A Comparative Analyzing of SMS Spam Using Topic Models

Er. Garima Jain

Abstract

Mobile phones or smart phones have changed or revolutionized the way we live. These days the short message service (SMS) is becoming fashionable. For spammers, the success of the mobile messaging channel has become a very attractive target to attack. To impose an additional level of security in the pervasive environment, we will create a system which is more authenticated for SMS. This system will have impact on user's usability from the point of view of user's safety. In modern era, the financial industries and other related agencies are seeing the SMS as an important aspect to communicate with their customers which somehow opens the easy flap for spammers, and customer's safety measures is at hazard. The digital encryption methodologies are useful to support the SMS formation which needs two nodes to swap over digital signed SMS message. These two nodes are protected by the public key cryptography and authentication is done with the help of the ECDSA signature scheme. These two nodes are recognized as sender and receiver, and when a sender sends an SMS to any receiver, the unencrypted text is sent means that there is possibility of loss of information. In this paper, we propose the technique called Gaussian Naive Bayes Classification (GNBC) for the filtering of spam by SMS that solves the message topic model (MTM) problems. It is believed that some pre-processing rules and background terms make it the most appropriate model to completely represent spam by SMS. Finally, we have concluded that GNBC is more accurate for the SMS spam filtering activity.

Keywords

Pervasive computing • Spam filtering tool • MTM (Message Topic Model) • LDA (Latent Dirichlet allocation) • GNBC (Gaussian Naive Bayes classification) • *K*-means

1 Introduction

There are a few different ways to make spam messages like email, SMS/MMS sent tumultuously to your PDA, short code, different remote numbers, and so on. As per the Text Retrieval Conference (TREC), the term "spam" is—a spontaneous, undesirable email that was sent aimlessly (Cormack, 2008). These undesirable and unnecessary spam messages are named as pervasive spam (Spam, 2015). A GSMA pilot spam reporting program, (GSMA Launches SMS Spam Reporting Service, 2011). The improvement of Open Mobile Alliance (OMA) (Efficient Support Vector Machines for Spam Detection: A Survey, 2015) morals for versatile spam revealing. The Internet is responsible for email spam though versatile organization is utilized for SMS spam (Kim et al., 2013; Torabi et al. 2015). The immense volume of spam sends moving through the PC networks effectively affect the memory space of email workers, correspondence transmission capacity, and CPU force and client time. To effectively handle the danger which is presented by email spams, fundamental essential email suppliers, for instance, Gmail, Yahoo mail and Outlook have chipped away at the gathering of various AI (ML) methods which are neural networks in its spam channels. The way that email is an extremely modest method for coming to a great many potential clients fills in as a solid inspiration for novice publicists and direct advertisers (Cranor & Lamacchia, 1998). One potential answer for improving spam characterization calculation is utilizing a spam channel named LingerIG actualized in 2003 out of an email arrangement framework named Linger (Chae et al., 2017).

Er. G. Jain (✉)
Swami Vivekanand Subharti University,
Meerut, UP 250005, India
e-mail: jaingarima2011@gmail.com

P. K. Singh et al. (eds.), *Innovations in Information and Communication Technologies (IICT-2020)*,
Advances in Science, Technology & Innovation,
https://doi.org/10.1007/978-3-030-66218-9_10

1.1 Market Inclinations Resultant in an Increase of SMS Attacks

The SMS takes remained used as dollar-making machine by mobile operators over the years. As per a survey, uncountable SMSs are sent on daily basis. Communication through SMS has its own benefits like all the GSM mobile companies use SMS communication. Nowadays, it is possible to send ringtones, animations, business cards, logos, and WAP configuration settings easily by a SMS which leads to increase the SMS attack by sending malicious malwares along with these setting SMS. The SMS market or mobile messaging market is an extremely beneficial production for mobile operator and is growing speedily (What YOU can do to control cell phone spam, 2012).

As per Fig. 1, people from developed countries like USA, UK, and Japan prefer communication through SMS instead of other mode of contact. Electronically sent messages are at higher risk and can easily be sensed by spammers. Spammers may use these messages to play with user's personal data or may impairment with the users by using theirs premium tariff facilities (Spam News, 2015; Sao & Prashanthi, 2015). In comparison of email spamming, SMS spam has an exponential growth measured up to more than 500% yearly (Benevenuto et al., 2010; GSMA, 2011; Guzella & Caminhas, 2009).

As per Table 1, Cloud mark report states that, in 2019, the SMS spam counting varies from region to region (Text Message (SMS) Spam Reporting, 2012). Asia has highest rate of SMS spam up to 30%, while North America reports the figure up to 1%. The volume of spam emails containing malware and other malicious codes between the fourth quarter of 2019 and first quarter of 2020 is depicted in Fig. 2 (Dada et al., 2019; Fonseca et al., 2016).

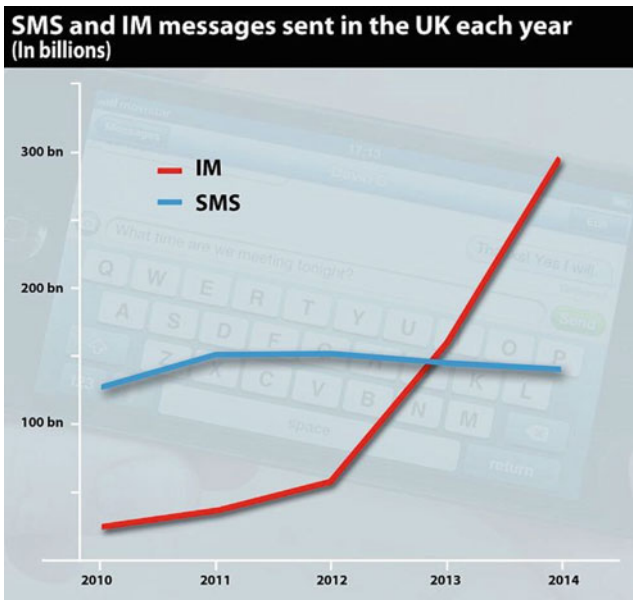

Fig. 1 SMS and IM messages sent in UK each year in billions (Delany et al., 2004)

2 Related Works

In recent approach, SMS spam is a serious security threat in lots of countries which badly destroy the individual privileges and still harm the public safety measures. Pervasive SMS spam filtering can be carried out using various approaches and methodologies on different programming framework. This section aims to analyze previous work which is related to spam detection filtering the spam messages in pervasive environment (Jaswal & Professor, 2013; Satish kumar, 2013; Malarvizhi & Saraswathi, 2013). This section also focused on the motivation and findings through the previous research papers.

2.1 Background

As we have go through the previous research work, we had treasure that mostly email spam clarifying was taken into consideration (Kim et al., 2013; GSMA, 2011; GSMA, 2011a; Sharma et al., 2012; Education & Science, 2013; Johnson et al. 2014) which can be defined as: Operator's server can catch the messages of any frequency triggered from phone numbers. This way has two defects related to pseudo-base stations and possible delay the mass SMS. Users can define black lists or fair lists or secure keywords in their mobile phone at the cost of their own negative impact.

Author (Kou et al., 2020) discuss another most effective method content-based SMS filtering, but it is cost-effective. Finally, MTM is compared with the SVM; the result shows that the MTM is more powerful tool to protect from SMS spam. Author examined about feature determination which is a significant segment in AI and an important advance for text and order. Author You et al. (2020) creator examine and depict a solo technique focusing on astutely distinguishing on the web survey spams. Their investigations on TripAdvisor exhibit the high adequacy and knowledge of the proposed model, which can possibly altogether help the online web business. Author Gopi et al. (2020) in their proposed thought of this paper is to improved RBF piece of SVM-performed with 98.8% of precision when contrasted and the current SVM-RBF classifier and different models. Author Barushka and Hajek (2020) the methodology followed in their paper is to utilize cost-touchy gathering learning strategies with regularized profound neural organizations as base students. Their methodology beats other well-known calculations utilized in interpersonal organization spam sifting, for example, arbitrary woods, Naïve Bayes

Table 1 Graph of mobile behavior shown in developed countries from September 2019 to March 2020 (age group 16+)

	USA	European countries	Japan	India
Messaging (Used)				
Send SMS	72	89.7	62.6	74
Direct messaging	31.2	25.2	6.6	19.9
Financial messaging services				
Bank financial messages	32.4	19	18	21
Share market financial messages	21.2	18	21.5	31

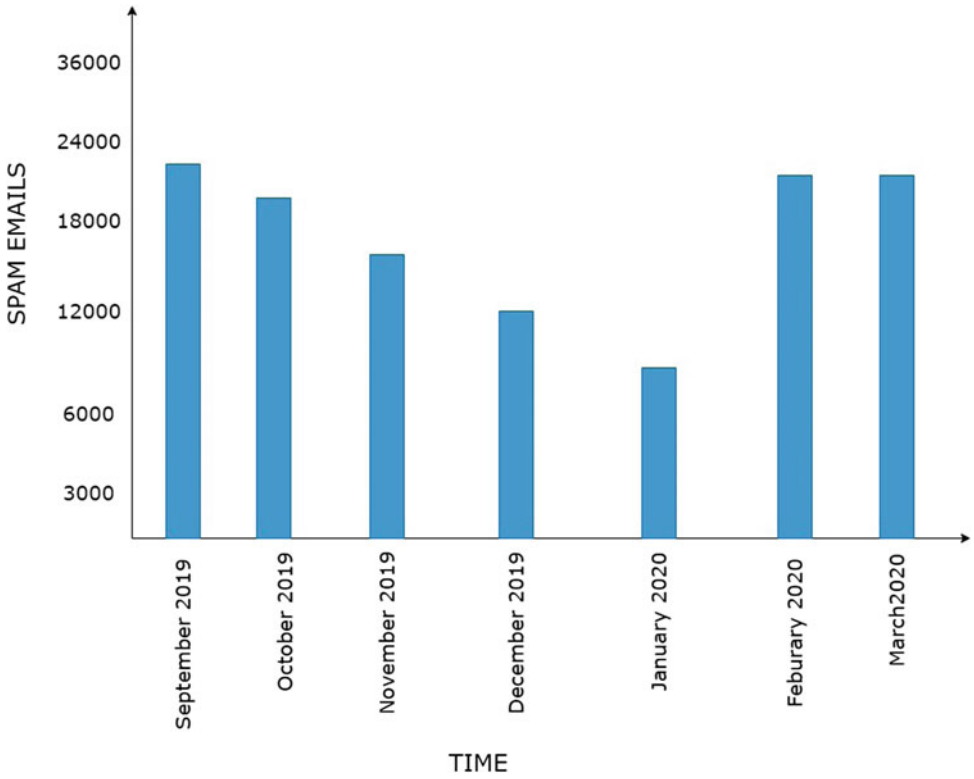

Fig. 2 Volume of spam emails 4th quarter 2019 to 1st quarter 2020

or backing vector machines. Gaurav et al. (2020) their paper proposed a novel, spam mail discovery strategy dependent on the archive naming idea which arranges the new ones into ham or spam. The experimental aftereffects of this paper delineate that RF has higher exactness when contrasted and different strategies. Author Cekik and Uysal (2020) experimental outcomes showed that the PRFS offers either better or serious execution regarding other component determination strategies as far as Macro-F1.

The Author Abayomi Alli et al. (2019) examined investigation that closes with fascinating discoveries which show that most of existing SMS spam separating arrangements are still between the "Proposed" status and "Proposed and Evaluated" status. Likewise, the scientific categorization of existing best in class techniques is created, and it is presumed that 8.23% of Android clients really use this current SMS against spam applications. Their investigation likewise presumes that there is a requirement for specialists to misuse all security strategies and calculation to make sure about SMS consequently improving further characterization in other short message stages. Author Bahassine et al. (2020) their paper give the blend fundamentally improves the presentation of Arabic content order model. The best f-measures got for this model are 90.50%, when the quantity of highlights is 900. Author Asghar et al. (2020) in their paper the work show that joining spam-related highlights with rule-based weighting plan can improve the presentation of even gauge spam location strategy. This improvement can be useful to Opinion Spam recognition frameworks, because of the developing enthusiasm of people and organizations in detaching counterfeit (spam) and certifiable (non-spam) surveys about items. Author Jain et al. (2020) this paper additionally presents a similar examination of various calculations on which the highlights are executed. Furthermore,

it presents the commitment of various highlights in spam identification. After execution and according to the arrangement of highlights chosen, artificial neural network algorithm utilizing back propagation strategy works in the most effective way. Author Bhat et al. (2020) explore and propose two profound neural organization variations (2NN DeepLDA and 3NN DeepLDA) of existing subject displaying method Latent Dirichlet Allocation (LDA) with explicit intend to deal with huge corpuses with less computational endeavors. Two proposed models (2NN DeepLDA and 3NN DeepLDA) are utilized to copy the measurable cycle of latent Dirichlet allocation. Reuters-21578 dataset has been utilized in the examination. Results registered from LDA are contrasted, and the proposed models (2NN DeepLDA and 3NN DeepLDA) utilize Support Vector Machine (SVM) classifier. Proposed models have demonstrated noteworthy exactness other than computational adequacy in contrast with conventional LDA.

3 Message Topic Model (MTM)

The MTM follows the latent semantic analysis enriched by probability. MTM can reduce sparsely problems up to higher extent in comparison to other filtering technologies. The multinomial distributions like document topic or topic-word distributions are governed by several parameters like α and β which show the hyperparameters prior to θ and φ obtained through a Gibbs sampling (Saxena and Payal 2011).

4 Problem Statement

Today, network security is more unpredictable contrasted with before. Huge increment in the quantity of spontaneous business notices being sent to client's cell phones has been watched by means of text informing. SMS spam is an eminent issue for the mobile phone customers. Among the network, the ongoing increments in the spam rate had caused an extraordinary concern. To manage this spam issue, there are numerous methods utilizing diverse sort of spam channels. Essentially, every one of these channels arrange the messages into the classification of spam and non-spam (Ham). The majority of the classifiers choose the destiny of an approaching message based on certain words in information part and sort it. There are two sections, known as test information and preparing information that function as the information base the spam classifier to characterize the messages. The issue of spam has been tended to be as a straightforward two-class record arrangement issue where primary point is to sift through or separate spam from non-spam (Ham). As archive grouping assignments are driven by enormous ineffective information, so choosing most separating highlights for improving exactness is one of the fundamental destinations, and this theory work focuses on this undertaking. The essential point of this work is to focus on various arrangement strategies and to look at their exhibitions on the space of spam message discovery. To lookout, which one is more effective under which set of highlights, various pre-characterized messages are prepared with the strategies. Auxiliary point of the proposition is to progress in the direction of actualizing the strategy which spam can undoubtedly be distinguish with no information superseding and which additionally increment the prior exhibition by finding the best couple of included decrease procedure and characterization calculation. As a huge number of such couples as of now exists, however this work can be considered as a stage toward that objective.

5 Methodology

To achieve all the aims, objectives and overcome the problems as discussed in Sect. 3, various algorithms need to be used for clustering, classification, tokenization and more. In this section, we discussed some of the important algorithms need to be used in this paper or research work. Also the algorithm defines to form some of the associate correct prediction which is a key challenges for facing meteorologist at all planets (https://www.developershome.com/sms/smsIntro.asp; Jain & Mallick, 2016,2017; Failed, 2017). The security algorithm will provide a great level of security in which we have lesser key size as associated with other cryptographic techniques (Jain, 2018).

6 *K*-means

There are several algorithms for data clustering. To achieve simplification, we do clustering which is nothing but the partitioning of data into groups. Although clustering simplifies the dataset, it loses few details. Clustering is not suitable for infinite streams. Working of *K*-means algorithms can be described as *K* input parameter with n set of objects can be partitioned into K clusters in l iterations. The time complexity of *K*-means algorithm is $O(nkl)$.

6.1 Term Frequency-Inverse Document Frequency

Term frequency-inverse document frequency is also known as TF-IDF. To create tokens or categorize documents, text mining techniques like TF-IDF are used. Term frequency can be described as the occurrence of a particular word in an individual document (Bønes, et al. 2007; www.securelist.

com). There is possibility that the same word can be occurred in multiple documents many times TD-IDF uses inverse document frequency which is nothing but the balancing of the occurrence count a particular word.

$$\begin{aligned}\mathrm{TF}(t) &= (\text{Number of times term } t \text{ appears in a document}) \\ &\quad /(\text{Total number of terms in the document}) \\ \mathrm{IDF}(t) &= \log\_\mathrm{e}(\text{Total number of documents} \\ &\quad /\text{Number of documents with term } t \text{ in it}).\end{aligned}$$

$$\text{Value} = \mathrm{TF} * \mathrm{IDF}$$

6.2 GNBC (Gaussian Naive Bayes Classifier)

In our study, we present a classifier named as Gaussian Naive Bayes Classifier (GNBC) which is a combination of Naive Bayes algorithm and Gaussian distance function. The probability theory of semantic analysis has been used for GNBC, and research tells that it is more suitable algorithm for SMS spam filtering. The basic difference between GNBC and MTM is given as follows: The number of tokens is able to find the appropriate status class (SPAM or HAM) due to which identification of spam message is accurate. Number of tokens which are already fixed for spam filter would not cover for the sparse matrix due to which data does not over ride and spam and ham messages can easily be identified.

In Table 2, the Gaussian distribution shown is standardized so that the sum over all values of x gives a probability of 1. Within one standard deviation of the mean, the nature of the Gaussian gives a probability of 0.683. The Gaussian distribution is also termed as "normal distribution" and is often described as "bell-shaped curve." The mean value is $a = yz$, where

$$\begin{aligned}y &= \text{number of events} \\ z &= \text{probability of any integer value of } x\end{aligned}$$

7 Experimental Results

GNBC classifier is applied on the dataset: the large corpus SMS Spam Collection Dataset created by T.A. Almeida et al. By applying classifier on the dataset, the best classifier can be judged as compared with MTM. From the statistics, while applying GNBC and MTM on the dataset, GNBC classifier has the best accuracy than MTM and consumed less time without overriding the data. This paper has incorporated datasets which were specified by T.A. Almeida et al. can be downloaded from https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection. This is accessible online for study sources and is used generally. For SMS Spam research, the SMS Spam collection v.1 (hereafter the corpus) is a set of SMS tagged messages which have been collected. It contains one set of SMS messages.

Table 2 Gaussian distribution function

Distribution	Functional form	Mean	Standard deviation
Gaussian	$f_g(x) = 1/\sqrt{2}\pi\ \sigma^2 e^{\frac{-(x-a)^2}{2\sigma^2}}$	a	ρ

7.1 Evaluation Metrics

The metrics measured the percentage of spam detected by the system and how many misclassifications it makes. Few of the evaluation metrics are:

- True Positive (TP): When positive occurrences are effectively arranged, it is spoken to by a number called genuine positive.
- False Positive (FP): When positive occurrences are mistakenly characterized, it is spoken to by a number called bogus positive.
- False Negative (FN): When negative cases are inaccurately characterized, it is spoken to by a number called bogus negative.
- True Negative (TN): When negative occasions are effectively arranged, it is spoken to by a number called genuine negative.
- Accuracy (ACC): It very well may be characterized as the extent of effectively ordered classes to be specific True Positive and True Negative over the complete number of arrangements.

$$\frac{\text{True Negative} + \text{False Positive}}{\text{True Negative} + \text{False Positive} + \text{True Positive} + \text{False Positive}} * 100 \tag{1}$$

- Precision (P): It is the fraction of the messages retrieved that are related to the end client.

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{2}$$

- Recall (R): It is the fraction of the positively retrieved messages that are related to the client (Figs. 3, 4, 5 and 6).

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

```
     1     2     2     1     2     2     2

  Columns 460 through 486

     1     1     2     1     1     1     1

  Columns 487 through 500

     2     1     2     2     1     1     1


err =

    0.0560

fx >>
```

Fig. 3 Snapshot showing error on command window

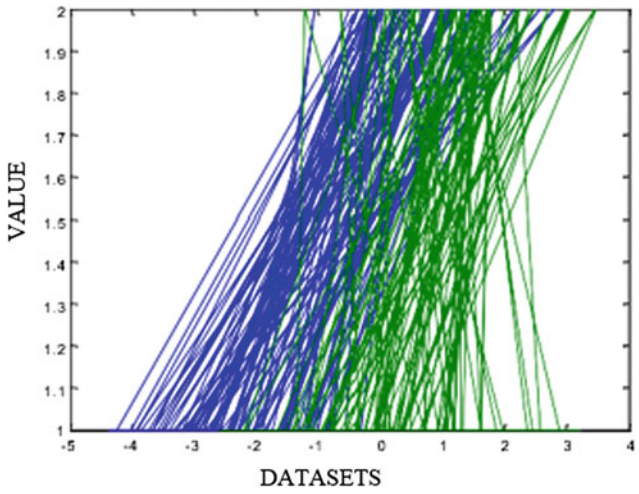

Fig. 4 Data without filtering of ham, spam and noisy data

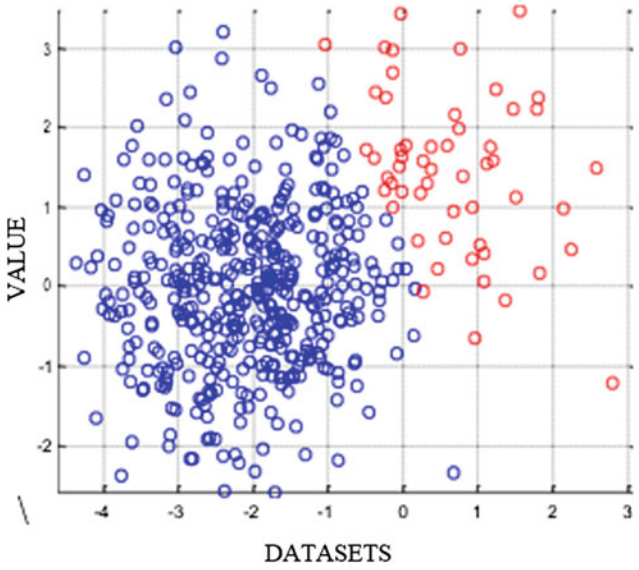

Fig. 5 Data after filtering in which blue color represents noisy data, red color represents spam data

7.2 Experiment 2

Again running a classifier on a separate dataset, it has been shown that Fig. 7 shows the data without filtering which consist of ham, spam and noisy data. Figure 8 shows the data in which red color represented as Spam data, green color represented as Ham data and blue color represented as noisy data. Figure 9 shows the accuracy in MTM and GNBC models, and it has shown that GNBC gives more accurate result when we compared with MTM. Figure 10 shows the error coming on command window.

And the error we had received is 0.0320.

7.3 Experiment 3

Again running a classifier on a separate dataset.

And the error we had received is 0.1340 (Fig. 11).

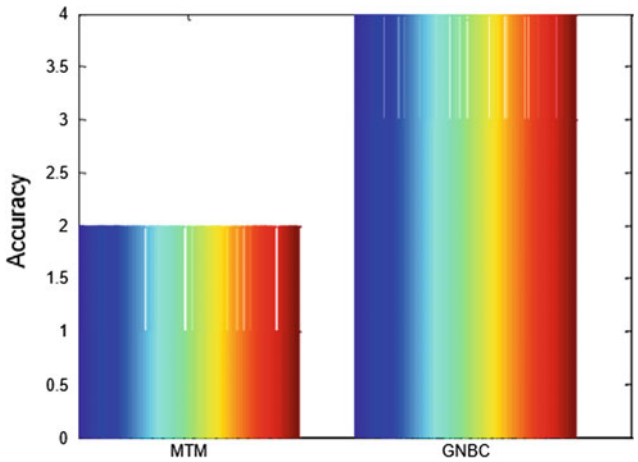

Fig. 6 Comparison between MTM and GNBC, i.e., accuracy

```
     1     2     1     1     2     1     2

  Columns 460 through 486

     2     2     2     1     2     2     2

  Columns 487 through 500

     2     2     2     2     2     2     2


err =

    0.0320

fx >>
```

Fig. 7 Snapshot showing error on command window

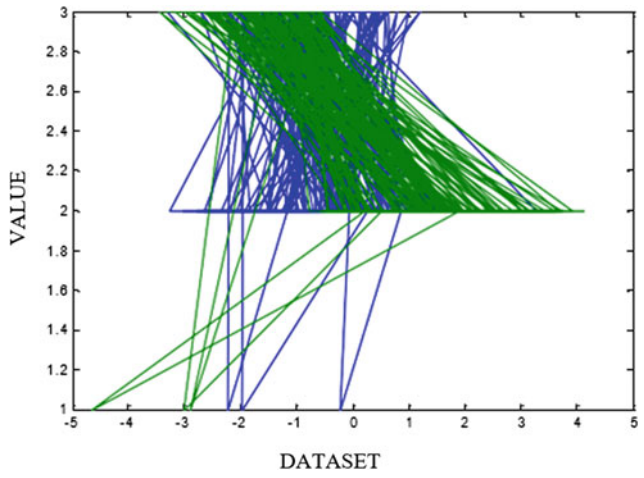

Fig. 8 Data without filtering of ham, spam and noisy data

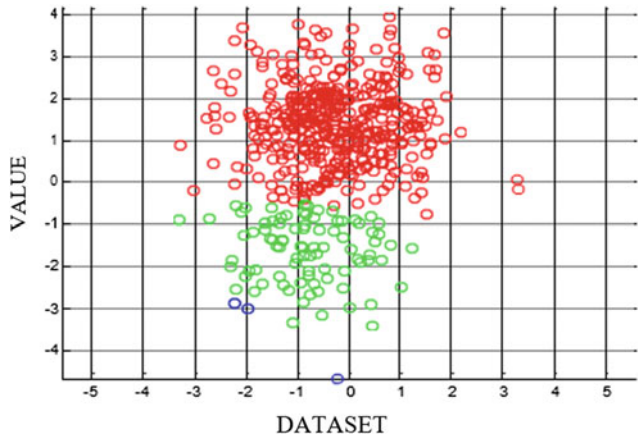

Fig. 9 Data after filtering which blue color represents noisy data, green color represents ham data, red color represents spam data

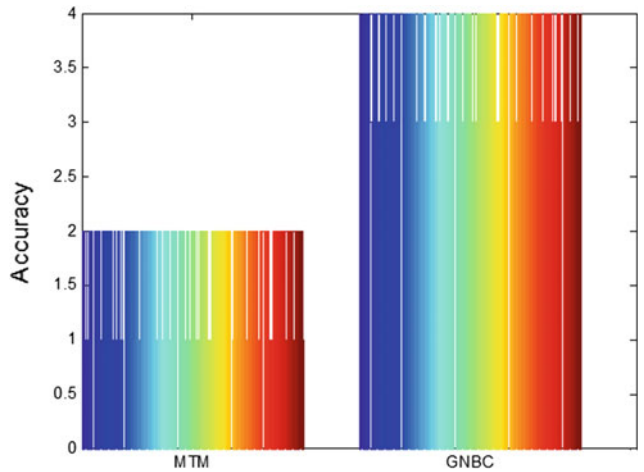

Fig. 10 Comparison between MTM and GNBC, i.e., accuracy

```
     1     2     2     1     1     1     2

  Columns 460 through 486

     2     2     2     2     2     1     2

  Columns 487 through 500

     1     1     2     1     2     1     1

err =

    0.1340

fx >>
```

Fig. 11 Showing error on command window

8 Results and Discussion

On performing experiment on the dataset, trying to classify the data as spam or ham using GNBC and MTM classifier, we observe that GNBC is more accurate as compared with MTM. Also we had observed that when data is filtered by MTM, it is overriding, whereas when we filtered the same by GNBC, it is clearer. In the above experiment, data with red color represented as Spam data, green color represented as Ham data and blue color represented as noisy data. Also we had calculated the error which changes as per the dataset and spam messages.

9 Conclusion

Nowadays, the undertaking of automatic SMS spam clarifying in pervasive environment is stagnant a real task. The major problem handled in detection of spams in SMS is due the total character small in number in short text message and the usual practice of idioms and acronyms. The immediate conclusion from the results is that GNBC has the best performance considering accuracy and overriding of the data. It requires fewer input features to achieve the same results produced by other classifiers. The main aim of our research to refine spam token precisely as compared to existing technique which increase accuracy of the system. Furthermore, our aim will be using classifiers of Gaussian-based NBC to increase the efficiency of spam detection system.

10 Future Scope

The feature plot can increase the aspect of the future work which gives practice in numerous methods. If we are adding more significant features like in the given certain thresholds for the measurement and for the evaluation of the knowledge and the given arcs can also contribute to the development in results. In future, using this technique, we can make and application for smartphones (iPhone, android, Windows) for protecting them from spam message. This will also be compared like we did in DND (Do not disturb) in which we can also block numerous annoying communications, but in the given future we can also make an attempt to block all the communicated messages from given undesirable figures as well as junk bases.

References

Abayomi-Alli, O., Misra, S., Abayomi-Alli, A., & Odusami, M. (2019). A review of soft techniques for SMS spam classification: Methods, approaches and applications. *Engineering Applications of Artificial Intelligence, 1*(86), 197–212.

Asghar, M. Z., Ullah, A., Ahmad, S., & Khan, A. (2020). Opinion spam detection framework using hybrid classification scheme. *Soft Computing, 24*(5), 3475–3498.

Bahassine, S., Madani, A., Al-Sarem, M., & Kissi, M. (2020). Feature selection using an improved Chi-square for Arabic text classification. *Journal of King Saud University-Computer and Information Sciences, 32*(2), 225–231.

Barushka, A., & Hajek, P. (2020). Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks. *Neural Computing and Applications, 32*(9), 4239–4257.

F. Benevenuto, G. M., Rodriques, T., & Almeida, V. (2010). Detecting spammers on Twitter. In: *Proceedings of the 7th Annual Collaboration Electronic Messaging, Anti-Abuse and Spam Conference*.

Bhat, M. R., Kundroo, M. A., Tarray, T. A., & Agarwal, B. (2020). Deep LDA: A new way to topic model. *Journal of Information and Optimization Sciences, 41*(3), 823–834.

Bønes, E., et al. (2007). Risk analysis of information security in a mobile instant messaging and presence system for healthcare. *International Journal of Medical Informatics, 76*(9), 677–687.

Cekik, R., & Uysal, A. K. (2020). A novel filter feature selection method using rough set for short text data. *Expert Systems with Applications, 1*(160), 113691.

Chae, M. K., et al. (2017). Spam filtering email classification (SFECM) using gain and graph mining algorithm. In *2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*. IEEE.

Dada, E. G., et al. (2019). Machine learning for email spam filtering: Review, approaches and open research problems. *Heliyon, 5*(6), e01802.

Delany, S. J., et al. (2004). A case-based technique for tracking concept drift in spam filtering. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. London: Springer.

Torabi, Z. S., Nadimi-Shahraki, M. H., & Nabiollahi, A. (2015). Efficient support vector machines for spam detection: A survey. *International Journal of Computer Science and Information Security, IJCSIS, 13*(1).

Fonseca, O., et al. (2016). Measuring, characterizing, and avoiding spam traffic costs. *IEEE Internet Computing, 20*(4), 16–24.

Garima Jain, E., & Mallick, B. (2017). The weather forecasting using sliding window algorithm. *IJRCCT, 6*(4), 099–105.

Gaurav, D., Tiwari, S. M., Goyal, A., Gandhi, N., & Abraham, A. (2020). Machine intelligence-based algorithms for spam filtering on document labeling. *Soft Computing, 24*(13), 9625–9638.

Gopi, A. P., Jyothi, R. N., Narayana, V. L., & Sandeep, K. S. (2020). Classification of tweets data based on polarity using improved RBF kernel of SVM. *International Journal of Information Technology, 1,* 1–6.

GSMA. (2011a). Operator FAQs. GSMA Spam Reporting Service.

GSMA.(2011b). SMS spam and mobile messaging attacks—Introduction, trends and examples. *GSMA Spam Reporting Service*.

GSMA Launches SMS Spam Reporting Service—PC World Business Center. (2011). PC World. Retrieved January, 13 2011.

Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to spam filtering. *Expert Systems with Applications*, 10206–10222.

https://www.developershome.com/sms/smsIntro.asp

Jain, G. (2018). Time-series analysis for wind speed forecasting. *Malaya Journal of Matematik (MJM), 1*(2018), 55–61.

Jain, G., & Mallick, B. (2016). A review on weather forecasting. *IJARCCE*.

Jain, G., & Mallick, B. (2017). A study of time series models ARIMA and ETS. *IJMECS*.

Jain, A. K., Goel, D., Agarwal, S., Singh, Y., & Bajaj, G. (2020). Predicting spam messages using back propagation neural network. *Wireless Personal Communications, 110*(1), 403–422.

Jaswal, V., & Sood, N, (2013). Spam detection system using Hidden Markov model. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(7).

Johnson, D., Menezes, A., & Vanstone, S. (2014). The elliptic curve digital signature algorithm (ECDSA). *International Journal of Information Security, 1*, 36–63.

Kim, M., Kim, J., & Cheon, J. H. (2013). Compress multiple cipher texts using ElGamal encryption schemes. *Journal of Korean Mathematical Society, 50*, 361–377.

Kou, G., Yang, P., Peng, Y., Xiao, F., Chen, Y., & Alsaadi, F. E. (2020). Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing, 1*(86), 105836.

Kumar, S. (2013). *How to activate do not disturb (DND) India registration—All network*, May 26 2013.

Malarvizhi, R., & Saraswathi, K. (2013). Content-based spam filtering and detection algorithms—An efficient analysis & comparison. *International Journal of Engineering Trends and Technology (IJETT), 4*(9).

Nadimi-Shahraki, M. H., Torabi, Z. S., & Nabiollahi, A. (2015). Using J48 tree partitioning for scalable SVM in spam detection.*Computer and Information Science*, *8*(2), 37.

Spam News. (2015). https://en.wikipedia.org/wiki/Acision,(2015)

Priyanka, S., & Prashanthi, K. (2015). E-mail spam classification using Naïve Bayesian classifier. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4*(6).

Rathi, M., & Pareek, V. (2013). Spam mail detection through data mining—A comparative performance analysis. *International Journal of Modern Education and Computer Science, 12*, 31–39. Published Online December (2013) in MECS (https://www.mecs-press.org/). https://doi.org/10.5815/ijmecs.2013.12.05.

Saidani, N., Adi, K., & Allili, M. S. (2020). A semantic-based classification approach for an enhanced spam detection. *Computers & Security, 94*, 101716.

Saxena, N., & Payal, A. (2011). Enhancing security system of short message service for MCommerce in GSM. *International Journal of*

Computer Science & Engineering Technology (IJCSET), 2(4), 126–133. ISSN: 2229-3345.

Sharma, S., Yadav, J. S., & Sharma, P. (2012). Modified RSA public key cryptosystem using short range natural number algorithm. *International Journal, 2*.

Spam. (2015). https://www.kaspersky.com/about/news/spam.

Text message (SMS) spam reporting. *T-Mobile Support Community*. Retrieved December 8, 2012.

What YOU can do to control cell phone spam (PDF). *AT&T Consumer Guide*. Retrieved December 8, 2012.

www.securelist.com.

You, L., Peng, Q., Xiong, Z., He, D., Qiu, M., & Zhang, X. (2020). Integrating aspect analysis and local outlier factor for intelligent review spam detection. *Future Generation Computer Systems, 1*(102), 163–172.