DISCOVER OF MALICIOUS CONTENT AND TEXT IN SOCIAL MEDIA

V Subbulakshmi,
Assistant Professor,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Ujjwal Anand,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Rajesh kannan T,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Rajesh N,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Abstract: Twitter slants, an opportune refreshed arrangement of best terms in Twitter, can influence people in general plan of the group and have pulled in much consideration. Lamentably, in the wrong hands, Twitter patterns can likewise be mishandled to deceive individuals. In this paper, we endeavour to explore whether Twitter patterns are secure from the control of malignant clients. We gather in excess of 69 million tweets from 5 million records. Utilizing the gathered tweets, we first direct an information investigation and find confirmation of Twitter incline control. At that point, we learn at the subject level and deduce the key factors that can decide if a theme begins drifting because of its fame, scope, transmission, potential scope, or What we find is that aside from transmission, all of elements above are firmly identified with drifting. At long last, we additionally explore the drifting control from the point of view of bargained and phony records and talk about counter measures. We used an algorithm such as naive Bayesian algorithm is used for machine learning without any explicitly programmed .Behavioural algorithm, finding the behaviour of the person and their emotions. Linguistic algorithm focuses on problem of processing efficiency and ambiguity.

Keywords: Twitter trend, Security.

I. INTRODUCTION

Twitter is an online social networking and micro blogging service that enables users to send and read 140-character messages. Registered users can read and post tweets, interact with other users, and all other features of Twitter. -Unregistered have limited access and can only read them. The Internet has subverted the autocratic way of disseminating news by traditional media like newspapers. Online trends are different from traditional media as a method for information propagation. For instance, Google Hot Trends ranks the hottest searches that have recently experienced a sudden surge in popularity [2]. Meanwhile, these trends may attract much more attention than before due to their appearance on Google Hot Trends. We examine Twitter slanting at theme level, thinking about points' fame, scope, transmission, potential scope, and notoriety. The comparing dynamics for each factor above are separated, and after that Help Vector Machine (SVM) classifier is utilized to check how precisely a factor could foresee drifting. Previous research have studied trend taxonomy [10], trend detection [14], and real events extraction from Twitter trends [6].It is reported that attackers manipulate Google trends by simply employing large group of people to visit Google and search for a specific keyword phrase [15].

We examine that what the user can do and don't in twitter. The instructions to be followed, Use a strong password that you don't reuse on other websites and Use login verification. Be cautious of suspicious links and always make sure you're on twitter.com before you enter your login information. Do create a password at least 10 characters long. Longer is better. Do use a mix of uppercase, lowercase, numbers, and symbols. Do not use personal information in your password such as phone numbers, birthdays, etc. Make sure your computer software, including your browser, is up-to-date with the most recent upgrades and anti-virus software.

II.RELATED WORK

We present the threat of malicious manipulation of Twitter trending, given compromised and fake accounts in the suspect spamming infrastructure we observed. Then we demonstrate how compromised and fake accounts could threaten Twitter trending by simulating the manipulation of dynamics as compromised and fake accounts would do. Corresponding countermeasures are then discussed. If the system encounters any vulnerable context within the tweet, it will encrypt with special characters. Machine generated hash tags and tweets can be identified. It discusses the manipulation of Twitter trends and the corresponding counter measures. Fake activity, often bought for publicity purpose, Influence trending topics[1].Track political memes in twitter and detect astroturfing, smear campaigns & other misinformation in context of U.S political elections.[5].Analysing Twitter messages to distinguish between those covering real world events and non-event messages[6].Classifying and categorizing the topics based on i) network based classification and ii) text classification.[10]

We describe the factors and the corresponding dynamics as follows:

Popularity and Tweet Dynamics. The ubiquity of a theme speaks to the subject's essentialness. We utilize the tweet number of a point to catch the theme's notoriety. The tweet progression of a subject record the varieties of the quantity of tweets about the theme. It is the most every now and again utilized metric for estimating the advancement of occasions and identifying trending themes.

Coverage and Account Dynamics. Scope of a theme implies the interest rate of the subject. We can utilize the record number of a point to evaluate its scope. Record progression mirror the varieties of the quantity of records engaged with the subject. Contrasted and tweet flow, account elements reject the effect of those to a great degree dynamic records in the trending.

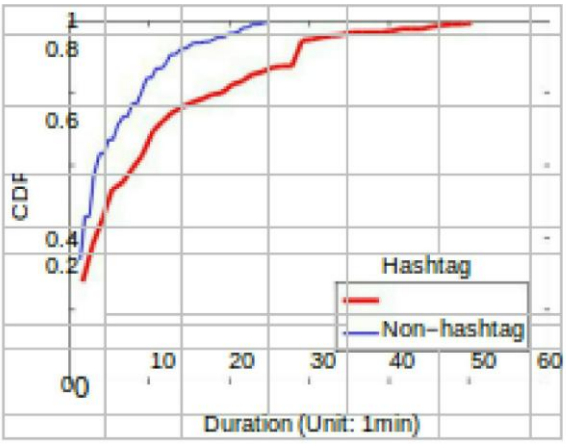

Figure.1 Topics showing the trending durations.

Manipulation of Trends

Manipulating Trends, Spammers in Twitter conduct malicious activities mainly through compromised and fake accounts.

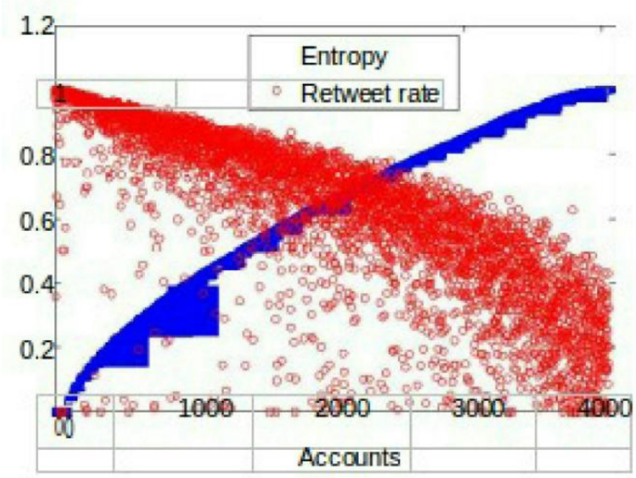

Figure.2 Rate of the retweeted accounts and Entropy.

In this section, we first evidence the involvement of compromised and fake accounts in the manipulation of trends, and then we simulate the manipulation of dynamics as compromised and fake accounts would do. Finally, we discuss the possible countermeasures against the manipulation of trends.
Entropy is used for sorting the element in the ascending order and the retweet rate of a person is show in the fig2.Retweet rate of a person will be shown in the red scattered way on the Fig2.

Fake Accounts

As per a current report, a huge number of phony records on Twitter are controlled by bot-experts. They are sold and purchased through a secret market. To confirm the presence of phony records in the manipulation of Twitter slanting, we ponder the conduct profile of those records that show up in the spike, in which the confirmation of drifting control is found. There are of aggregate 4,055 records (5,193 tweets) in the spike. Utilizing the web creeping technique, we extricate a gathering of tweets posted by each record and the related data (e.g., supporter number) for each record. There are by and large 180 tweets for one record, and the tweet histories most recent 334 days all things considered.

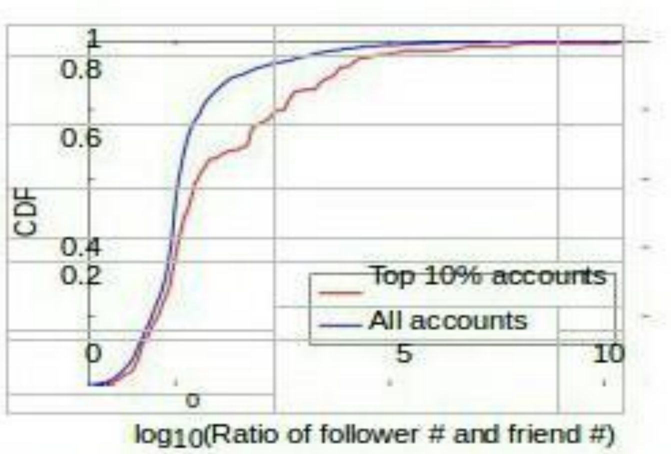

Figure.3 Shows the entropy (climbing request) of the records

We initially investigate the entropy of time interims between posting tweets for each record. The entropy of time interims between posting tweets of a record can demonstrate the consistency of the record's posting conduct. When all is said in done, the littler entropy esteem a record has, the more probable it is a bot.
In Fig.4 we analyse the reports of all tweets, we estimated trends between original and manipulated dynamics. This diagram is showing the predicted report of all the trends which have been manipulated and which is still original. The manipulated dynamics shows the difference in sum of two peaks and sum of three peaks.

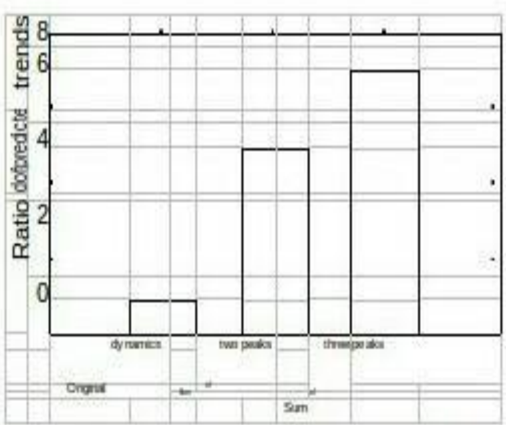

Figure.4 The ratio of estimated trends between the original dynamics and manipulated dynamics (sum of two peaks and sum of three peaks).

III.PROPOSED WORK

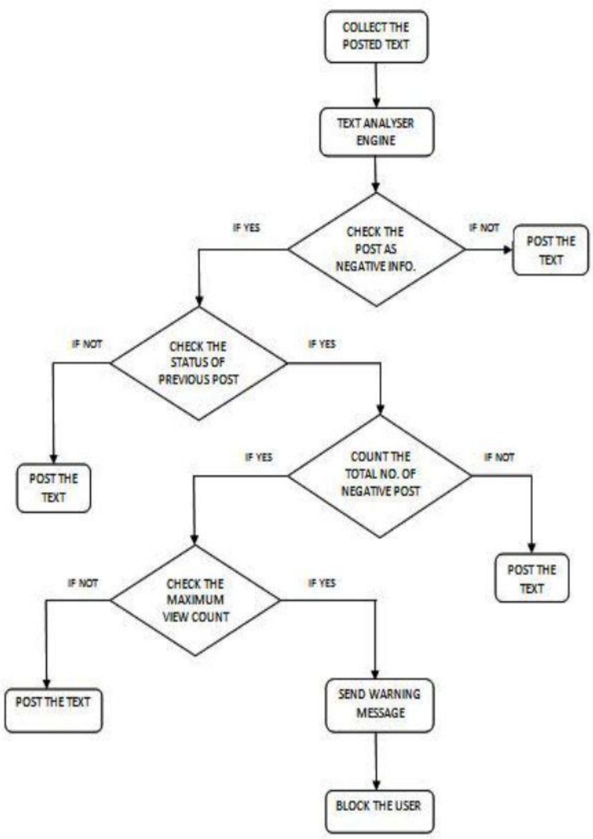

Figure5: Architecture flow diagram

To the best of our knowledge, this is the first effort to investigate whether Twitter trends could be manipulated. Research on trending topics in Twitter includes real event recognization, realtime trending topic detection, the evolution of trending topic characterization, and the taxonomy of trending topics. Becker et al. analyzed the stream of Twitter messages and distinguished the messages about real events from non-event messages based on a clustering method. Zubiaga et al. categorized different triggers that leverage the trending topics by using social features rather than content-based approaches.

In the Trend manipulation, the overall process of the manipulation will be in the architecture is shown below

In the detection of real-time trending topics. Kasi viswanathan et al. presented a dictionary learning- based framework for detecting emerging topics in social media via the user-generated stream. Lu et al. used an energy function to model the life activity of news events on Twitter and proposed a news event detection method based on online energy function. Lehmann et al. classified the popular hash tags by the temporal dynamics of hash tags. Irani et al. focused on the trend stuffing issue and developed a classifier to automatically identify the trend-stuffing in tweets. Whether a topic begins trending is closely related to (1) the influence of users who are involved with the topic and (2) the topic adoption for users who are exposed to the topic. Cha et al. performed a comparison of three different measures of influence: in-degree, retweet, and mention. Weng et al. proposed a topic-sensitive Page Rank measure for user influence. Romero et al. proposed an algorithm to measure the relative influence and passivity of each user from the viewpoint of a whole network. Bakshy et al. measured the influence from the diffusion tree. The studies of topic adoption in Twitter mainly concentrate on hash tag adoption. Lin et al. classified the adoption of hash tags into two classes and proposed a framework to capture the dynamics of hash tags based on their topicality, interactivity, diversity, and prominence. Yang et al. studied the effect of the dual role of a hash tag on hash tag adoption.

IV.IMPLEMENTATION RESULTS

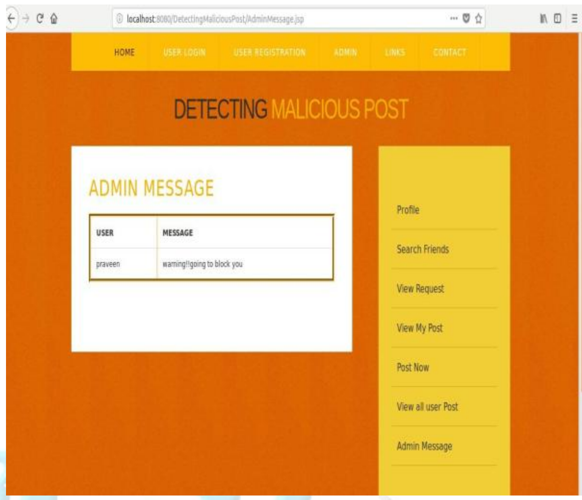

Figure6: Warning Message from Admin

In Implementation results, by the use of two algorithms we discover that the user is done some Malicious activity. So admin sent the Warning message to the user. It shown in the fig.6

The malicious text and content posted person will be seen in the admin login. Admin is the only person will see this type of activities in social media. This show clearly in the fig.7.

If the person/user is continuously doing the malicious activity and he will be Blocked permanently for the social media platform. How many post the particular person is posted related to malignant words. In the fig.8 shown the Blocking function.

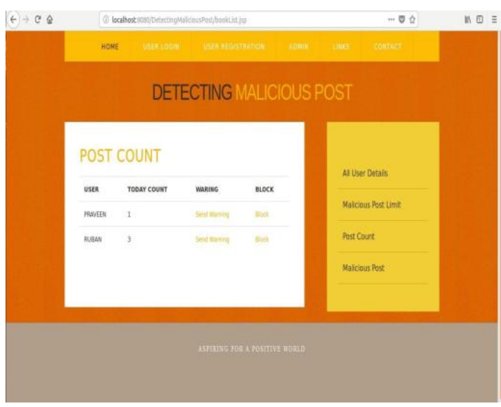

Figure7: Number of Malicious Content Posted

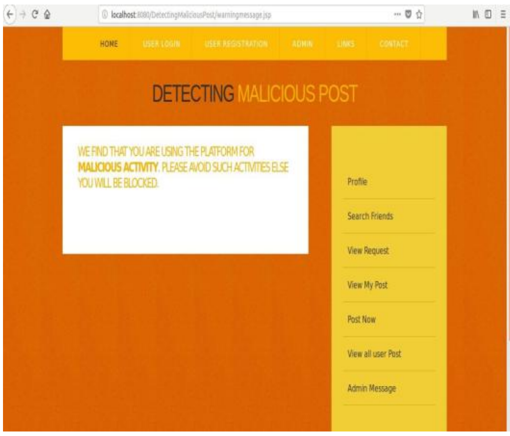

Figure.8 Blocking of the user

V.CONCLUSION

In this algorithms, we analysis the user behaviour based on the history and tweets. Naive Bayesian for malicious content identification on the tweets. Warning and blocking the user based on the algorithm result.

In future image and video trends will be manipulates and integrated with the text analysis in order to reduce the computational complexity for the Trends manipulation.

VI.REFERENCES

[1]. Jeff Elder,"Inside a Twitter robot factory",Wall Street Journal,2013.

[2]. Ginsberg.J,Mohebbi M. H, Patel R.S,Brammer.L , Smolinski M. S., and Brilliant.L," Detecting influenza epidemics using search engine query data",Nature vol457, pp.7232, 1012-4.

[3]. Nikolov, S. Trend or No Trend," A Novel Nonparametric Method for Classifying Time Series",Doctoral dissertation, Massachusetts Institute of Technology,2012.

[4]. Just.M,Crigler.A,Metaxas.P, and Mustafaraj.E,"It's Trending on Twitter-An Analysis of the Twitter Manipulations",APSA 2012 Annual Meeting Paper,2012.

[5]. Ratkiewicz.J,Conover.M, and Meiss.M,"Detecting and tracking the spread of astroturf memes in microblog streams",5th International Conference on Weblogs and Social Media, 2010.

[6]. Becker.H,Naaman.M,and Gravano.L," Beyond trending topics: Real-world event identification on twitter", ICWSM,2011.

[7]. Zubiaga.A,Spina.D,and Martinez.R,"Classifying Trending Topics: A Typology of Conversation Triggers on Twitter", CIKM, 2011.

[8]. Agarwal M.K ,Ramamritham.K , and Bhide.M , "Identifying Real World Events in Highly Dynamic Environments", VLDB,2012.

[9]. Naaman.M,Becker.H,andGravano.L,"Hip and trendy: Charac-terizing emerging trends on Twitter",Journal of the American Society for Information Science and Technology, vol62(5),pp.902-918,2011.

[10]. Lee.K,Palsetia.D,Narayanan.R,Patwary M.M.A,Agrawal.A,and Choudhary.A," Twitter Trending Topic Classification",IEEE 11th International Conference on Data Mining Workshops,pp.251-258,2011.

[11]. Kasiviswanathan S.P,Melville.P,Banerjee.A,and Sindhwani.V," Emerging topic detection using dictionary learning",CIKM, 2011.

[12]. Altshuler.Y,and Pan.W,"Trends Prediction Using Social Diffusion Models", Social Networks,2011.

[13]. Fang.F,Pervin.N,Datta.A,Dutta.K,and VanderMeer.D,"Detecting Twitter Trends in Real-Time",WITS,pp.49-54,2011.

[14]. Mathioudakis.M,and Koudas.N,"Twittermonitor: trend detection over the twitter stream",Proceedings of the 2010 ACM SIGMOD International Conference on Management of data,pp.1155-1158,ACM.

[15]. Egele.M,Kruegel.C,and Vigna.G,"COMPA:Detecting Compromised Accounts on Social Network",NDSS,2013.