International Journal of Engineering & Technology, 7 (1.3) (2018) 61-65

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Comparison of twitter spam detection using various machine learning algorithms

M. Sangeetha [1]*, S. Nithyanantham [2], M. Jayanthi [3]

[1] *Assistant Professor/CSE, Kongu Engineering College, Perundurai, Erode.*
[2] *Assistant Professor/CSE, Karpagam College of Engineering, Coimbatore.*
[3]*Student/CSE, Kongu Engineering College, Perundurai, Erode.*
**Corresponding author E-mail: sangeetham@kongu.ac.in*

Abstract

Online Social Networks(OSNs) have mutual themes such as information sharing, person-to-person interaction and creation of shared and collaborative content. Lots of micro blogging websites available like Twitter, Instagram, Tumblr. A standout amongst the most prominent online networking stages is Twitter. It has 313 million months to month dynamic clients which post of 500 million tweets for each day. Twitter allows users to send short text based messages with up to 140-character letters called "tweets". Enlisted clients can read and post tweets however the individuals who are unregistered can just read them. Due to the reputation it attracts the consideration of spammers for their vindictive points, for example, phishing true blue clients or spreading malevolent programming and promotes through URLs shared inside tweets, forcefully take after/unfollow valid clients and commandeer drifting subjects to draw in their consideration, proliferating obscenity. Twitter Spam has become a critical problem nowadays. By looking at the execution of an extensive variety of standard machine learning calculations, fundamentally expecting to distinguish the acceptable location execution in light of a lot of information by utilizing account-based and tweet content-based highlights.

Keywords: *Twitter; Spammer; tweet; machine learning algorithm; account; tweet content –based.*

1. Introduction

Online social networks (OSNs), for example, Twitter, Facebook and LinkedIn, have had huge effect on our life and have reshaped the best approach to mingle and impart. Individuals can connect with our loved ones anyplace, whenever. Take Twitter for instance, Users can post messages with pictures, recordings, message and tail others whom they are occupied with and nurture. Up until now, Twitter has increased enormous ubiquity and have had up to 313 million dynamic clients [11]. Nonetheless, with the expanding number of clients on Twitter, the spamming exercises are developing also. Twitter spams normally allude to tweets containing promotions, drugs deals or messages diverting clients to outer malignant connections including phishing or malware downloads [1]. Spams on Twitter influence the online social experience, as well as debilitates the security of the internet.

Twitter has connected standards to control clients practices, for example, confining clients from sending copy substance, from specifying different clients over and again or from posting URL-just substance. Then, the spamming issues have pulled in the consideration of the examination group. Specialists have put numerous endeavors to enhance Twitter spam discovery proficiency and precision by proposing

different novel approaches [1], [10], [12], [21].

There are three noteworthy kinds of highlight related answers for Twitter spam identification. The first type is depends on the features of user account and second type is depends on tweets content (such as account age, the number of followers/ followings and the number of URLs contained in the tweet, etc.) [2]. These highlights can be straightforwardly separated from tweets with close to nothing or without calculation. In view of the watched certainties that the content-based features highlights could be manufactured effectively. Different analysts proposed to utilize strong highlights got from the social graph, which is the second sort of arrangement [3]. By utilizing directed graph model to investigate the relationship of senders and recipients [4]. By and by, graph based highlights are observationally hard to gather, in light of the fact that producing an expansive social/relationship graph can be time and asset expending thinking about that as a client may connect with a substantial yet erratic number of clients. The third kind of arrangement concentrates on tweets with URLs. As indicated by areas and IP boycotts are utilized to channel tweets containing noxious URLs [22].

In any case, there is an absence of relative work benchmarking the execution of machine learning algorithms on Twitter spam recognition to show the relevance and plausibility of these machine learning approaches in reality situations. It cross over any barrier by directing an observational investigation of 6 normally utilized machine learning algorithms to assess the location execution as far as discovery exactness, the genuine/false positive rate (TPR/FPR) and the f-measure. The proposed framework recreated the reasonable like condition by utilizing the uneven proportion of spam and non-spam datasets for execution assessment of the chose calculations.

The commitments of the proposed framework are the accompanying: By using account and tweet content based features the performance is compared for various machine learning algorithms[12]. By tests, found that two decision tree based calculations Random

Forest and C5.0 accomplished the best execution in different conditions [13], [23].

1.1. Types of Spammers

Spammers are the cruel users who infect the information accessible by valid users and thus represent a hazard to the security and mystery of interpersonal organizations. Spammers have a place with one of the accompanying classifications [24].

1. Phishers: The clients who carry on like a typical client to get individual information of other certified clients.
2. Fake Users: The users who imitate the profiles of authenticate users to send spam content to the friend's of that user or other users in the network.
3. Promoters: The user who send spiteful links of advertisements or other promotional links to others in order to obtain ones personal information.

1.2 Twitter Spam Detection

1. Account-based spam detection
2. Tweet-based spam detection
3. Graph-based spam detection
4. Hybrid spam detection

1.2.1 Account-based spam detection

Account-based spam detection methods are based on the features (or combination of them) of twitter account such as username, Biography, Profile photo, Header photo, Theme color, Birth Date, Homepage, Location, Creation date, Number of tweets, Number of following, Number of followers, Number of likes, Number of retweets, Number of lists, Number of moments.

1.2.2 Tweet-based spam detection methods

Tweet-based spam detection methods are based on the features of sender, Mentions, Hash-tags, Link, Number of Likes, Number of retweets, Number of replies, sent date, Location. The conventional approaches to channel spam depend on IP boycotting, domain and URL boycotting. Since spammers tend to utilize abbreviated URLs, conventional URL or IP boycotting techniques are not ready to channel malevolent URLs in Twitter.

1.2.3 Graph-based spam detection methods

Graph based spam detection techniques utilize graph information structures to display highlights of Twitter as nodes and edges. Graph-based spam detection methods are based on the highlights of Distance and Connectivity. The separation between a spammer and a genuine client is more remote than the separation between two true blue clients.

The connectivity and distance to break down how these records are associated each other and to gauge qualities of their associations so as to uncover the likelihood of a spam association. Graph-based features provide the most robust performance to detect spam and spammers since they are difficult to control and not client controlled.

1.2.4 Hybrid spam detection methods

Hybrid spam detection methods utilize a mix of spam discovery strategies keeping in mind the end goal to give more vigorous spam detection.

2. Proposed System

The Spams in twitter are detected with Data Sets collected and analysed by using various Machine Learning algorithms for its performance, stability and scalability. The data set has two features: Account - based features like account_age, no_follower, no_following, no_userfavourites, no_lists, no_tweets and Tweet content-based features like no_retweets,no_hashing, no_usermention, no_url, no_char, no_digits. The Machine Learning has various algorithms; here the algorithms used are KNN, k-kNN, Random Forest, c5.0, Stochastic GBM, Naïve Bayes.

The selected algorithms are widely used both in modern and scholarly fields. kNN-based algorithms, is picked, because of their appropriateness for information tests with moderately modest number of measurements. It presents weighting mechanism for the closest neighbors in view of their similitude to an unclassified example. The likelihood of a recently watched test having a place with a class is impacted or weighted by its likeness to the examples in the training set in Figure - a.

The Random Forest and C5.0 are selected as representatives of the decision tree-based algorithms. The Naive Bayes, classic probabilistic classifier which expands on suspicion that all highlights of information autonomous. The data set collected from twitter are analyzed with the above selected ML algorithms, also computed the performance, stability and scability of spammers and non-spammers. The graph is also generated by the accuracy, TPR, FPR and F-Measure.

System Architecture

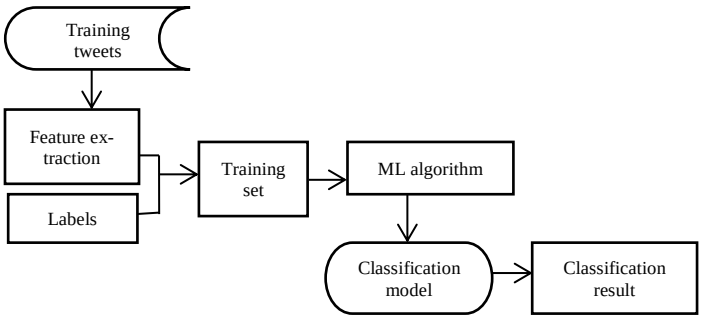

Figure - a

2.1. KNN-Based

2.1.1 KNN

KNN is picked because of their appropriateness for information tests with generally modest number of measurements. KNN algorithm is robust to noisy training data. All occurrences compare to focuses in a n-dimensional Euclidean space. Grouping is postponed till another occurrence arrives. Arrangement done by looking at highlight vectors of the diverse focuses [19].

2.1.2 WEIGHTED KNN(K-KNN)

Aside from kNN, likewise chose k-kNN which is an enhanced kNN algorithm. It presents a weighting mechanism for estimating the closest neighbors in light of their comparability to an unclassified sample [25].

2.2 Decision Tree-Based

2.2.1 Random Forest

Random forest (or random forests) is an joint classifier that consists of many decision trees and outputs. one of the most accurate learning algorithms available today is random forest algorithm. It produces a highly accurate classifier for data sets. It runs efficiently on large databases. Large number of decision trees are created using

random forest approach. Each observation is nourished into each choice tree. The regular result for every perception is utilized as the last yield. Another observation which is taken is encouraged into every one of the trees and taking a greater part vote in favour of every grouping model. A mistake assessment is made for those cases which were not utilized while building the tree. That is called an OOB (Out-of-bag) blunder appraise which is said as far as rate. To create random forest tree in R, "Random Forest" package is needed. [13].

2.2.2 C5.0

C5.0 gives more accurate and efficient result compared to any other classifiers. By using the C5.0 as the base classifier, the proposed framework will sort out the outcome set with high exactness and low memory use. The grouping procedure produces less guidelines contrast with all techniques. Accuracy in the result set is high due to low error rate. Due to the construction of pruned tree, the system generates fast results compared to other techniques [23].

2.3 Boosting Algorithm

Gradient boosting algorithm is mainly used to obtain regression and classification problems, it produces a expectation model in the method of an collaborative of weak calculation models. It forms the model in a stage-wise fashion like other boosting models [18].

2.4 Naive Bayes

Naive Bayes Algorithm is a quick, profoundly versatile calculation. Naive Bayes can be use for Binary and Multiclass grouping. It provides different types of Naive Bayes Algorithms like Gaussian NB, Multinomial NB, Bernoulli NB. Naive Bayes classifier accepts that the occurrence of a exact feature in a class is unlike to the presence of any other feature. Naive Bayes works fine with high dimensions [14].

3. Experimental Results

3.1 Performance Metrics

The measure of performance is to use the accuracy, the true positive rate (TPR), the false positive rate (FPR) and the F-measure as metrics. The accuracy is the level of effectively recognized cases (the two spams and non-spams) in the aggregate number of inspected cases, which can be figured using equation (1). The TPR demonstrates the proportion of effectively distinguished spams to the aggregate number of real spams. It can be computed using equation (2). The FPR refers to the amount of non-spams incorrectly classified as spams in the total number of actual non-spams, as equation (3) shows. The precision is defined as the ratio of correctly classified spams to the total number of tweets that are classified as spams, as shown by the equation (4). Lastly, the F-measure F1 score or F-score, is another extent of prediction accuracy combining both the precision and recall. F-measure can be

calculated by the equation (5).

$$\text{ACCURACY} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{TPR} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{FPR} = \frac{FP}{FP+TN} \quad (3)$$

$$\text{PRECISION} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{F-MEASURE} = 2 \cdot \frac{\text{PRECISION.RECALL}}{\text{PRECISION+RECALL}} \quad (5)$$

3.2 Accuracy

Table 1: Comparison of classifiers on accuracy values for dataset 1,2 and 3

	ACCURACY					
DATASET	KNN	K-KNN	RF	c50	GBM	NB
D1	75.44	83.14	90.95	87.44	89.44	58.18
D2	69.03	78.14	90.05	84.54	85.84	53.18
D3	71.14	79.54	89.59	83.79	85.94	54.79

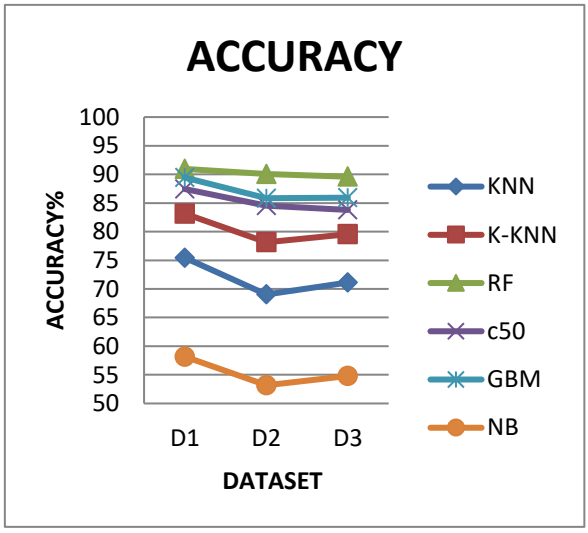

Fig. 1: Detection accuracy (%) of algorithms using dataset 1, 2 and 3

From the figure 1 random forest achieved highest accuracy observed among all. c5.0 and GBM achieves more than 85% of correctness.

3.3 True postive rate

Comparison of True Positive Rate (TPR) for different dataset using Machine Learning algorithms.

Table 2: Comparison of classifiers on True Positive Rate (TPR) for dataset 1 VS 4

DATASET	KNN	K-KNN	RF	c50	GBM	NB
D1	52.98	48.85	50.66	49.82	50.89	17.02
D4	48.51	47.08	50.65	50.75	50.11	14.92

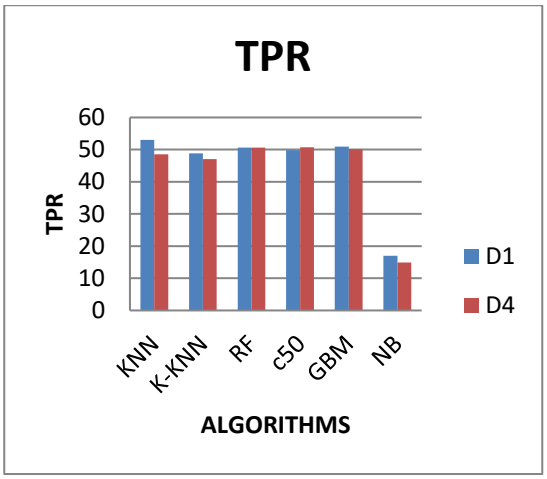

Fig. 2: TPR values on dataset 1 VS 4

Figure 2 shows the comparison of True Positive Rate (TPR) for dataset D1 and D4 using Machine Learning algorithms.

3.4 False Positive Rate

Comparison of False Positive Rate (FPR) for different dataset using Machine Learning algorithms

Table 3: Comparison of classifiers on FPR values for dataset 1 VS 4

DA-TASET	KNN	K-KNN	RF	c50	GBM	NB
D1	38.28	48.07	36.46	46.21	36.49	94.37
D4	51.84	41.13	35.32	40.65	43.93	93.81

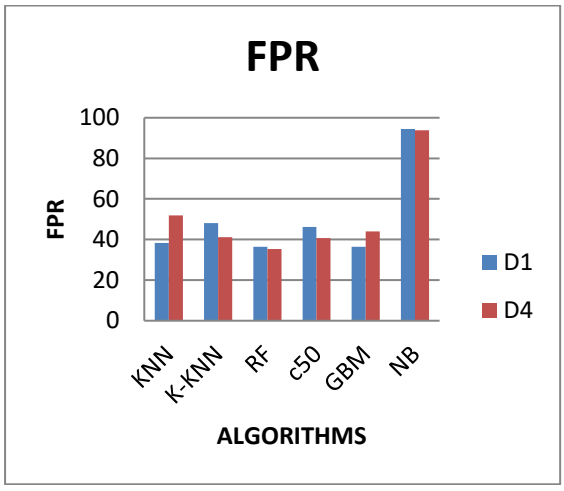

Fig. 3: FPR values on dataset 1 VS 4

Figure 3. shows the comparison of False Positive Rate (FPR) for dataset D1 and D4 using Machine Learning algorithms

3.5 F-Measure

Table 4: Comparison of classifiers on F-Measure values for dataset 1 and 4

DA-TASET	KNN	K-KNN	RF	c50	GBM	NB
D1	64.04	61.60	65.66	63.69	65.58	18.41
D4	58.27	59.96	65.83	64.57	64.21	15.90

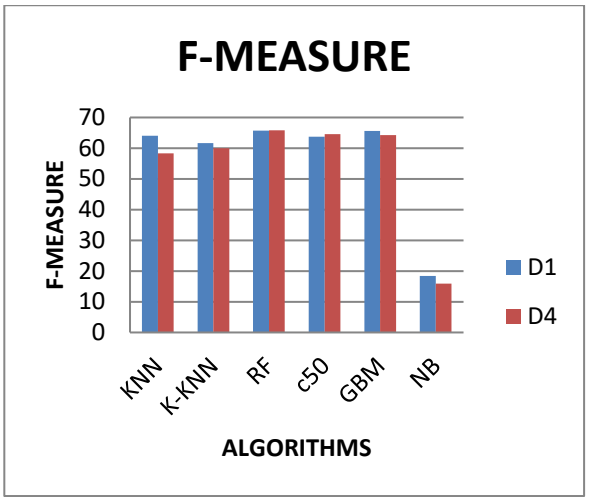

Fig. 4: F-measure Values on Dataset 1 VS 4

Figure 4 shows comparison of F-Measure for dataset D1 and D4 using Machine Learning algorithms

4. Conclusion

For different Machine Learning algorithms, the performance of detecting Twitter spams in terms of accuracy, the TPR/FPR and the F-measure has been calculated. The outcome of our experiment shows that Random Forest and C5.0 has get the high detection accuracy, and Random Forest performed more stable than other algorithms. As a future work the performance of the algorithms can be improved by training more tweets.

References

[1] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting spammers on Twitter,” in Proc. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf. (CEAS), vol. 6. 2010, p. 12.

[2] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, “6 million spam tweets: A large ground truth for timely Twitter spam detection,” in Proc.

[3] C. Yang, R. Harkreader, and G. Gu, “Empirical evaluation and new design for fighting evolving Twitter spammers,” IEEE Trans. Inf. Forensics Security, vol. 8, no. 8, pp. 12801293, Aug. 2013.

[4] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, “Design and evaluation of a real-time URL spam filtering service,” in Proc.

[5] Cran R-Project, R Project Website. (Aug.6, 2015). A Short Introduction to the Caret Package.

[6] M. Kuhn, “Caret package,” J. Statist. Softw., vol. 28, no. 5, pp. 1_26, 2008.

[7] Z. Chu, S. Gianvecchio, H. Wang, S. Jajodia, Who is Tweeting on Twitter: Human, Bot, or Cyborg?, in: 26th Annu. Comput. Secur. Appl. Conf. (ACSAC 2010), Austin, Texas, USA, 2010: pp. 21–30. doi:10.1145/1920261.1920265.

[8] P. Kaur, A. Singhal, J. Kaur, Spam Detection on Twitter: A Survey, in: 2016 Int. Conf. Comput. Sustain. Glob. Dev., IEEE, New Delhi, India, 2016: pp. 2570–2573.

[9] C.D. Gowri, V. Mohanraj, A Survey on Spam Detection in Twitter: A Review, Int. J. Comput. Sci. Bus. Informatics. 14 (2014) 92–102.

[10] J. Song, S. Lee, and J. Kim,“Spam filtering in Twitter using sender receiver relationship,” in Proc. Int.Workshop Recent Adv. Intrusion Detection, 2011, pp. 301317.

[11] Statista. Number of Monthly Active Twitter Users Worldwide from 1st Quarter 2010 to 2nd Quarter 2016 (in millions), accessed on Aug. 9, 2016.

[12] G.Stringhini, C. Kruegel, and G. Vigna, “Detecting spammers on social networks,” in Proc. 26th Annu. Comput. Secur. Appl. Conf., 2010,pp. 1-9.IEEE Int. Conf. Commun. (ICC), Jun. 2015, pp. 70657070.

[13] G. Biau, “Analysis of a random forests model,” J. Mach. Learn. Res.,vol. 13, pp. 1063_1095, Apr. 2012.
[14] C.M.Bishop,“Pattern recognition and machine learning,” New York,NY,USA: Springer, 2006.
[15] D. Conway and J. White, Machine Learning for Hackers. Newton, MA,USA: O'Reilly Media, 2012.
[16] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, “COMPA: Detecting compromised accounts on social networks,” in Proc. NDSS, 2013.
[17] J.Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: A statistical view of boosting,” Ann. Statist., vol. 28, no. 2, p. 2000, 1998.
[18] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” Ann. Statist., vol. 29, no. 5, pp. 1189_1232, 2001.
[19] K. Ghosh, P. Chaudhuri, and C. A. Murthy, “On visualization and aggregation of nearest neighbor classifiers,” IEEE Trans. Pattern Anal.Mach. Intell., vol. 27, no. 10, pp. 1592_1602, Oct. 2005.
[20] K. Hechenbichler and K. Schliep, “Weighted K-nearest-neighbor techniques and ordinal classifcation,”Ludwigs_Maximilias Univ. Munich,Munich, Germany, Discussion Paper 399, SFB 386, 2004, p. 16
[21] H. Wang, “Don't follow me: Spam detection in Twitter,” in Proc. Int. Conf. Secur. Cryptogr. (SECRYPT), 2010, pp. 1_10.
[22] D. Wang, S. B. Navathe, L. Liu, D. Irani, A. Tamersoy, and C. Pu, “Click traffic analysis of short URL spam on Twitter”.
[23] J. R. Quinlan. Data mining tools See5 and C5.0, accessed on Jun. 10, 2017.[Online]. Available: http://www.rulequest.com/see5-info.html
[24] Abdullah Talha Kabakus , Resul Kara,”A Survey of Spam Detection Methods on Twitte”.
[25] K. Hechenbichler and K. Schliep, “Weighted K-nearest-neighbor techniques and ordinal classification,” LudwigsMaximilians Univ. Munich, Munich, Germany, Discussion Paper 399, SFB 386, 2004, p. 16