ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 11, November 2024

An SMS Spam Detection Meta Classifier Model Using MultinomialNB - LinearSVC Algorithms

[1] K. Egneswari, [2] Vaishnavi Priya N, [3] Marieshwari P
[1] Department Of Computer Science And Engineering, Er. Perumal Manimekalai College Of Engineering, Hosur, India.
[1] egneshwari1112@gmail.com, [2]vaaishu1990@gmail.com, [3]marieswari1181987@gmail.com

Abstract: ***A new means of communication known as short messaging services (SMS) has evolved alongside the proliferation of mobile devices, networks, and data transmission in the last several decades. Even SMS users face the issue of spam. Bulk texts or SMS spam is any unimportant message sent through a mobile network [2]. There are a lot of causes for the excess of spam messages, The sheer number of mobile phone users raises the stakes when it comes to spam, or unsolicited bulk messages [1]. Also, the attackers could be ecstatic to hear that sending spam is cheap. In particular, there are a number of well-established algorithms for spam identification, which is a very active area of research [15]. This approach investigates a Multinomial Naive Bayes-Linear SVC methodology [4] to accurately detect spam data or communications. In order to remove irrelevant or inappropriate characters and information from the input dataset, pre-processing is performed [5]. The model is trained using a Multinomial Naive Bayes-Linear SVC approach for spam message prediction [10]. At 98.38% accuracy, the Multinomial Naive Bayes-Linear SVC model outperforms previous models in spam detection, including LSTM, SVM, and naive bayes.***

Keywords— Hash Vectorizer, Deep Learning Algorithms, LSTM, Naïve Bayes, SMS Spam, SVM.

I. INTRODUCTION

Mobile messaging is a form of inter-person communication, and billions of individuals use mobile devices. However, the absence of appropriate message-filtering methods makes this sort of communication insecure. Spam contributes to this risk by making mobile SMS communication insecure. Spam emails or messages are delivered to users without their knowledge and are unwanted by the receivers. Nowadays, spam makes up more than 85% of all emails and communications that people receive. The sender spends very little time on the transaction, but the recipient or service providers are responsible for the majority of the expenses. Spam SMS fills up the inbox with unwanted messages, using up network capacity and storage space while favoring the rapid spread of malicious code and false information. Social media, private networks, and public networks all contain spam. The user's private and sensitive information may be leaked or destroyed. A person that sends unwanted messages is known as a spammer. Spammers attack people with numerous messages for their own personal or professional gain. Due to the tremendous growth in SMS usage, it is now relatively simple for spammers and hackers to attack a user's mobile device by just sending them a malicious link. If the user

clicks the link or messages the hacker or spammer has sent, it will be automatically detected. Even though there are different kinds of techniques available but still there is a need to handle these techniques in an advanced way. Most real-world problems in all industries can now be solved using machine learning and deep learning. Computers can now learn from the past and make predictions because of technology called deep learning. To overcome these two levels of filtering are done, the first level of filtering is done by removing stop words and then by the correlational coefficient. To label the messages as spam or ham, the refined dataset is translated to the classifiers Naïve Bayes, Long Short-Term Memory (LSTM), Support Vector Machine (SVM), and Multinomial Naive Bayes-Linear SVC Model.

II. LITERATURE SURVEY

GHADA AL-RAWASHDEH et.al. [1] is about a new technique called water cycle feature selection (WCFS) Simulated annealing is employed for the feature selection along with three different hybridization methods and meta-heuristic optimization. for choosing optimal solutions. Here they have used a cross-validation technique for training and validating the datasets. The authors used seven different datasets for testing spam detection. For the classification, they have used many classifiers such as Support Vector Machine (SVM), Naïve Bayes, and K Nearest Neighbour (KNN), for training and testing the data. Even used accuracy and f-measurement for the exact output. The hybridization works by considering initial data as raindrops, the best raindrop is considered as a sea. The remainder of the raindrops are classed as streams that run into the sea or rivers, and a selection of good raindrops that goes under feature selection are combined to create a river. The above

process is a traditional WCA algorithm that used the initial state and had some issues finding optimal features, so they have employed the hybrid WCA – SA (Simulated Annealing) algorithm to eliminate the issues of WCA. The 3 ways of hybridization which are low-level hybridization, interleaved hybridization, and high-level hybridization surpassed other hybridization feature selection techniques with an accuracy of 96.3%. As a result, there are now less than half as many characteristics, and content classification may be done better by utilising all of the best features produced by the layered hybridization of WCA and SA. The limitation of this research is the accuracy achieved here is 96.3% which is not optimal for accurate spam detection and the process which is used here is lengthy and time taking compared to normal detection of spam messages.

Sjarif, N.N et.al. [2] is about the attack of SMS spam increasing daily SMS is a way to send text messages to one another. Feature Extraction, Pre- processing, selection, and classification are steps in the process of detecting spam. Pre-processing the data is where the unstructured data is converted into structured data. Term Frequency Inverse Document Frequency (TFIDF) is used for the feature extraction and selection phase. TFIDF is the vector space model's often weighting mechanism. Finally for the classification of messages as ham or spam, Random Forest (RF)algorithm has been used. The ensemble learning method developed by RF to average data can be applied to classification problems. To solve the overfitting issue in decision trees, this technique mixes multiple decision tree models. With the help of the RF algorithm, each tree can produce unique prediction outcomes as each tree performs differently, it is necessary to generalize and get the average of performances. The limitations are the RF algorithm may become very slow and ineffective for real-time predictions as there are a lot of trees. The classifiers may perform better and train data more effectively if additional features, such as message lengths, are added.
Saidani, N et.al. [3] provide a technique of two semantic-level analyses. To classify emails in the first level according to certain domains (such as health, education, finance, etc.). For spam identification in each domain, we integrate a group of explicitly provided and automatically retrieved semantic features at the second level. These features give a detailed description of each domain of spam, making it possible to target the detection of this spam more effectively. And then classifiers are used for separating the messages as ham or spam messages, techniques used in this proposed method are Naïve Bayes, Decision tree, and K-Nearest neighbour. The limitation is that the decision tree is largely unstable and KNN with large data the protection phase might be slow.

Gauri Jain et.al. [4] proposed a specialized approach called Recurrent Neural Networks (RNN) which are a particularly efficient deep learning architecture. It makes use of a unique kind of RNN called Long Short-Term Memory (LSTM), which does exceptionally well in classifying spam. It can learn intricate patterns, unlike conventional classifiers that require hand-crafted features. When using LSTM, information can be stored in multiple layers and is dependent on earlier levels. Prior to running LSTM, the text must be converted into word vectors to increase the efficiency of learning complex patterns. Analysing the data demonstrates that LSTM can identify spam considerably more precisely than any traditional machine-learning method. The limitations of this research are the Slang and acronyms are frequently used in data, and proper grammar rules are not always followed. The grammatical rules are not observed.

Pavas Navaney et.al. [5] authors implemented different types of supervised machine-learning algorithms for detecting the ham and spam messages including maximal Entropy, Support Vector Machine (SVM), and Naive Bayes algorithm which contradicts how well they work at filtering spam and junk mail. SVMs partition data into subsets of related components using a linear boundary known as a hyperplane, often as suggested by the class values. Before training the model for machine learning techniques they firstly remove the data of all digits, punctuation, and blank spaces then tokenize messages using Document Term Matrix. It is concluded that the SVM model performs better than the naive Bayes model, classifying spam with 96.4% accuracy and ham with 98.4% accuracy for a total accuracy of 97.4%. The limitations of this research are, it is unsuitable for large amounts of datasets, takes more time for training the data, and gives a bad performance on highly noisy data.

S. Mishra et.al. [6] about smishing is a fraudulent, sending SMS to mobile devices in order to obtain personal information, so to prevent this smishing detector is introduced to detect the smishing attacks. The smishing detector combines methods for URL inspection and SMS content analysis to distinguish between legitimate and spam communications. The message's contents are examined using an SMS Content Analyzer, and the URL, source code, and APK download behaviours are examined using URL Filter, Source Code Analyzer, and APK Download Detector. Messages are categorized using machine learning algorithms based on smishing keywords and the Naive Bayes classifier shows the best accuracy for the classifications of the keywords. The limitations of this research are that the APK Download Detector module lacks security in validating the legitimacy of the application downloaded and all the attributes are assumed to be independent of one another by Naive Bayes.

Gustavo Jose de Sousa et.al. [8] is about One of the unsupervised learning methods used to discover the words that are most relevant to a given word is skip-gram. Text embedding technique for SMS messages that make use of patterns relevant for spam characterization. As a result, spam and non-spam messages are clearly projected in the embedding area, making categorization more accurate. Each SMS message from a training dataset is transformed into a series of sets of characteristics through the process of tokenization. The resulting sequences are then subjected to a skip-gram pattern mining technique, which locates patterns pertinent to the classification objective. An embedding model is developed to map text messages into a vector space given a set of pertinent patterns. Using the generated vectors from the UCI Spam Collection dataset, and evaluate the proposed method. The limitations of this research are this model fails to identify the combined word phrases and it is less sensitive to overfit frequent words.

e Tian Xia et.al. [9] is about the Discrete hidden Markov model (HMM) to apply the word order data and fix the low-frequency SMS spam detection problem. The machine learning repository's SMS spam dataset is performed to examine the proposed HMM method's performance. By using CNN and LSTM models, the total performance is suitable with deep learning. Many informal words, short and abbreviated words, social media acronyms, and even odd character sequences frequently emerge in SMS, especially for SMS. After the pre-processing, Each SMS has been condensed into a meaningful word sequence. HMM is a simple machine-learning model It can be used to handle the high throughput requirements of the spam filtering sector. Due of the extremely low word frequency, standard feature extraction algorithms like TFIDF do not perform well for SMS spam detection. The limitations of this research are it becomes very complicated when more states and more interactions among states are included.

III. METHODOLOGY

At present, all grievances should be raised in the grievance cell, where a separate grievance needs to be allocated for every department in the city. There is a need for a lot of resources to maintain them, and it is very hard to listen to and take one after another.
Spam SMS is a significant factor in the explosive growth of cybercrime. Spam detection systems are used to prevent cybercrimes. Multiple machine-learning techniques are used to recognise spam texts, but the outcomes are not entirely precise. For the detection of spam messages, a Multinomial Naive Bayes-Linear SVC methodology combining naive bayes and SVM algorithms is employed to achieve improved accuracy and precision.

IV. PROPOSED SYSTEM

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 11, November 2024

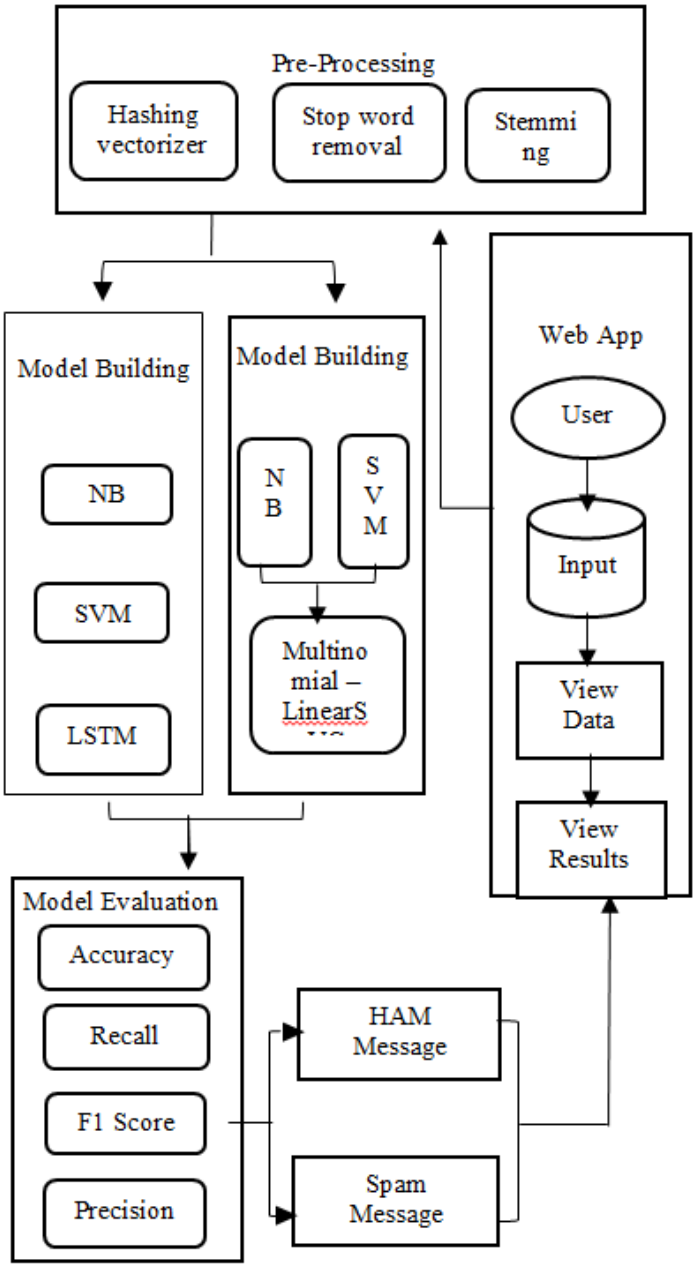


Fig. 1. Proposed System

Various deep learning and machine learning methods have been proposed, however, none have been able to categorize spam messages effectively.
Furthermore, the various types and a large amount of data are typically not taken into account in comparable research which has suggested methods for evaluation of spam messages classification. Using a dataset as an input and performing the following five stages to identify spam messages.

i. Taking an SMS spam collection dataset from Kaggle.com as an input

ii. Replacing missing values and Null values with a numerical value or a categorical value.

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 11, November 2024

iii. The dataset is subjected to stemming, which separates the sentences into words and eliminates white spaces. The special characters are all removed using the stop words removal approach, and the words are then converted to numerals using a hash vectorizer from NLP.

iv. Model training is done by applying a Multinomial Naive Bayes-Linear SVC methodology combining Naive Bayes and SVM algorithms.

The data will be trained as (m*n) during the model training phase. The dataset was then applied to 2 models, with the results serving as a new training set for second-level training which uses a meta-classifier.

Fig. 2. Multinomial Naive Bayes-Linear SVC Model

This kind of training is used to predict data from a first-level model, and the procedure results in a second-level model. Using these findings, a final prediction is created with the highest degree of accuracy.

v. Accuracy, Precision, Recall, and F1 score are used to evaluate the model, and it is then compared to various deep learning and machine learning methods like Naive Bayes, SVM, and LSTM.

vi. The final stage is prediction of SMS as spam or ham.

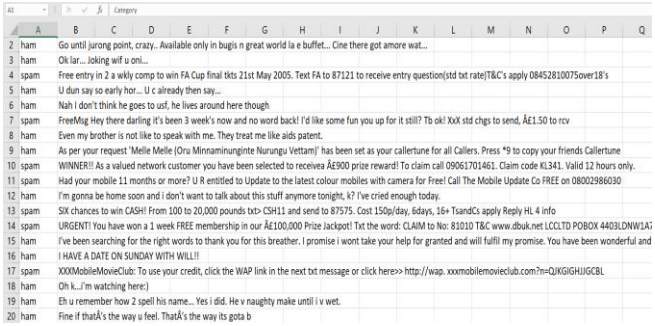
Algorithm Pseudo-Code of Meta Classifier for Spam Detection

```
input Df Dataset split into t1, t2
begin
// t1= training data
// t2 = testing data
M=[MNB]
for i in M do
        var->i
        res1=var.fit(t1)
end for
N=[LSVC]
for i in M do
        var->i
        res2=var.fit(res)
end for
//Meta classifier
M=[MNB]
For i in M do
      var->i
       var.fit (res1, res2)
end for
/MNB-Multinomial Naive Bayes
//LSVC -Linear Support Vector Classifier
end
```

V. Results

The dataset consists of different types of text based on the messages received by the user. Below is a data collected from SMS spam collection dataset.

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 11, November 2024



	A	B
2	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
3	ham	Ok lar... Joking wif u oni...
4	spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
5	ham	U dun say so early hor... U c already then say...
6	ham	Nah I don't think he goes to usf, he lives around here though
7	spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, Â£1.50 to rcv
8	ham	Even my brother is not like to speak with me. They treat me like aids patent.
9	ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
10	spam	WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
11	spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
12	ham	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.
13	spam	SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info
14	spam	URGENT! You have won a 1 week FREE membership in our Â£100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7
15	ham	I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and
16	ham	I HAVE A DATE ON SUNDAY WITH WILL!!
17	spam	XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL
18	ham	Oh k...i'm watching here:)
19	ham	Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.
20	ham	Fine if thatÂ's the way u feel. ThatÂ's the way its gota b

Fig. 3. Spam Dataset

Using the review dataset, the classification report of four algorithm models

Table 1

classification report of Accuracy, F1 score, Precision and Recall

ALGORITHM	Accuracy	F1 Score	Precision	Recall
NB	96.54%	90	85	95
SVM	97.48%	89	82	98
MultinomialNB - LinearSVC	98.38%	93	93	96
LSTM	86.28%	--	--	--

The classification report for the four models can be found in the above table along with their corresponding accuracy, F1 score, precision, and recall values in which Multinomial Naive Bayes-Linear SVC has a higher level of accuracy. It can provide comprehensive details on classification reports on metric values.

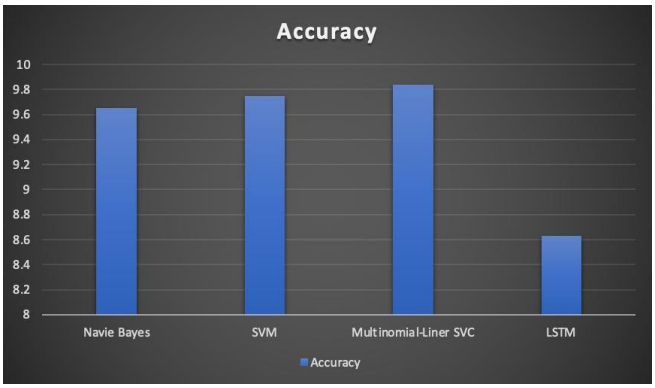


Fig. 4. Comparison of Accuracies of models

The above figure shows the accuracy of four algorithms in which Multinomial Naive Bayes -Linear SVC has highest accuracy compared to the remaining models

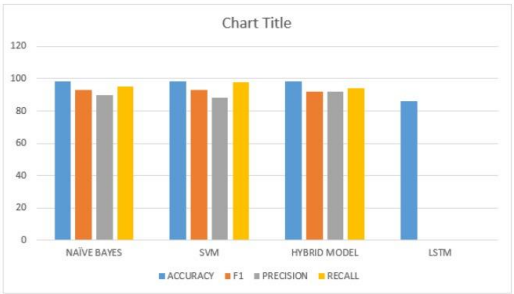


Fig. 5. Comparison of performance metrics

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 11, November 2024

The above graph, shows the comparison of four algorithms based on the metrics such as accuracy, precision, recall and F1 score. In this bar plot the Multinomial Naive Bayes-Linear SVC model has highest accuracy with a range of 98.38%.

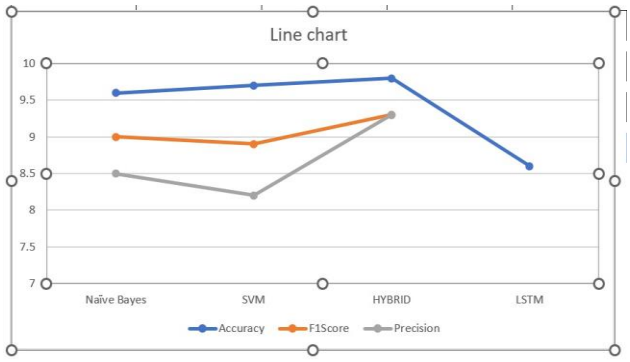


Fig. 6. Trained and Test accuracy, F1 score and precision results

The above graph shows that the Multinomial Naive Bayes-Linear SVC model performs better than other models both in terms of F1 score, accuracy, and precision. However, the precision and F1 Score of the Multinomial Naive Bayes-Linear SVC model have the same values.

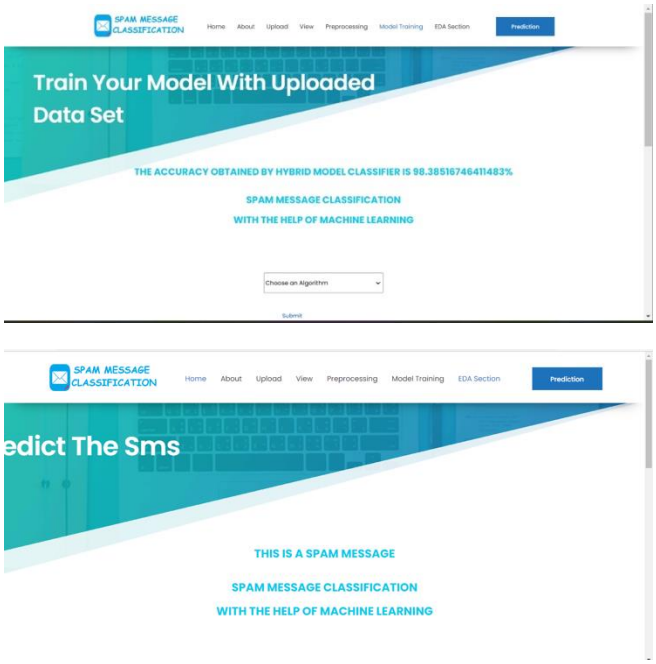


Fig. 7. Spam message detected by the system

The accuracy of each model was assessed during model training, and the Multinomial Naive Bayes-Linear SVC model was found to have the highest accuracy. This model has been used to predict if messages are spam or ham Messages The last step is to predict the model; during this step, a user's random message is used to determine whether or not it is spam.

VI. CONCLUSION AND FUTURE WORK

The SMS spam detection was developed to identify spam messages with the help of different classifiers. In the proposed model, Pre-processing is done to remove the extraneous data before we extract the most pertinent features from the spam dataset, which contains over 5000 messages. The Filtered Feature set classifies messages as ham and spam using a classification algorithm and model evaluation. On the basis of the data gathered, SMS spam is categorised using Naive Bayes (NB), long short-term memory (LSTM) and support vector machines (SVM). The proposed method's experimental evaluation revealed that the Multinomial Naive Bayes-Linear SVC model outperforms it in terms of classifying SMS spam.

The Multinomial Naive Bayes-Linear SVC model had an F1-Score of 91.84%, precision of 95.39%, accuracy of 98.37%, and recall of 87.87% according to the trial data, and as a result, Multinomial Naive Bayes-Linear SVC is a method that produces the best results for categorising spam messages. This solution can considerably improve mobile phone security by decreasing the risks associated with smishing attacks in mobile environments and screening spam messages.

In future, In order to more accurately filter spam messages in smartphones, we intend to build a robust framework. The objective is to introduce new features, like the ability to examine phone numbers found in messages and to analyse URLs or files that are sent along with messages, as well as to deal with more difficult issues, such the management and analysis of report data in the storage of spam SMS filters. Future effort will also centre on finding a solution to this issue.

References

[1]. Ghada al-Rawashdeh,Rabiel Mamat,and Noor Hafhizah Bintiabd Rahim, "Hybrid Water Cycle Optimization Algorithm With Simulated Annealing for Spam E-mail Detection", IEEE Transactions on Knowledge and Data Engineering, Volume 7, pp. 143721 – 14373, 26 September 2019, DOI:10.1109/ACCESS.2019.2944089.

[2]. Amir Sjarif, N. N., Mohd Azmi, N. F., Chuprat, S., Sarkan, H. M., Yahya, Y., & Sam, S. M. (2019). "SMS Spam Message Detection using Term Frequency-Inverse Document Frequency and Random Forest Algorithm". Procedia Computer Science, 161, 509–515. doi:10.1016/j.procs.2019.11.15

[3]. Saidani, N., Adi, K., & Allili, M. S. (2020). "A Semantic-Based Classification Approach for an Enhanced Spam Detection". Computers & Security, 101716. doi:10.1016/j.cose.2020.10171

[4]. Gauri Jain, Manisha Sharma, Basant Agarwal," Optimizing semantic LSTM for spam detection.", in International Journal of Information Technology, Volume 11, pp. 239-250, 12 April 2019, DOI: https://doi.org/10.1007/s41870-018-0157-5

[5]. Pavas Navaney, Gaurav Dubey, Ajay Rana," SMS Spam Filtering using Supervised Machine Learning Algorithms.", in 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Volume 13, pp. 112-124, 23 August 2018, DOI: 10.1109/2018.

[6]. Mishra, S., & Soni, D. (2020)." Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis". Future Generation Computer Systems, 108, 803–815. doi:10.1016/j.future.2020.03.0

[7]. A. Ghourabi, M. A. Mahmood, and Q. M. Alzubi, ''A hybrid CNN-LSTM model for SMS spam detection in arabic and English messages.'', in future internet, vol. 12, pp. 156, 8 September 2020, DOI: doi.org/10.3390/fi12090156.

[8]. Gustavo Jose de Sousa and Ivan Rizzo Guilherme, "SMS Spam Detection Through Skip-gram Embeddings and Shallow Networks", in research gate, vol. 10, pp. 4193-4201, January 2021, DOI 10.18653/v1/2021.findings-acl.367.

[9]. Tian Xia and Xuemin Chen, "A Discrete Hidden Markov Model for SMS Spam Detection", in applied science, vol. 10, pp. 5011, July 2020, DOI: 10.3390/app10145011.

[10]. Joe, I., & Shim, H. (2010). "An SMS Spam Filtering System Using Support Vector Machine". Lecture Notes in Computer Science, 577–584. doi:10.1007/978-3-642-17569-5_56

[11]. Duan, L., Li, N., & Huang, L. (2009). "A New Spam Short Message Classification". 2009 First International Workshop on Education Technology and Computer Science. doi:10.1109/etcs.2009.299

[12]. Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011). "Contributions to the study of SMS spam filtering". Proceedings of the 11th ACM Symposium on Document Engineering - DocEng '11. doi:10.1145/2034691.2034742

[13]. Becker, B. G. (n.d.). "Visualizing decision table classifiers. Proceedings" IEEE Symposium on Information Visualization (Cat. No.98TB100258). doi:10.1109/infvis.1998.729565

[14]. Jain, G., Sharma, M., & Agarwal, B. (2018). "Optimizing semantic LSTM for spam detection". International Journal of Information Technology. doi:10.1007/s41870-018-0157-5

[15]. Bosaeed, S., Katib, I., & Mehmood, R. (2020). "A Fog-Augmented Machine Learning based SMS Spam Detection and Classification System". 2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC). doi:10.1109/fmec49853.2020.9144833