comparative study of existing technique used to detect the spams. The comparative analysis of various techniques helps in deciding which approach is best suitable for detection of spam. On comparing the accuracy of all the model, model which is perform best is consider ELM and Random Forest with accuracy 99.9%.

VII. FUTURE SCOPE

In the future, the evaluation of the performance on different machine learning techniques based on the different parameter i.e. Precision, Recall, F-Measure by using different machine learning tools. Deep Learning techniques also detect the spams in social media. Further the more features are used with the combination of machine learning algorithm. So, in future find the best clustering algorithm for detection of spam with real time datasets having low errors and high accuracy.

REFERENCES

[1] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, "Detecting spammers on social networks," *Neurocomputing*, vol. 159, no. 1, pp. 27–34, 2015, doi: 10.1016/j.neucom.2015.02.047.

[2] A. Mathur and P. Gharpure, "Spam Detection Techniques: Issues and Challenges," *Int. J. Appl. Inf. Syst.*, vol. 2013, no. Icwac, pp. 39–41, 2013.

[3] A. H. Wang, "Detecting spam bots in online social networking sites: A machine learning approach," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6166 LNCS, pp. 335–342, 2010, doi: 10.1007/978-3-642-13739-6_25.

[4] D. Wang, D. Irani, and C. Pu, "A social-spam detection framework," *ACM Int. Conf. Proceeding Ser.*, pp. 46–54, 2011, doi: 10.1145/2030376.2030382.

[5] M. Soiraya, S. Thanalerdmongkol, and C. Chantrapornchai, "Using a Data Mining Approach : Spam Detection on Facebook," vol. 58, no. 13, pp. 26–31, 2012.

[6] X. Hu, J. Tang, Y. Zhang, and H. Liu, "Social spammer detection in microblogging," *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 2633–2639, 2013.

[7] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, "Twitter spammer detection using data stream clustering," *Inf. Sci. (Ny).*, vol. 260, no. November, pp. 64–73, 2014, doi: 10.1016/j.ins.2013.11.016.

[8] Y. Liu, B. Wu, B. Wang, and G. Li, "SDHM: A hybrid model for spammer detection in Weibo," *ASONAM 2014 - Proc. 2014 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, no. Asonam, pp. 942–947, 2014, doi: 10.1109/ASONAM.2014.6921699.

[9] A. Gupta and R. Kaushal, "Improving spam detection in Online Social Networks," *Proc. - 2015 Int. Conf. Cogn. Comput. Inf. Process. CCIP 2015*, 2015, doi: 10.1109/CCIP.2015.7100738.

[10] X. Zheng, X. Zhang, Y. Yu, T. Kechadi, and C. Rong, "ELM-based spammer detection in social networks," *J. Supercomput.*, vol. 72, no. 8, pp. 2991–3005, 2016, doi: 10.1007/s11227-015-1437-5.

[11] S. Goyal, R. K. Chauhan, and S. Parveen, "Spam detection using KNN and decision tree mechanism in social network," *2016 4th Int. Conf. Parallel, Distrib. Grid Comput. PDGC 2016*, pp. 522–526, 2016, doi: 10.1109/PDGC.2016.7913250.

[12] M. Mateen and M. Aleem, "A Hybrid Approach for Spam Detection for Twitter," pp. 466–471, 2017.

[13] H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, "A framework for real-time spam detection in Twitter," *2018 10th Int. Conf. Commun. Syst. Networks, COMSNETS 2018*, vol. 2018-Janua, pp. 380–383, 2018, doi: 10.1109/COMSNETS.2018.8328222.

[14] N. S. Murugan and G. U. Devi, "Detecting spams in social networks using ML algorithms – a review," *Int. J. Environ. Waste Manag.*, vol. 21, no. 1, pp. 22–36, 2018, doi: 10.1504/ijewm.2018.091308.

[15] A. Kumar, M. Singh, and A. R. Pais, *Fuzzy string matching algorithm for spam detection in twitter*, vol. 939. Springer Singapore, 2019. doi: 10.1007/978-981-13-7561-3_21.

[16] Z. Alom, B. Carminati, and E. Ferrari, "A deep learning model for Twitter spam detection," *Online Soc. Networks Media*, vol. 18, p. 100079, 2020, doi: 10.1016/j.osnem.2020.100079.

[17] P. Pirozmand, M. Sadeghilalimi, A. A. R. Hosseinabadi, F. Sadeghilalimi, S. Mirkamali, and A. Slowik, "A feature selection approach for spam detection in social networks using gravitational force-based heuristic algorithm," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2021, doi: 10.1007/s12652-021-03385-5.