ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org

Vol 10, Issuse.2 April 2022

Using a Supervised Deep Learning Machine for Online Social Spam Detection

[1]Omni Jyothi Gudiguntla., [2]Shaik Karishma, [3]Yamini Kasanaboina ,[4]Bhoomika Vangala,

[5]Dr Vaka Murali Mohan

ABSTRACT:

As the online social network has developed rapidly, platforms like Twitter have become more important in everyday life and the primary target of spammers. An imbalanced distribution of data on Twitter has been created by both spam and non-spam, making spam identification a challenging operation that requires consideration of many factors. This research analyses the attributes of Twitter spam, including the user, content, activity, and relationship, and uses that information to inform the development of a novel spam detection algorithm, the Improved Incremental Fuzzy-kernel-regularized Extreme Learning Machine (I2FELM), which can accurately identify such posts as spam. The results of the practical validation show that the proposed I2FELM can accurately distinguish between balanced and unbalanced datasets. Furthermore, the I2FELM can more accurately identify spam with less features, demonstrating the efficacy of the algorithm.

INTRODUCTION:

In recent years, both the Internet and the use of smart terminals have advanced at a dizzying rate. As a result, OSNs become an important conduit via which individuals learn new things, share what they know with others, create social bonds, and pass the time. User adoption, content creation, group interaction, and information dissemination on online social networks have far-reaching effects on social stability, organisational management models, and people's day-to-day work and life [1, 2]. This is because of the complex nature of the online social network structure, the scale of the group, and the massive, rapid, and difficult traceability of information generation. The identification of Twitter spam, for instance, may improve the process of evaluating, directing, and monitoring activities inside social networks and the administration of networks.

Currently, feature selection and detection algorithm selection are the two most difficult aspects of Twitter spam research. Below, we've described the specifics.

Department of CSE, Malla Reddy College of Engineering for Women

Maisammaguda, Medchal, Hyderabad, Telangana

E-Mail: krishomni@gmail.com

algorithms are detected on unbalanced dataset, their performance will decline. Accordingly, an algorithm capable of effectively exploiting multi-dimensional characteristics and exhibiting continuous feasibility in the face of imbalanced datasets should be adopted. By understanding and summarizing the research achievements of predecessors, four novel characteristics are proposed to express the Twitter datasets accurately and improve supervised machine learning algorithm to deal with unbalanced datasets to detect Twitter spam effectively. The details are illustrated below: 1) How to select the full category feature and pay attention to the correlation between the characteristics of the social network account helps enhance the accuracy of identifying spam users. This study considers the Twitter spam attributes composed by the user attribute, content, activity and relationship to express the user characteristic and detect the spam accurately. 2) This study proposes a novel incremental Twitter spam assessment algorithm, termed as the Improved Incremental Fuzzy-kernel-regularized Extreme Learning Machine (I2FELM) to enhance the accuracy in dealing with the unbalanced data. 3) I2FELM is capable of enhancing the performance using Cholesky factorization without square root and composite kernel function. Besides,, it can automatically determine the optimal number of hidden layer nodes by gradually adding new hidden nodes one by one. 4) The I2FELM introduces the fuzzy weight as a method to address the unbalanced problem, which can apply to each input and facilitate the learning of output weights. 5) On the public dataset and the collected dataset, a range of index parameters and experimental verification methods are adopted to ascertain the performance of I2FELM, and spam is assessed based on the imbalance data problem and few characteristics.

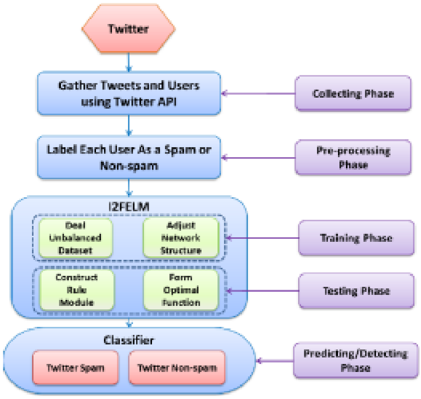
RELATED WORK:

Extensively studied, several approaches related to social spam detection have been proposed (e.g., spam characteristics and assessment algorithm). Benevenuto et al. [3] considered two attribute sets, namely, content attributes and user attributes, to distinguish one user class from the other and exploited the mentioned characteristics as attributes of SVM process to classify users as either spam or non-spam. Lee et al. [4] conducted the statistical analysis of the properties of the mentioned spam profiles to create spam classifier to actively filter out existing and novel spam. Based on the mentioned profile characteristics, the authors developed meta-classifiers (Decorate, Logit Boost, etc.) to identify previously unknown spam. Stringhini et al. [5] initially created a set of honey net accounts (honey-profiles) on Twitter and then identified multiple characteristics that allow authors to detect spam. Lastly, the RF model was built to detect spam and employed in a Twitter dataset. Wang [6] developed the novel content-based characteristics and graph-based characteristics to facilitate spam detection; besides, a Bayesian classification algorithm was adopted to distinguish suspicious behaviors from normal ones. Chu et al. [7] presented the collective perspective and focused on identifying spam campaigns that manipulate multiple accounts to spread spam on Twitter. An automatic classification system was designed based on RF and a variety of characteristics, i.e., individual tweet/account levels to classify spam campaigns. In Meda et al.’s work [8], a standard Principal Component Analysis (PCA) algorithm was exploited to reduce the dimensionality of the 62 feature to the 20 characteristics, 10 characteristics, and 5 characteristics, and then three different machine learning algorithm (SVM, ELM, RF) were adopted to support spam detection in Twitter. Wang et al. [9] studied the suitability of five classification algorithms of Bayesian, KNN, SVM, DT, and RF at the detection stage; they took four different feature sets of user characteristics, content characteristics, n-grams, and sentiment characteristics to the social spam detection task. Zheng et al. [10] extracted a set of characteristics from content-based and user-based feature and applied into SVM based spam detection algorithm. Chen et al. [11] built a hybrid model that uses SVM and NB to distinguish suspect users from normal ones based on the user-based characteristics and content-based characteristics. During the assessment, the authors assessed the impact of different factors on spam detection performance, covering discretization of functionality, size of learning data, and data related to time. Chen et al. [12] proposed an Lfun approach to identify the ‘‘Spam Drift’’ problem in statistical features based Twitter spam detection. They compared Lfun to four traditional machine learning algorithms and evaluated the performance of Lfun approach in terms of overall accuracy, F-measure and Detection Rate. He et al. [13] proposed an analysis approach based on information entropy and incremental learning to study how various features affect the performance of an RBF-based SVM spam detector, through this effort, they attempted to increase the awareness of a spam by sensing the features of a spam. Teng et al. [14] proposed a selfadaptive and collaborative intrusion detection model is built by applying the Environments classes, agents, roles, groups, and objects (E-CARGO) model. Wu et al. [15] found that most of current spam detection techniques are based on feature selection and machine learning classification (e.g. DT, RF and NB). Liu et al. [16] reviewed the schemes and systems proposed to deal with an increasing number of cyber security threats. The work can extract information from data sources and applied analytics/algorithm (e.g. machine learning) to make a decision. Sun et al. [17] presented an overview and research outlook of the emerging field, i.e., cybersecurity incident prediction. They also extracted and summarized the research methodology at critical phases of predicting cybersecurity incident. In the research of Coulter et al. [18], a new research methodology of data-driven cyber security (DDCS) was demonstrated, and its application in social and Internet traffic analysis was studied. DDCS shows the strong link between data, model, and methodology during the review of key recent works in Twitter spam detection and IP traffic classification.

MODEL:

Nowadays, the Twitter spam attributes primarily focus on the tweet-based characteristics and user-based characteristics. The assessment algorithms refer to general machine learning methods based on the relationship between spam characteristics and detection. For instance, the methods adopted are primarily SVM, DT, RF, BP, RBF, ELM, and XGBoost, etc. In the face of the multi-dimensional characteristics and the imbalance dataset, the performance of the mentioned algorithms requires enhancement. This study proposes the Twitter spam attributes consisting of user attribute, content, activity and relationship characteristics to detect spam exactly; each feature can be captured in the Twitter to ensure its integrity and reliability. Besides, I2FELM is designed to address the multi-dimension and non-balance problem to achieve the high assessment accuracy. The procedure of Twitter spam detection is illustrated in Fig. 1. First, dataset from Twitter is collected to form user attribute, content, activity and relationship characteristic sets. Second, the collected dataset is preprocessed for labeling each user as a spam or non-spam. Third, the proposed I2FELM is adopted to tackle down the unbalanced problem. An optimal function of I2FELM is formed by training and testing phase. Based on the formed optimal function, I2FELM can effectively assess Twitter spam of novel dataset in the classification phase.

In this study, the feature set is composed of user attribute, content, activity and relationship in the online social network, and the details are listed in Table 1. To be specific, the user attribute feature refers to the period of the existence of the account, the number of registered locations, the number of lists added by the user, and the number of tweets sent by the user. Besides, the content feature covers the numbers of retweets this tweet, favorites this tweet received, hashtags and URLs this tweet included, characters and digits in this tweet, the mentioned time of this tweet, as well as spam words in this tweet, content similarity score.



The procedure of Twitter spam detection.

DETECTION ALGORITHM:

Unbalanced data is an issue in online social networks since the number of non-spam posts is far higher than the number of spam posts. The I2FELM is thus designed to implement RELM to find a solution to the issue. Since the suggested technique is based on Cholesky factorization without a square root and a composite kernel function, it is able to compensate for the non-balance of datasets by employing fuzzy membership to enhance accuracy.

Huang et al. [34] introduced the equality constrained optimization-based ELM to improve the generalisation capabilities of the standard ELM-based SFLNs. Their method introduces structural risk as a regularisation term. Using the parameter C, the so-called RELM controls the balance between structural risk and empirical risk. One such formulation of the suggested restricted optimization is

$$\begin{aligned} &\min \Gamma_{RELM} = \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{i=1}^{N}\|\xi_i\|^2 \\ &s.t.\ h(x_i)^T\ \beta = t_i - \xi_i \quad i = 1, \ldots, N \end{aligned}$$

I2FELM:

In order to train SLFNs with a single hidden layer using RELM, the I2FELM is presented. The core idea behind I2FELM is that generalised SLFNs' hidden layer should not be adjusted. As a result, it may be used directly in regression and multiclass classification.

The input layer, hidden layer, and output layer are the three groups of nodes that make up I2FELM. The input dataset's uneven weights must be addressed at the input layer.

$$\begin{aligned} &\min \Gamma_{I2FELM} = \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{i=1}^{N}S_i\|\xi_i\|^2 \\ &s.t.\ h(x_i)^T\ \beta = t_i - \xi_i \quad i = 1, \ldots, N, \end{aligned}$$

EXPERIENCE:

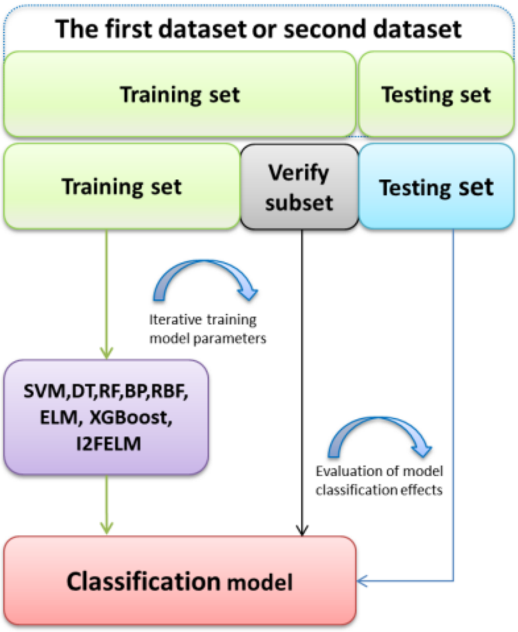
DATASET COLLECTION:

In order to evaluate the performance of the experiment, we use two datasets.

The Aponador dataset is the first available public dataset. Each record in this dataset has 59 characteristics and 2 classifications, and it was obtained using Brazil's popular location-based social network.

This research also uses the Twitter API and the Twitter4J software to collect a second dataset consisting of 43 million tweets from around 16 million users that include In the month of June in 2017, prominent trends were updated every day. During our dataset's pre-processing step for I2FELM, we labelled Twitter accounts as spam or non-spam using the approach described in [40]. To solve the issue of recognising spam and dangerous Tweets, the method [40] presented a hybrid approach, which combines a blacklist with algorithms tailored to social networks. According to the data, it has been shown that blacklisting, in combination with other analytical techniques, may successfully detect harmful Tweets.

Consequently, blacklists may stop the spam from getting through. Graph theory tells us that a group of users using a round-robin protocol will form a bipartite clique. Because of this, bipartite cliques in such a network should raise serious red flags; it's very unlikely that genuine users would act in this manner by chance. In addition to the traditional blacklist, a clique-discovery technique is used to help spot spam. In the end, 0.81 million accounts were sorted into spam and non-spam categories, and each record had 62 features.


The first dataset or second dataset
Training set
Testing set
Training set
Verify subset
Testing set
Iterative training model parameters
SVM,DT,RF,BP,RBF, ELM, XGBoost, I2FELM
Evaluation of model classification effects
Classification model

EXPERIMENTAL RESULT AND COMPARISON:

Here we do the three experiments on the first and second datasets.

Every evaluation is based on averaging the results of 10 separate attempts at the same task, which eliminates the possibility of random outcomes.

In-depth procedures for conducting experiments

The effectiveness of eight algorithms on the first and second datasets is validated in the first experiment by establishing balanced datasets.

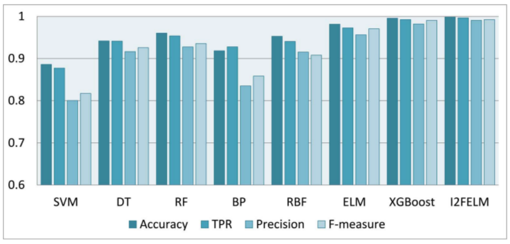
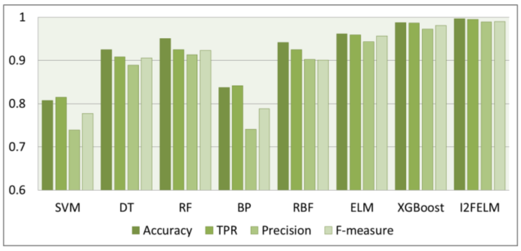
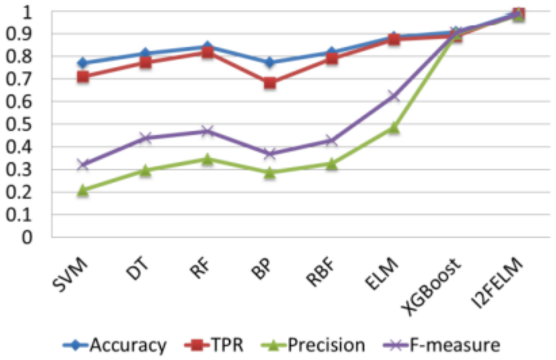


FIGURE 3. The comparison results of the first dataset.

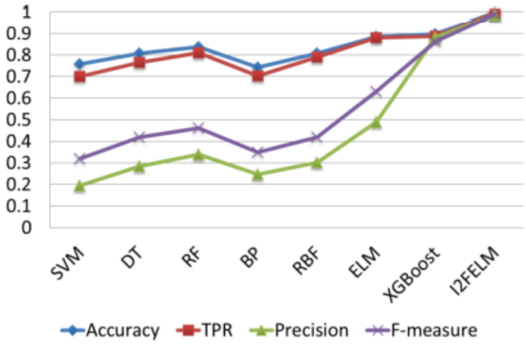


In the first and second datasets, the performance of the eight methods is evaluated using the accuracy, TPR, precision, and F-measure.

Each of the eight methods shows excellent performance on the balanced datasets, with I2FELM achieving the highest index parameter values.



(a) the first dataset



(b) the second dataset

CONCLUSION:

This paper introduces a unique approach to detecting spam on Twitter by using a feature set that takes into account information about users' attributes, content, activities, and relationships in the virtual world. To further improve accuracy, the I2FELM method is used to evaluate spam. This technique makes use of fuzzy weights to deal with the issue of imbalanced data.

In addition, a composite kernel function and Cholesky factorization without the square root are used to improve performance. To add, an appropriate threshold for the percentage of concealed nodes may be computed mechanically. Experience validation shows that the proposed I2FELM works well for evaluating spam in online social networks, and it may be used to multi-dimensional balanced or unbalanced datasets.

The following lines of inquiry will serve as the focal points of the next phase of this project. In the first place, additional elements (such as semantic analysis and emotion analysis) would be evaluated in order to accurately detect spam. In addition, we want to use oversampling and a feature selection approach to choose appropriate feature sets and enhance model adaptability [21, 28, 29]. However, since there isn't a large enough labelled data set in the social network, we'll be using a semi-supervised learning approach in place of the traditional supervised learning in the I2FELM model to automatically identify Twitter spam.

REFERENCES:

[1] M. Chakraborty, S. Pal, R. Pramanik, and C. Ravindranath Chowdary, ‘‘Recent developments in social spam detection and combating techniques: A survey,’’ Inf. Process. Manage., vol. 52, no. 6, pp. 1053–1073, Nov. 2016.

[2] R. K. Dewang and A. K. Singh, ‘‘State-of-art approaches for review spammer detection: A survey,’’ J. Intell. Inf. Syst., vol. 50, no. 2, pp. 231–264, Apr. 2018.

[3] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ‘‘Detecting spammers on Twitter,’’ in Proc. CEAS, vol. 6, 2010, p. 12.

[4] 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. 435–442.

[5] G. Stringhini, C. Kruegel, and G. Vigna, ‘‘Detecting spammers on social networks,’’ in Proc. 26th Annu. Comput. Secur. Appl. Conf. (ACSAC), 2010, pp. 1–9.

[6] A. H. Wang, ‘‘Don’t follow me: Spam detection in Twitter,’’ in Proc. Int. Conf. Secur. Cryptogr. (SECRYPT), Jul. 2010, pp. 1–10.

[7] Z. Chu, I. Widjaja, and H. Wang, ‘‘Detecting social spam campaigns on Twitter,’’ in Proc. Int. Conf. Appl. Cryptogr. Netw. Secur. Cham, Switzerland: Springer, 2012, pp. 455–472.

[8] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, ‘‘A machine learning approach for Twitter spammers detection,’’ in Proc. Int. Carnahan Conf. Secur. Technol. (ICCST), Oct. 2014, pp. 1–6.

[9] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, ‘‘Making the most of tweet-inherent features for social spam detection on Twitter,’’ 2015, arXiv:1503.07405. [Online]. Available: http://arxiv.org/abs/1503.07405

[10] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, ‘‘Detecting spammers on social networks,’’ Neurocomputing, vol. 159, pp. 27–34, Jul. 2015.

[11] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian, ‘‘A performance evaluation of machine learning-based streaming spam tweets detection,’’ IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65–76, Sep. 2015.

[12] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ‘‘Statistical features-based real-time detection of drifted Twitter spam,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.

[13] H. He, A. Tiwari, J. Mehnen, T. Watson, C. Maple, Y. Jin, and B. Gabrys, ‘‘Incremental information gain analysis of input attribute impact on RBFkernel SVM spam detection,’’ in Proc. IEEE Congr. Evol. Comput. (CEC), Jul. 2016, pp. 1022–1029.

[14] S. Teng, N. Wu, H. Zhu, L. Teng, and W. Zhang, ‘‘SVM-DT-based adaptive and collaborative intrusion detection,’’ IEEE/CAA J. Automatica Sinica, vol. 5, no. 1, pp. 108–118, Jan. 2018.

[15] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ‘‘Twitter spam detection: Survey of new approaches and comparative study,’’ Comput. Secur., vol. 76, pp. 265–284, Jul. 2018.

[16] L. Liu, O. De Vel, Q.-L. Han, J. Zhang, and Y. Xiang, ‘‘Detecting and preventing cyber insider threats: A survey,’’ IEEE Commun. Surveys Tuts., vol. 20, no. 2, pp. 1397–1417, 2nd Quart., 2018.

[17] N. Sun, J. Zhang, P. Rimba, S. Gao, L. Y. Zhang, and Y. Xiang, ‘‘Datadriven cybersecurity incident prediction: A survey,’’ IEEE Commun. Surveys Tuts., vol. 21, no. 2, pp. 1744–1772, 2nd Quart., 2019.

[18] R. Coulter, Q.-L. Han, L. Pan, J. Zhang, and Y. Xiang, ‘‘Data-driven cyber security in perspective–intelligent traffic analysis,’’ IEEE Trans. Cybern., early access, Oct. 15, 2019, doi: 10.1109/TCYB.2019.2940940.

[19] R. Dayani, N. Chhabra, T. Kadian, and R. Kaushal, ‘‘Rumor detection in Twitter: An analysis in retrospect,’’ in Proc. IEEE Int. Conf. Adv. Netw. Telecommuncations Syst. (ANTS), Dec. 2015, pp. 1–3.

[20] J.-J. Sheu, Y.-K. Chen, K.-T. Chu, J.-H. Tang, and W.-P. Yang, ‘‘An intelligent three-phase spam filtering method based on decision tree data mining,’’ Secur. Commun. Netw., vol. 9, no. 17, pp. 4013–4026, Nov. 2016.

[21] H. Liu, M. Zhou, and Q. Liu, ‘‘An embedded feature selection method for imbalanced data classification,’’

IEEE/CAA J. Automatica Sinica, vol. 6, no. 3, pp. 703–715, May 2019.

[22] 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, Aug. 2016.

[23] C. Meda, E. Ragusa, C. Gianoglio, R. Zunino, A. Ottaviano, E. Scillia, and R. Surlinelli, ‘‘Spam detection of Twitter traffic: A framework based on random forests and non-uniform feature sampling,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 811–817.

[24] G. Lin, N. Sun, S. Nepal, J. Zhang, Y. Xiang, and H. Hassan, ‘‘Statistical Twitter spam detection demystified: Performance, stability and scalability,’’ IEEE Access, vol. 5, pp. 11142–11154, 2017.

[25] S. Liu, Y. Wang, C. Chen, and Y. Xiang, ‘‘An ensemble learning approach for addressing the class imbalance problem in Twitter spam detection,’’ in Information Security and Privacy, vol. 9722. Sydney, NSW, Australia: Springer, 2016, pp. 215–228.

[26] S. Liu, Y. Wang, J. Zhang, C. Chen, and Y. Xiang, ‘‘Addressing the class imbalance problem in Twitter spam detection using ensemble learning,’’ Comput. Secur., vol. 69, pp. 35–49, Aug. 2017.

[27] X. Wang, ‘‘Ladle furnace temperature prediction model based on largescale data with random forest,’’ IEEE/CAA J. Automatica Sinica, vol. 4, no. 4, pp. 770–774, 2017.

[28] W. Tang, Z. Ding, and M. Zhou, ‘‘A spammer identification method for class imbalanced weibo datasets,’’ IEEE Access, vol. 7, pp. 29193–29201, 2019.

[29] X. Wang, Q. Kang, J. An, and M. Zhou, ‘‘Drifted Twitter spam classification using multiscale detection test on K-L divergence,’’ IEEE Access, vol. 7, pp. 108384–108394, 2019. [30] H. He, T. Watson, C. Maple, J. Mehnen, and A. Tiwari, ‘‘A new semantic attribute deep learning with a linguistic attribute hierarchy for spam detection,’’ in Proc. Int. Joint Conf. Neural Netw. (IJCNN), May 2017, pp. 3862–3869.

Student Details:

Omni Jyothi Gudiguntla,

19RG1A0520, CSE Department, Malla Reddy College of Engineering for Women Maisammaguda, Medchal, Hyderabad, Telangana

Shaik Karishma,

19RG1A0550, CSE Department, Malla Reddy College of Engineering for Women Maisammaguda, Medchal, Hyderabad, Telangana

Yamini Kasanaboina,

19RG1A0527, CSE Department, Malla Reddy College of Engineering for Women Maisammaguda, Medchal, Hyderabad, Telangana

Bhoomika Vangala,

19RG1A0558, CSE Department, Malla Reddy College of Engineering for Women Maisammaguda, Medchal, Hyderabad, Telangana

Guide Details:

Dr Vaka Murali Mohan,

Guide, Principal and Professor, CSE Department, Malla Reddy College of Engineering for Women Maisammaguda, Medchal, Hyderabad, Telangana