# Heterogeneous Feature Analysis on Twitter Data Set for Identification of Spam Messages

Valliyammai Chinnaiah and Cinu C Kiliroor Department of Computer Technology, MIT Campus Anna University, India

**Abstract:** Spam is an undesirable content that present on online social networking sites, while spammers are the users who post this content on social networking sites. Unwanted messages posted on Twitter may have several goals and the spam tweets can interfere with statistics presented by Twitter mining tools and squander users' attention.. Since Twitter has achieved a lot of attractiveness through-out the world, the interest towards it by the spammers and malevolent users is also increases. To overcome the spam problems many researchers proposed ideas using machine learning algorithms for the identification of spam messages. Not only the selection of classifiers but also the variegated feature analysis is essential for the identification of irrelevant messages in social networks. The proposed model performs a heterogeneous feature analysis on the twitter data streams for classifying the unsolicited messages using binary and continuous feature extraction with sentiment analysis on social network datasets. The features created are assessed using significant stratagems and the finest features are selected. A classifier model is built using these feature vectors to predict and identify the spam messages in Twitter. The experimental results clearly show that the proposed Sentiment Analysis based Binary and Continuous Feature Extraction model with Random Forest (SA-BC-RF) approach classifies the spam messages from the social networks with an accuracy of 90.72% when compared with the other state-of-the-art methods.

Keywords: Spam filter, random forest, heterogeneous feature analysis, online social network.

Received November 9, 2019; accepted February 15, 2021 https://doi.org/10.34028/iajit/19/1/5

## **1. Introduction**

Twitter is one of the famous social networking platforms and its users' communicate with other users by posting tweets, and procure the hottest information from other users' tweets by following them. The spam performs a variety of forbidden behaviours and creates an adverse impression on users. Generally, the basic features are collected directly from the datasets and the derived features are inferred from the prevailing data attributes. The feature engineering process extract the relevant data from the twitter dataset based on the data analysis for designing of effective models for spam message classification and it develops a big impact for useful predictions when it matches with the existing classification model. The feature selection process performs the extraction of required and important features from the existing features. Feature selection is performed for model simplicity, to reduce the training times, to avoid the dimensionality issues, enhanced generalisation by reducing overfitting. The needless features are abolished without much information loss in the feature selection process. Sentiment analysis is a text mining technique to analyse the sentiment of a text message [13]. The sentiment analysis also called opinion mining may be positive, negative or neutral and is used for analysing the conversations, opinions and sharing of views for planning business strategy, political analysis and also for assessing public opinions. Sentiment analysis has two major aspects to

be concerned with-polarity and subjectivity. Polarity sentiment analysis takes into account the positive or negative terms that occur in a sentence. The subjectivity sentiment analysis aims at classifying the given text into two broad classes: objective and subjective. The words and phrases in subjectivity depend on their context and an objective document may also contain subjective sentences. Hence, sometimes it is difficult to find and compared to polarity. Being subjective refers to personal opinions, interpretations, emotions and judgement and objective analysis is fact-based, measurable and observable.

Presence of undesirable messages in the online social networks [23] is an existing issue and different research works are concentrated on several classification algorithms for uninvited message classification in social networks. The concern of feature selection and sentiment analysis are also vital in addition to the selection of a classifier, and to perform the spam message identification more effective. Binary and continuous feature extraction and text sentiment feature extraction are done in the proposed model for effective spam message classification in social networks.

The rest of this paper is organised as follows. Section 2 delivers the literature survey, section 3 presents the proposed methodology. Section 4 comprises the results and their analysis. Section 5 concludes with future work.

# 2. The Related Works

Spammers attracted towards Online Social Networks (OSNs) because OSNs have attained an excellent growth in past decades and thereby spammers use OSN to spread the spam. Initially, the spammers spread unsolicited messages for advertisement purposes. Later spam messages were used to gain illegal profit and sometimes it may be malicious also. Spam message can be an advertisement message or phishing message. Using spam, spammers can collect private information from users. Spam messages also affect the performance of OSN which reduces the users' trust towards OSN. So, it is very crucial to find an effective method to identify the unwanted messages in OSN. A literature review of various existing methods related to spam messages identification in emails and OSNs is presented in the following section.

The importance of feature selection for the improvement in the performance of the algorithms was discussed in [27]. The different approaches for feature selection were vividly presented and compared. The merits and the demerits were clearly presented for each feature selection approaches. Al-Qurishi et al. [4] discussed that finding the difference between quiet users and mischievous users is difficult in. The user can be verified using captcha. Mutual confidence is maintained using social graph-user conversations. The feature selection process for text classification was discussed in [9]. Mutual Information (MI), correlation co-efficient, maximum discrimination, linear measure were deployed for the analysis of the features. The selected features were fed to wrapper methods for further refinement. MI is performed poorly as compared to other measures due to its bias toward favouring rare terms and its sensitivity to errors in probability estimation. The performance comparison of different machine learning algorithms was analysed and discussed in [18].

An attribute selection approach was used to elect the attributes for uninvited message identification in OSN is presented in [10]. The identification of spam-tweets without knowing the previous background of the user and with the aid of language analysis on trending topics was proposed in [14]. Halawi et al. [15] proposed spam detection model using ontology-based approach. The ontologies are derived and used for generation of a dictionary that validates real tweet messages from random topics. The ratio among dictionary and tweets is used to reflect the legitimacy of messages. The statistical properties of spam tweets vary over time; thus the performance of existing machine learning classifiers decreases. It is referred to as twitter spam drift. The deep analysis on statistical features of one million spam tweets and one million non-spam tweets was performed in [7]. The grouping of short messages using a weighting scheme to improve the discriminating power was proposed in [3]. To improve the discriminating power of terms the weighting scheme is used. Murugan and Devi [22] proposed a hybrid technique was used to identify spam messages from twitter. Mohammed N proposed a deep learning based method to extract synonyms from the text [21]. Fazil and Abulaish [12] proposed a spammer detection model based on features such as content, interaction, and community which are selected for identification of spammers in twitter. A survey is conducted for text classification based on feature selection was discussed in [9]. Content-based feature selection and was used for the identification of spam messages in the phone using a genetic algorithmic methodology is proposed in [2].

Sentiment analysis is performed in the OSN to know the users' characteristics in a personalised way to get a better result from the analysis. The emotions about the users are easily identified from the sentiment analysis. The characteristics of the spam and ham messages are determined with the aid of sentiment analysis. Sentiment analysis based polarity detection of airline tweets was proposed in [1]. The major drawback is that the presence of positive and negative words may not give accurate results in case of sarcastic tweets. A survey on twitter research is performed in [5].

The analysed key characteristics of tweets were message length, writing technique, availability, topics, real-time, emoticons, target, hash tags, and special symbols. El-Alaoui *et al.* [11] discussed the classification of tweets into several classes by introducing new features that strongly tune the polarity degree of a post. A new sentiment ontology was used to conduct context sensitive sentiment analysis of online opinion post in the stock market. Schouten and Frasincar [24] explained the detailed assessment of the different perspective of sentiment analysis on twitter datasets.

The message filtering systems have three layers: user interface, support for external social network applications, and social network manager. The essential OSN functionalities were discussed in [26]. text represented was analysed by The considering some factors like right words, number of exclamatory marks, number of question marks, number of capital words. The spammers message contains irrelevant texts. Hence the users OSN wall were scanned and checked for spam. Class imbalance problem was addressed with the support of random undersampling, oversampling and fuzzy based approach [20]. Shi and Xie [25] discussed a collaborative approach was used here with the fingerprint technique for spam message filtering. Neural network based filtering of spam messages was proposed in [6].

Ji and Zhang [16]. discussed that web pages might contain spam data which and those data were identified and filtered out efficiently in The contents of web pages were analysed and then spam data were filtered out. Usually, non-spam web pages have statistical properties whereas spam web pages do not have statistical properties. The content feature distributions are distinct for various web spam data sets. So, spammers randomly create spam pages. Using this feature, spam web pages were found and filtered out in the analysis part. The correlation coefficient of the two features was the measure of the degree of correlation between two features.

Liu *et al.* [19] explored about the identification of spam messages using some patterns. If new words arrived as spam, then it added to the dictionary. A self-extensible spam dictionary is used for this operation.

# 3. Variegated Feature Analysis on Twitter Datastreams for Unwanted Message Classifications

The proposed model for twitter spam message detection is given in Figure 1 which performs a feature engineering process to find significant features to help in the designing of an effective spam classifier. Class Imbalance problem occurs in the dataset may lead to a high bias and such classification will generally follow a majority rule. Therefore, the dataset is balanced by randomly sub sampling the same number of ham tweets as spam tweets. Exploring the existence of null values in the dataset and eliminating them is also performed in the dataset. Binary and continuous feature extraction and text sentiment feature extraction are proposed in this paper for better classification of unwanted messages in social networks.

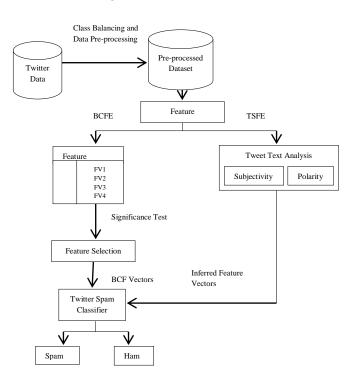
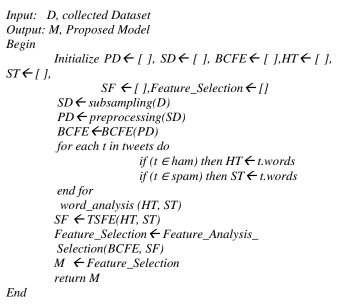


Figure 1. Proposed model with variegated feature analysis for twitter spam message classification.

Algorithm (1) describes the overall workflow of the

creation of proposed model from the dataset. It describes the creation of Subsampled Dataset (SD) and Preprocessed Dataset (PD) and the Binary and Continuous Feature Extraction (BCFE) from the existing features. For the Text Sentiment Feature Extraction (TSFE), tweet words are extracted and created as Ham Text (HT) and Spam Text (ST). Word Analysis is performed to infer the significance of words in both ham and spam corpus. Sentiment Analysis is performed to generate the Sentiment Features (SF) using polarity and subjectivity score calculation. Feature Selection approach is done to select the significant features and these features are coupled to generate the proposed model.

Algorithm 1: Spam Message Classification with Variegated Feature Analysis



#### **3.1. Binary and Continuous Feature Extraction**

BCFE approach creates a new set of binary and continuous features from the existing features. Profound data analysis is done to find every feature in the data set. The continuous features are analysed using Kernel Density Estimation (KDE) and the binary features are analysed using Count Bar-Plot (CBP) to find the distribution of features in the dataset. The features are examined and tested for significance and the best feature vectors are used for building supervised machine learning model. BCFE features extracted from the dataset and it combined and stored in to a list (Lbc). The feature vectors created are tested for validation using Sample T-Test (STT) is shown in Equation (1).

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \tag{1}$$

Where,  $\bar{x}$  is the sample mean,  $s^2$  is the sample variance, *n* is the sample size,  $\mu$  is the specified population mean and *t* is the test statistic value. The significant features selected are used to develop the

model.

The significant features are nominated for the proposed model design to identify the spam messages from OSN effectively. In order to make features more meaningful, new binary and continuous values are created from features. The binary features and continuous features considered in this work is shown in Tables 1 and 2 respectively. The binary and continuous feature analysis is performed in our previous work [17].

Table	1.	Binary	features.
-------	----	--------	-----------

<b>Binary Features</b>	Explanation	
Has-hashtags	Check the presence or absence of hashtag in tweets	
Has-media	Check the presence or absence of has media content in	
	tweets	
Has-urls	Check the presence or absence of URL in tweets	
Has-favorite-count	Check the presence or absence of favourite count in	
	tweets	
Has-place	Check the presence or absence of location in tweets	
Has-retweet-count	Check the presence or absence of has retweet count in	
	tweets	
Has-user-name	Check the presence or absence of user name in tweets	
Has-user-	Check the presence or absence of user description in	
description	tweets	
is-retweet	Check whether a tweet is a re-tweet	
is_user_verfied	Check whether the user is a verified user or not	

Table 2. Continuous features.

<b>Continuous Features</b>	Explanation	
Digits	Count of digits	
Cap	Count of capital words	
length	Span of a tweet	
f-ratio	Ratio between Followers and Followings	
num_hashtag	Count of hashtags	

The Exploratory Data Analysis (EDA) is performed on every feature in the data set. For continuous attributes, the dissemination of the attributes are analysed using KDE to visually recognize any important differences in the dispersal across each type of tweet. The probability distribution of spam or ham messages in the datasets vividly depends on the features. So, not only the choosing of machine learning based classification models but also the selection of features is crucial. STT computation is performed for the confirmation of the input features to find the most significant set of features and that subset of features predicts the tweet label.

#### **3.2. Text Sentiment Feature Extraction**

Nowadays spammers are doing a very tactical approach by adding genuine words to the spam posts. Not only the direct words comparison, it is also essential to gather additional information from the messages for better spam message identification. So, the TSFE is performed in real time datasets. The stop words removal and stemming of words in the tweet datasets performs in the pre-processing stage. Algorithm 2 describes the Text Sentiment Feature Engineering task. The empty lists are storing Polarity Score (PS), Subjectivity Score (SS), Polarity Label (PL) and Subjectivity Label (SL). The created and ST are used for calculating polarity and subjectivity scores. These scores are used for creating polarity label (positive, negative, neutral) and subjectivity label (subjective, objective, neutral). Using these labels, new sentiment features has-positive and has-neutralsubjective are created. Polarity score, subjectivity score, has-positive and has-neutral-subjective is present in the List of Sentiment Features (LSF).

Algorithm 2: TSFE

```
Input: HT, ST

Output: Lsf

Begin

Initialize PS \leftarrow [], SS \leftarrow [], PL \leftarrow [], SL \leftarrow []

for each tweets in HT, ST

PS \leftarrow polarity(HT, ST)

SS \leftarrow subjectivity(HT, ST)

PL \leftarrow polarity\_label(PS)

SL \leftarrow subjectivity\_label(SS)

end for

Lsf \leftarrow PS, SS

Lsf \leftarrow binary\_sentiment\_feature(PL, SL)

return Lsf
```

End

The identification of frequency distribution of words in the labelled data set is essential to find out the most frequently used words in the spam and ham data sets. The sentiment analysis is performed on tweets to generate two new sentiment scores namely Polarity (P) and Subjectivity (S). The P value is categorised into positive, negative or neutral and the S value is categorised into subjective, objective or neutral. The polarity and subjectivity score distribution in the dataset is shown in Figures 2 and 3 respectively.

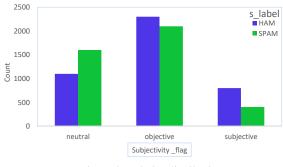


Figure 2. Polarity distribution.

Figure 2 clearly shows that positive words are mostly used in spam tweets when compared with the negative words in spam tweets.

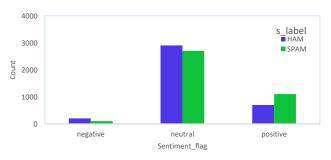


Figure 3. Subjectivity distribution.

Figure 3 shows that most of the spam tweets are neutral. KDE analysis for polarity in Figure 4 shows that most of the spam and ham tweets are neutral. The positive words are used mostly in spam tweets. The ham tweets contain a high amount of negative words than spam tweets.

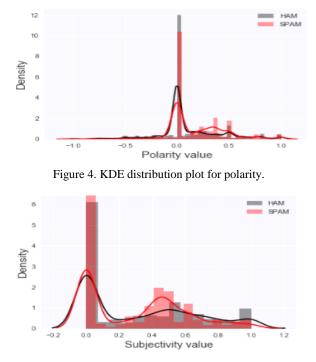


Figure 5. KDE distribution plot for subjectivity.

KDE analysis for subjectivity in Figure 5 shows that ham tweets are mostly objective which deals with common state expressions. Spam tweets are more neutral than ham tweets, which are deal with one or more private state expressions.

```
Algorithm 3: Positive feature extraction from tweets
```

```
Input: DF, MW
Output: DF_U
Begin
       Initialize PL \leftarrow 0, W \leftarrow 0, flag
       for each column c in DF do
                  flag = false
                  W \leftarrow DF.text
                 for each word (w_i \in W) do
                      if word (w_i \in MW) then
                         flag = true
                      end if
                  end for
                  if(flag == true) then
                      PL \leftarrow 1
                  else
                      PL \leftarrow 0
                  end if
                 DF_U \leftarrow PL
       end for
End
```

The polarity analysis of the tweets clearly shows that positive words are mostly used in spam tweets. So, a new feature called has-positive feature is extracted from the tweets and the extraction procedure is given in Algorithm 3.

The Most Words (MW) list contains the frequently used positive words in the spam dataset. A comparison check is performed for each word (w) in Word List (W) from the DF with the MW list. If match occurs the PL get changed and the corresponding feature value gets updated in the Updated Data frame (DFU). Likewise, subjectivity analysis clearly shows that the neutral words are highly present in the spam tweets. The new features are introduced in the proposed model based on the sentiment analysis which increases the accuracy for spam message filtering.

#### 3.3. Spam Message Classification

The proposed model is built with the related significant features for unwanted message classification in the OSN and is implemented based on the Sentiment Analysis with BCFE model. Random Forests (RF) is an ensemble method for classification of data by constructing a multitude of decision trees at training time and produce the class. Random forest classifiers, in general, tend to reduce the variance by training the model on different samples of data.

### 4. Experimental Setup and Results

The twitter Application Programming Interface (API) is used to collect the data streams from twitter. The responses of the APIs are available in Java Script Object Notation (JSON). To collect data from twitter, the authenticated information is used that is supplied by a twitter developer in order to connect and extract the needed data from Twitter's API. The dataset is balanced by randomly subsampling same number of ham tweets as spam tweets. The proposed model is implemented using Python. The tweet IDs and corresponding labels are collected from [8]. The evaluated features are given as input to the classifier. The proposed BCFE with RF (BC-RF) filtering approach provides an accuracy score of 85.23%. ROC analysis of the proposed BC-RF model is shown in Figure 6.

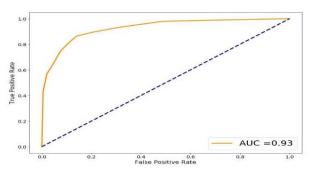


Figure 6. ROC analysis of the proposed BC-RF Filter model.

The proposed model with Sentiment Analysis based BC-RF (SA-BC-RF) filtering approach provides an accuracy score of 90.72%.

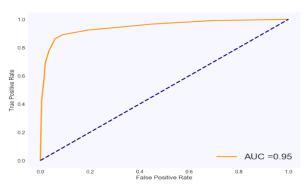


Figure 7. ROC curve of the proposed SA-BC-RF filter model.

ROC analysis of the proposed SA-BC-RF model for better classification of unwanted message from OSN is shown in Figure 7. The accuracy of the proposed Sentiment Analysis based BCFE with Random Forest Filtering Approach (SA-BC-RF) for spam classification in OSN with other models such as Content Based Filtering Approach (CB-FA), Sentiment Analysis based Filtering Approach (SA-FA), proposed BC-RF approach is shown in Figure 8 and it clearly depicts that the proposed SA-BC-RF approach outperformed when compared with other methods.

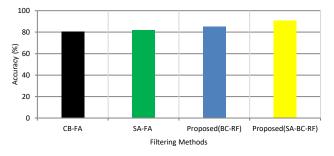


Figure 8. Accuracy of the proposed SA-BC-RF vs. existing models.

#### 5. Conclusions

The intrinsic nature of tweets is analysed based on the features and then it is used to design a strong classifier that helps in classifying a tweet as spam or not. The class balancing procedure is performed to make a balanced dataset. The data mining process is performed for extracting the significant data from a massive stream of Twitter. The main contribution of the proposed model concentrates on the SA-BC-RF approach in the process of identification of unsolicited messages from OSN. The binary and continuous feature distribution in the tweets are analysed to know the importance of the significant features. Feature Selection method STT is used for relevancy test and list of features are derived. The derived features are included to build the classification model. The experimental results clearly show that the proposed SA-BC-RF model attained a better spam message classification. In future, the proposed work can be extended as add-on for web browser so that the real time detection of spam messages can be done in twitter website itself by alerting the user.

#### References

- [1] Adarsh M. and Ravikumar P., "An Effective Method of Predicting the Polarity of Airline Tweets using Sentimental Analysis," in Proceedings of the 4<sup>th</sup> International Conference on Electrical Energy Systems, Chennai, pp. 676-679, 2018.
- [2] Al-Hasan A. and El-Alfy E., "Dendritic Cell Algorithm for Mobile Phone Spam Filtering," *Procedia Computer Science*, vol. 52, no. 1, pp. 244-251, 2015.
- [3] Alsmadi I. and Hoon G., "Term Weighting Scheme for Short-Text Classification: Twitter Corpuses," *Journal of Neural Computing and Applications*, vol. 31, pp. 1-13, 2018.
- [4] Al-Qurishi M., Hossain S., Alrubaian M., Rahman S., and Alamri A., "Leveraging Analysis of User Behavior to Identify Malicious Activities in Large-Scale Social Networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 799-813, 2018.
- [5] Antonakaki D., Fragopoulou P., and Ioannidis S., "A Survey of Twitter Research: Data Model, Graph Structure, Sentiment Analysis and Attacks," *Elsevier Journal of Expert Systems* with Applications, vol. 164, pp. 114006, 2021.
- [6] Barushka A. and Hajek P., "Spam Filtering using Integrated Distribution-Based Balancing Approach and Regularized Deep Neural Networks," *Applied Intelligence*, vol. 48, no. 3, pp. 3538-3556, 2018.
- [7] Chen C., Wang Y., Zhang J., Xiang Y., Zhou W., and Min G., "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 4, pp. 914-925, 2017.
- [8] Chen W., Yeo C., Lau C., and Lee B., "A Study on Real-Time Low-Quality Content Detection on Twitter from the Users' Perspective," *PLOS ONE Journal*, vol. 12, no. 8, pp. e0182487, 2017.
- [9] Deng X., Li Y., Weng J., and Zhang J., "Feature Selection for Text Classification: A Review," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 3797-3816, 2019.
- [10] Dutta S., Ghatak S., Dey R., Das A., and Ghosh S., "Attribute Selection for Improving Spam Classification in Online Social Networks: A Rough Set Theory-Based Approach," *Social Network Analysis and Mining*, vol. 8, no. 7, pp. 1-16, 2018.
- [11] El-Alaoui I., Gahi Y., Messoussi R., Chaabi Y., Todoskoff A., and Kobi A., "A Novel Adaptable Approach for Sentiment Analysis on Big Social Data," *Journal of Big Data*, vol. 5, no. 1, pp. 1-18, 2018.
- [12] Fazil M. and Abulaish M., "A Hybrid Approach for Detecting Automated Spammers in Twitter,"

*IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2707-2719, 2018.

- [13] Fiok K., Karwowski W., Gutierrez E., and Wilamowski M., "Analysis of Sentiment in Tweets Addressed to a Single Domain-Specific Twitter Account: Comparison of Model Performance and Explainability of Predictions," *Elsevier Journal on Expert Systems with Applications*, vol. 186, pp. 115771, 2021.
- [14] Gharge S. and Chavan M., "An Integrated Approach for Malicious Tweets Detection Using NLP," in Proceedings of the International Conference on Inventive Communication and Computational Technology, Coimbatore, pp. 435-438, 2017.
- [15] Halawi B., Mourad A., Otrok H., and Damiani E., "Few Are As Good As Many: An Ontology-Based Tweet Spam Detection Approach," *IEEE Access*, vol. 6, pp. 63890-63904, 2018.
- [16] Ji H. and Zhang H., "Analysis on the Content Features and Their Correlation of Web Pages for Spam Detection," *IEEE on China Communications*, vol. 12, no. 3, pp. 84-94, 2015.
- [17] Kiliroor C. and Valliyammai C., "Binary and Continuous Feature Engineering Analysis on Twitter Data Stream for Classification of Spam Messages," in Proceedings of 2<sup>nd</sup> International Conference on Communication, Devices and Computing, pp. 581-594, 2019.
- [18] Lin G., Sun N., Nepal S., Zhang J., Xiang Y., and Hassan H., "Statistical Twitter Spam Detection Demystified: Performance, Stability and Scalability," *Big Data Analytics in Internet of Things and Cyber-Physical Systems*, vol. 5, pp. 11142-11154, 2017.
- [19] Liu C., Wang J., and Lei K., "Detecting Spam Comments Posted in Micro-Blogs Using Self-Extensible Spam Dictionary," *in Proceeding of the International Conference on Communications*, Kuala Lumpur, pp. 1-7, 2016.
- [20] Liu S., Wang Y., Zhang J., Chen C., and Xiang Y., "Addressing the Class Imbalance Problem in Twitter Spam Detection using Ensemble Learning," *Computers and Security*, vol. 69, pp. 35-49, 2016.
- [21] Mohammed N., "Extracting Word Synonyms from Text Using Neural Approaches," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 45-51, 2020.
- [22] Murugan N. and Devi G., "Detecting Streaming of Twitter Spam Using Hybrid Method," *Wireless Personal Communication*, vol. 103, pp. 1353-1374, 2018.
- [23] Rao S., Verma A., and Bhatia T., "A Review on Social Spam Detection: Challenges, Open Issues, and Future Directions," *Elsevier Journal on Expert Systems with Applications*, vol. 186, pp. 115742, 2021.

- [24] Schouten K. and Frasincar F., "Survey on Aspect-Level Sentiment Analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813-830, 2016.
- [25] Shi W. and Xie M., "A Reputation-Based Collaborative Approach for Spam Filtering," *AASRI Procedia*, vol. 5, pp. 220-227, 2013.
- [26] Vanetti M., Binaghi E., Ferrari E., Carminati B., and Carullo M., "A System to Filter Unwanted Messages from OSN User Walls," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 2, pp. 285-297, 2013.
- [27] Xue B., Zhang M., Browne W., and Yao X., "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606-626, 2016.



Valliyammai Chinnaiah received her Ph.D in computer science and engineering at Anna University. She has eighteen years of teaching experience. Her area of interest includes Big Data, Network Management, Grid Computing,

Cloud Computing and Mobile Agents. She has around 77 publications in National and International conferences and journals.



**Cinu C Kiliroor** received her Ph.D in Computer Science and Engineering at Anna University. She has seven years of teaching experience. Her areas of research interests include Machine Learning, Social Network Analysis and Deep Learning.