JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Detection of Social Network Spam based on Improved Machine Learning

Rucha Kibe[1],Archana Deokate[2] ,Pooja Suryawanshi [3],Saloni Sonar[4]

Department of E&TC , SKNCOE ,SKNCOE,SPPU,Pune

[1]Ruchaakibe24@gmail.com , [2]d.archana455@gmail.com ,[3]poojasuryawanshi9595@gmail.com,

[4]sonarsaloni2001@gmail.com

***Abstract*—Social networking websites have become more and more popular recently. Users use them to meet new people and communicate their most recent thoughts and actions to their existing acquaintances. The website among these that is growing the quickest is social media. Due to its popularity, many spammers attempt to flood actual users' accounts with spam messages. This paper considers three social networks, Twitter, Facebook, and Instagram, for experimentation. The classification of the data into spam and non-spam using four machine learning techniques, including SVM, KNN, decision trees, and Random Forest. The results obtained from the experiments show that the proposed approach can accurately detect spam in social networks. Implementing such algorithms could help social network platforms improve user experience by reducing the prevalence of spam and fraudulent activity.**

***Keywords*—spam detection, social media, Machine learning, SVM, KNN, DT, RF**

I. Introduction

Social networks are all about bringing together a large user base or simply people who share their thoughts, information, and multimedia files they want to dedicate to or let others know about [1,2]. They communicate considerably more news, trivia, relevant scientific information, jokes, pictures, etc. Through these texts or videos, people can get to know and learn more about one another. Information can be sent and received from other users through social networks like Facebook, Twitter, LinkedIn, YouTube, and Snapchat.

Social networking sites like Facebook, LinkedIn, and Twitter [3,] users can connect with new people, keep in touch with friends, make business relationships, and much more. The fastest-growing social networking platform overall, according to the research, is Twitter. Social media users can send tweets, which are short messages, to other users via Twitter's microblogging services. Only text and HTTP connections are permitted; each tweet is limited to 140 characters. Friends and co-workers can communicate and remain in touch through tweet exchanges [4]. Micro-blogging platforms drew spammers as well as authorized users. On social media networks like Twitter, spam is getting worse. According to Grier et al., 0.13 percent of spam is posted on Twitter, twice as much as email spam. As the click through rate rises, Twitter becomes a more alluring platform for spammers.

Everyone is generally familiar with social networks and some specific instances that allow users to share anything [2]. The use of these social networks is growing fast daily, and real-time users of these social network websites face an enormous difficulty when utilizing them: intrusions known as spam, which pose a severe threat to their safety. When users accept or click on the messages we refer to as spam, they seek to steal their information. Spam comes in the form of emails, photos, and videos. It is an intrusive message that disrupts and bothers users using social networks like Facebook, Twitter, etc.

This approach presents the machine learning algorithms to classify Twitter, Facebook, and Instagram social platforms into spam and non-spam.

II.Literature Survey

The characteristics were employed by Benevenuto et al. [5] as attributes of the SVM algorithm to classify individuals as either spam or non-spam. To identify one user class from the other, they considered two attribute sets: content attributes and user attributes.

In order to develop a spam classifier that actively filters out both old and new spam, Lee et al. [6] performed a statistical investigation of the previously stated spam profile characteristics. The authors created meta classifiers (Decorate, LogitBoost, etc.) to identify previously unidentified spam based on the profile traits described.

Stringhini et al. [7] .'s initial creation of a collection of honey net accounts (honey profiles) on Twitter led to the discovery of numerous traits that enable authors to recognize spam. The RF model was also used in a Twitter dataset to identify spam.

Wang [8] created innovative content-based and graph-based characteristics to make spam identification easier. A Bayesian classification technique was also used to differentiate between suspicious and everyday activities.

The collective viewpoint was introduced by Chu et al. [9], who concentrated on identifying spam campaigns that employ numerous accounts to propagate spam on Twitter. An automatic categorization system was developed to categorize spam campaigns based on RF and different traits, such as individual tweet/account levels.

III. PROPOSED SYSTEM

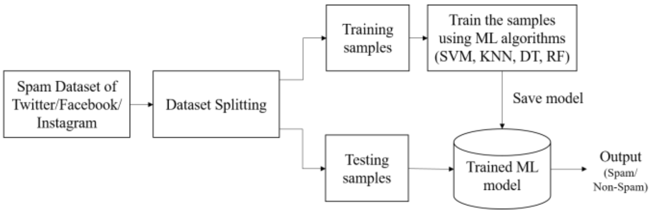

Spam Dataset of Twitter/Facebook/ Instagram
Dataset Splitting
Training samples
Train the samples using ML algorithms (SVM, KNN, DT, RF)
Save model
Testing samples
Trained ML model
Output (Spam/ Non-Spam)

Fig. 1. Block diagram of the proposed system

A. Dataset

This approach uses three social platforms, i.e., Twitter, Facebook, and Instagram. Each dataset is explained below.

a) Twitter spam dataset

The dataset required for evaluating the proposed system is retrieved from Kaggle. [https://www.kaggle.com/c/twitter-spam]. The dataset has the following attributes.

- Tweet: the tweet's text looked like this
- Following: The total amount of followers an account has on Twitter
- Followers: The number of followers for the tweeting account
- Actions: The total number of people who liked, commented on, and retweeted that tweet.
- is retweet: Binary value [0,1]: If 0, a retweet is not made; if 1, a retweet is made.
- Location: The self-described location that the person has listed on their profile may not be accurate, may be "Unknown," and is not standardized! (NY," "New York," "Upper East Side," etc.) as an example!
- Type: Either Quality or Spam

b) Facebook spam dataset

The dataset can be used for building machine learning models. Facebook API and Facebook Graph API are used to collect the dataset, which is collected from public profiles. There are 500 legit profiles and 100 spam profiles. The list of features is as follows with Label (0-legit, 1-spam).

- Number of friends
- Number of followings
- Number of Community
- The age of the user account (in days)
- Total number of posts shared
- Total number of URLs shared
- Total number of photos/videos shared
- Fraction of the posts containing URLs
- Fraction of the posts containing photos/videos
- The average number of comments per post
- The average number of likes per post
- The average number of tags in a post (Rate of tagging)
- The average number of hashtags present in a post

B. Preprocessing

There are two primary responsibilities for the Data Pre-processing module. They are first cleansing the data that the Data Gathering module has retrieved. Finding and fixing corrupt or false data is the process of cleaning data. Create new features for the data set, and then add them. Creating features is the process of changing the form of already-existing features. The Fault Detection module can use the data after these procedures.

By shifting and rescaling values to fall between 0 and 1, normalization is a scaling technique. Additionally called Min-Max scaling. Eq. 1 provides it.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Standardization, which employs a unit standard deviation and centers the data around the mean, is another scaling technique. The distribution that develops has a unit standard deviation and the attribute's mean changes to zero. It is given by

$$X' = \frac{X - \mu}{\sigma} \qquad (2)$$

C. Training using ML Algorithms

For classification issues, the proposed method employs a machine-learning approach. The system classifies the input PV parameters into faulty and normal states using the Support Vector Machine (SVM), K Nearest Neighbour (KNN), decision tree (DT), and Random Forest (RF) algorithms. The section below explains the machine learning algorithm in detail.

a) SVM

A robust machine learning technique called Support Vector Machine (SVM) is utilized for classification, regression, and outlier detection applications. In SVM, the objective is to identify the hyperplane that best separates two classes in a given dataset. The hyperplane is selected to maximize the margin between the two classes. Support vectors, utilized to specify the margin, are the data points closely related to the hyperplane. Both linearly separable and non-linearly separable datasets can be processed using SVM. For datasets that are not linearly separable, SVM uses a kernel function to move the data into a higher-dimensional space where it is linearly separable.

b) KNN

The K-Nearest neighbor (KNN) non-parametric machine learning approach is used for classification and regression issues. Here, the main emphasis will be on the KNN for categorization.

Based on their proximity to the training data points, the KNN algorithm classifies new data points. It does this by figuring out how far the new data point is from every training data point. The new data point is then assigned to the class that appears most frequently among its K closest neighbors after the algorithm chooses the K-nearest data points (the data points with the shortest distances).

An essential KNN algorithm parameter is the value of K. A low number of K may result in overfitting, whereas a high value may result in under-fitting. Hyper parameter tweaking can be used to find the ideal value of K.

KNN's ease of use and interpretability are two benefits. But it can be computationally expensive, and storing the training data in memory takes up a lot of space. KNN can also be sensitive to distance metric selection and may not function well in high-dimensional spaces.

Random forests, however, can be computationally expensive, particularly for big datasets with plenty of input features. The model's intricacy makes it challenging to evaluate the results as well.

IV. RESULTS

This project presents three Twitter, Facebook, and Instagram spam detection datasets. Four machine learning algorithms classify input samples into spam and non-spam. The dataset distribution of Twitter, Facebook, and Instagram spam datasets is shown in Table I. The dataset is spitted into training (80%) and testing (20%).

TABLE I: EVALUATION OF THE ML ALGORITHM ON THE TWITTER SPAM DATASET

Dataset	**No. of Total Samples**	**No. of Training Samples**	**No. of Testing Samples**
Twitter	10000	8000	2000
Facebook	600	480	120
Instagram	696	557	139

The evaluation of the machine learning algorithm on each dataset is discussed below.

A. Twitter Spam dataset

The data visualization of the account age vs. the number of tweets of the Twitter dataset is shown in Fig.2.

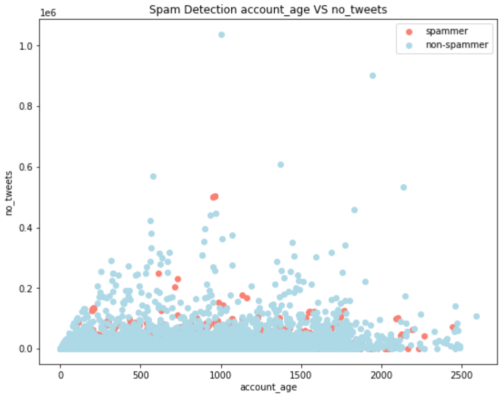


Fig 2: Scatter plot of Twitter Spam Detection account_age VS no_tweets

From the visualization of account age vs. the number of tweets, it is observed that the data is not that separated.

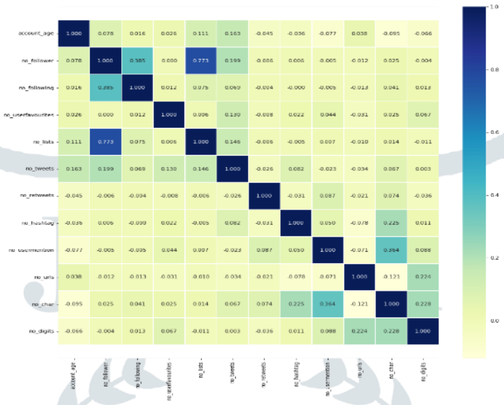


Fig 3:Correlation matrix of Twitter Spam Detection

The correlation matrix shows the relationship between each variable with each other variable. From Fig.3, it is observed that the number of the list has a strong positive correlation with the number of followers.

The performance analysis of different machine learning algorithms on the Twitter spam dataset is tabulated in Table II.

TABLE II: EVALUATION OF THE ML ALGORITHM ON THE TWITTER SPAM DATASET

	Accuracy	**Precision**	**Recall**	**F1 score**
SVM	0.9667	0.966667	0.966667	0.966667
KNN	**0.975**	**0.975673**	**0.975**	**0.975227**
DT	0.8917	0.898698	0.891667	0.894444
RF	0.9667	0.966667	0.966667	0.966667

Table II shows that the KNN performs better than SVM, DT, and RF for classifying the Twitter spam dataset.

B. Facebook Spam dataset

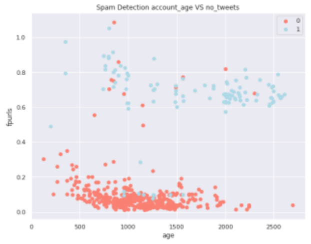


Fig 4: Scatter plot of Facebook Spam Detection friends VS following

From the visualization of Spam Detection friends VS following, it is observed that data is well separated from each other hence these parameters help to increase the classification accuracy.

The correlation matrix of the Twitter spam dataset is shown in Fig 5.5.

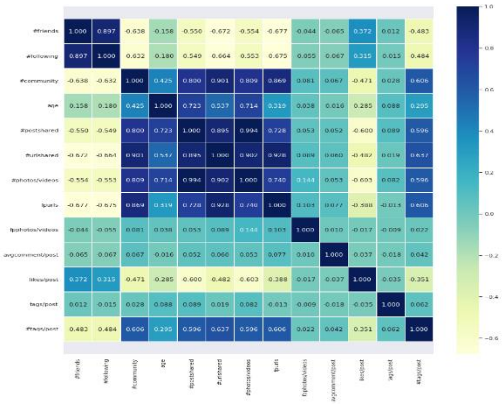


Fig 5:Correlation matrix of Facebook Spam Detection

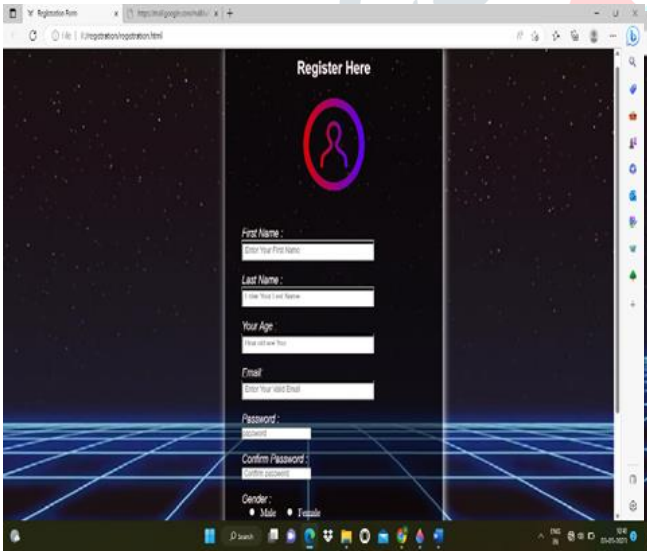
From Fig.5, it is observed that most of the variables have strong positive correlations.

The performance analysis of different machine learning algorithms on the Facebook spam dataset is tabulated in Table III.

TABLE III: EVALUATION OF THE ML ALGORITHM ON THE FACEBOOK SPAM DATASET

	Accuracy	Precision	Recall	F1 score
SVM	0.921429	0.92406	0.921429	0.921569
KNN	0.9	0.900083	0.9	0.899856
DT	0.908333	0.914281	0.908333	0.910546
RF	**0.942857**	**0.94336**	**0.942857**	**0.942916**

Table III shows that the RF classifier performs better than SVM, KNN, and DT for classifying the Facebook spam dataset.


Register Here

First Name :
Enter Your First Name

Last Name :
Enter Your Last Name

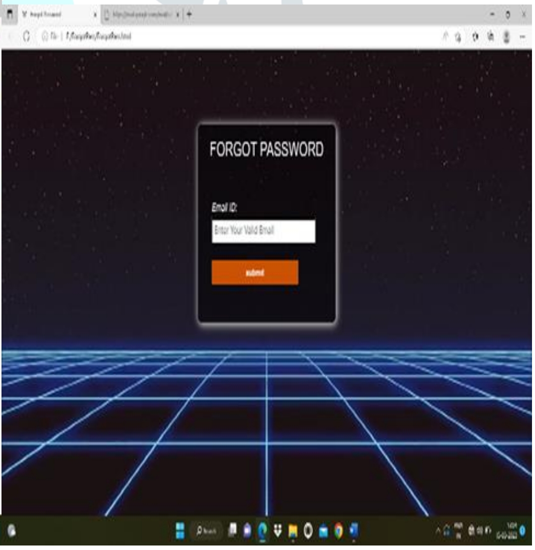
Your Age
What old are You

Email
Enter Your Valid Email

Password :
password

Confirm Password :
Confirm password

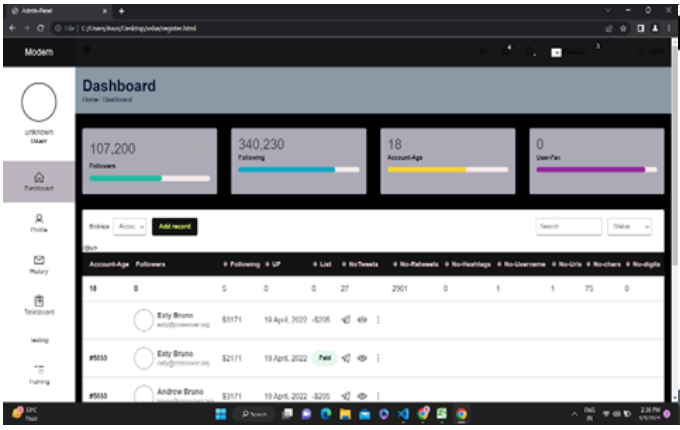
Gender :
● Male ● Female


FORGOT PASSWORD

Email ID:
Enter Your Valid Email

Fig.Registration

Fig.Forgot Password



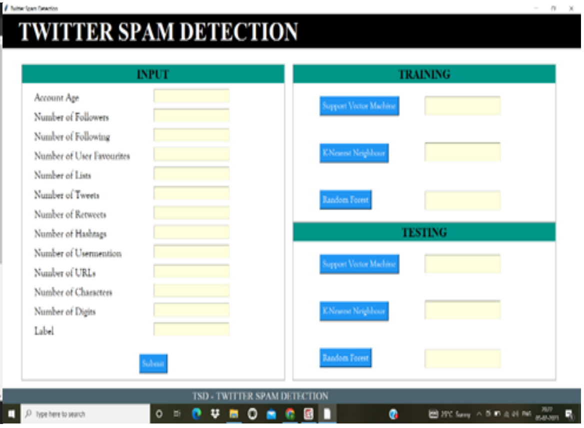
Account-Age	Followers	Following	UF	List	NoTweets	No-Retweets	No-Hashtags	No-Username	No-Urls	No-chars	No-digits
18	0	5	0	0	27	2001	0	1	1	75	0

Exty Bruno $3171 19 April, 2022 -$205

#5033 Exty Bruno $2171 19 April, 2022 Paid

#5033 Andrew Bruno $3171 19 April, 2022 -$205

Fig.Dashboard



Account Age
Number of Followers
Number of Following
Number of User Favourites
Number of Lists
Number of Tweets
Number of Retweets
Number of Hashtags
Number of Usermention
Number of URLs
Number of Characters
Number of Digits
Label

TRAINING: Support Vector Machine, K-Nearest Neighbour, Random Forest

TESTING: Support Vector Machine, K-Nearest Neighbour, Random Forest

Fig .Training & Testing

TABLE IV: EVALUATION OF THE ML ALGORITHM ON THE FACEBOOK SPAM DATASET

	Accuracy	Precision	Recall	F1 score
SVM	0.921429	0.92406	0.921429	0.921569
KNN	0.9	0.900083	0.9	0.899856
DT	0.908333	0.914281	0.908333	0.910546
RF	**0.942857**	**0.94336**	**0.942857**	**0.942916**

Table IV shows that the RF classifier performs better than SVM, KNN, and DT for classifying the Instagram spam dataset.

V. CONCLUSION

Spam detection using ML algorithms has been widely successful and is used in various industries to filter out unwanted and potentially harmful messages. Using machine learning techniques allows for automatically classifying messages as spam or non-spam based on their content, sender, and other features.

In this approach, three social media spam datasets, i.e., Twitter, Facebook, and Instagram, are considered for experimentation. The Four ML algorithms, SVM, KNN, Decision Tree, and Random Forest, are used to train the dataset. The system's performance is evaluated using precision-recall, F-measure, and accuracy parameters. The RF classifier outperforms the SVM, KNN, and DT algorithms for the Twitter spam dataset. RF classifier achieved a precision of 0.8795, recall of 0.8785, F1 score of 0.8787, and accuracy of 0.8785. The KNN classifier outperforms the SVM, DT, and RF algorithms for the Facebook spam dataset. KNN classifier achieved a precision of 0.9756, recall of 0.975, F1 score of 0.9752, and accuracy of 0.975. The RF classifier outperforms the SVM, KNN, and DT algorithms for the Instagram spam dataset. RF classifier achieved a precision of 0.94336, recall of 0.9428, F1 score of 0.9429, and accuracy of 0.9428.

REFERENCES

[1] Zhu, K., Zhi, W. and Zhang, L. (2016) “*Exploring mobile users and their effects in online social networks: a Twitter case study”*,IEEE 24th International Conference on Network Protocols (ICNP), DOI: 10.1109/icnp.2016.7785319.

[2] Xu, H., Sun, W. and Javaid, A. (2016) “*Efficient spam detection across online social networks*”, IEEE International Conference on Big Data Analysis (ICBDA), DOI: 10.1109/icbda.2016.7509829.

[3] Jain, G., Sharma, M., & Agarwal, B. (2018). “*Spam detection on social media using semantic convolutional neural network”.* International Journal of Knowledge Discovery in Bioinformatics (IJKDB), 8(1), 12-26

[4] Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010, October). @ “spam: the underground on 140 characters or less. In Proceedings of the 17th ACM conference on Computer and communications security (pp. 27-37). ACM

[5] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "*Detecting spammers on Twitter*," in Proc. CEAS, vol. 6, 2010, p. 12.

[6] K. Lee, J. Caverlee, and S. Webb, *"Uncovering social spammers: Social honeypots + machine learning,*" in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2010, pp. 435442.

[7] G. Stringhini, C. Kruegel, and G. Vigna, *"Detecting spammers on social networks,*" in Proc. 26th Annu. Comput. Secure. Appl. Conf. (ACSAC), 2010, pp. 19.

[8] A. H. Wang, ``*Don't follow me: Spam detection in Twitter*," in Proc. Int. Conf. Secure. Cryptogr. (SECRYPT), Jul. 2010, pp. 110.

[9] Z. Chu, I. Widjaja, and H. Wang, ``*Detecting social spam campaigns on Twitter,*" in Proc. Int. Conf. Appl. Cryptogr. Netw. Secure. Cham, Switzerland: Springer, 2012, pp. 455472.