Dear editor and dear reviewers,
Thank you very much for valuable comments and suggestions. We have updated the manuscript according to the comments below.
On behalf of all the co-authors,
Sincerely,
Syed Md. Minhaz Hossain
1st Author

Reviewer 1:

1. Abstract needs to be concise yet comprehensive reflection of what is in the article. Overall, it is ambiguous.

Thank you, sir for your suggestions. We rewrite our abstract and eradicate some unnecessary and ambiguous lines. Modified lines are highlighted with red color.

Short Message Service (SMS) is becoming the secure medium of communication due to large-scale global coverage, reliability, and power efficiency. As person--to--person (P2P) messaging is less secure than application-to-person (A2P) messaging, anyone can send a message, leading to the attack. Attackers mistreat this opportunity to spread malicious content, perform harmful activities, and abuse other people, commonly known as spam. Moreover, such messages can waste a lot of time, and important messages are sometimes overlooked. As a result, accurate spam detection in SMS and its computational time are burning issues. In this paper, we conduct six different experiments to detect SMS spam from the dataset of 5574 messages using machine learning classifiers such as Multinomial Naïve Bayes (MNB) and Support Vector Machine (SVM), considering variations of Term Frequency-Inverse Document Frequency (TF-IDF) features for exploring the trade-off among accuracy, F1-score and computational time. The experiments achieve the best result of the accuracy of 98.50%, F1--score of 98%, and area under roc curve (AUC) of 0.97 for multinomial naïve bayes classifier with TF-IDF after stemming.

2. The introduction section is not convincing. Try to structure the introduction section with four paragraphs as follows: i) State the motivation and clearly define the problem to be solved. ii) Make a thorough discussion of the state-of-the-art. iii) Describe your proposal in fair context to other published methods highlighting

advantages and disadvantages of these methods. iv) clearly pinpoint the novelty/contribution of your proposal and briefly describe your findings.

Thank you, sir for your comments. We have changed our introduction.

i) State the motivation and clearly define the problem to be solved.

The rising of 5G and cloud technology introduces a new ecosystem that incorporates the connectivity of devices and technologies. SMS is becoming the secure medium of communication for machine--to--machine (M2M) or machine--to--person (M2P), for large-scale global coverage, reliability, power efficiency, and reduction of SMS cost [2]. Besides, SMS has a vital role in making a decision based on the received data in Internet of Things (IoT) [5]. IoT users can get notifications and other alarms from IoT devices through SMS [6]. As the smart devices used in our everyday life activities are mostly directed by internet connectivity, the risk of data privacy or cyber-attacks is increasing day by day [16]. Cyber-attack causes vital infrastructural destruction with massive losses of \$345 per incident [1]. Spamming is an effective way to spread malware through the internet and mobile network. Usually, a mobile user faces a crisis of spamming through SMS. This fact led attackers to use SMS spam for spreading payload of cyber-attacks [3]. Moreover, SMS spam is considered the most straightforward technique to deploy phishing attacks [4]. As a result, security specialists are very devoted to developing an efficient SMS spam detection or classification method.

ii) Make a thorough discussion of the state-of-the-art.

Machine learning techniques such as SVM, decision tree (DT), logistic regression (LR), and MNB play an essential role in detecting anomalies or classifying SMS spam [15,17,18,19,24]. Moreover, these techniques have a vital contribution to detecting email spam messages [20,21]. Several studies have been done on the different types of proposed methods for the filtering of mobile SMS spam [25,26,27,28]. Different feature extraction methods such as word2vec, word n-gram, character n-gram, and combination of variable length n-gram are used to extract features in several works [22,23,31]. Moreover, in spite of having the highest accuracy in machine learning based spam SMS detection [29,30], the ratio of false positive (precision) and false negative (recall) is an issue.

Moreover, in Gupta et al. (2018), a deep learning algorithm is used to detect spam SMS and achieves 99.1\% of accuracy [32]. Using deep learning in detecting spam SMS is computationally expensive and time-consuming. Besides, deep learning models require a large amount of data. Despite having better accuracy in the deep

learning method, it is a burning question of high accuracy with less time. Machine learning techniques consume less time than deep learning based spam SMS detection. Bagging and boosting algorithms achieve better results than traditional machine learning methods [31]. However, there are still limitations of computational complexity and loss of interpretability.

iii) Describe your proposal in fair context to other published methods highlighting advantages and disadvantages of these methods.

Thanks for your suggestions. We have pointed out the limitations of existing system. Despite having high accuracy, having limitation of precision and computational time. To solve this, we investigate the various features and prove that some features like number of lengths do not have effect on filtering SMS. We execute six experiments with different features and classifiers. We use TF-IDF features, it has limitations in similarity of words in documents. However, TF-IDF is significant for its improvement in ratio of false positive (precision) and false negative (recall) for its formulation. In perspective of accuracy, deep learning algorithm and boosting algorithms are better than traditional machine learning algorithms. However, computational cost, interpretability and huge amount of data are factors. Therefore, we choose Multinomial Na¨ıve Bayes (MNB) and SVM using TF-IDF.

iv) clearly pinpoint the novelty/contribution of your proposal and briefly describe your findings.

Thanks, your suggestions. We have point out the contributions of our work with the trade-off among computational time, accuracy and F1 score using investigations of different features.

3. "… security is a significant threat to…" does it make sense? Also, "Not only has the number of internet users increased but so has the number of new malware". Paper needs a comprehensive proof-reading to fix English.

Thanks sir for your suggestions. These two lines are eradicated from this paper.

4. The use of SMS has been increased drastically due to the reduction of SMS cost. This kind of information should have a proper citation.

Thank you, sir, for your suggestion. We have placed a reference with this line.

SMS is becoming the secure medium of communication for machine--to--machine (M2M) or machine--to--person (M2P), for large-scale global coverage, reliability, power efficiency and reduction of SMS cost [2].

5. References need to be proper formatting. Authors may follow IEEE manuscript preparation guidelines for fixing.

Thank you, sir, for your comment. We fix this and format the references as depict.

Reviewer 2:

1. Why did the authors use SVM and MNB instead of other powerful algorithms such as CNN and XGBoost? Clarify it?

Thank you, sir, for your query. Deep learning algorithm is used to detect spam SMS and achieves 99.1% of accuracy (32). Using deep learning in detecting spam SMS is computationally expensive and time-consuming. Despite having better accuracy in deep learning method, it is burning question of high accuracy with less time. Machine learning techniques consumes less time than deep learning-based spam SMS detection. Bagging and boosting algorithms achieve better results than traditional machine learning methods. However, there are still limitations of computational complexity and loss of interpretability (31). As, our contribution is to trade-off accuracy, F1 score (ratio of false positive and false negative) and computational time, and having less amount of SMS, we prefer SVM and MNB for filtering SMS.

2. Why didn't authors consider AUC, SPE, MCC as an evaluation metric for checking the performance of their method?

Thank you, sir, for your query. We have considered AUC for evaluating our models performance. To evaluate the results further, we have drawn the ROC curves for both cases. The micro average area under the ROC curve (AUC) for SVM is 0.93, and for MNB is 0.97. Figure 6 shows the area under the ROC curve (AUC)for MNB classifier. It implies that MNB is more robust to filter the spam in SMS with stemming and the TF-IDF process