SMS SPAM FILTERING USING MACHINE LEARNING AND DEEP LEARNING

Krish Goel
Computer Science and Engineering (SCOPE)
Vellore Institute of Technology (VIT University)
Chennai, India
krish.goel2021@vitstudent.ac.in

Meghna Das
Computer Science and Engineering (SCOPE)
Vellore Institute of Technology (VIT University)
Chennai, India
meghna.das2021@vitstudent.ac.in

***Abstract*— Due to the ubiquity of mobile phones and the extensive usage of Short Message Service (SMS), SMS spam has become a major annoyance to smartphone users [1]. In this research, we propose a complete solution that combines state-of- the-art deep learning techniques with standard machine learning approaches to filter SMS spam. To extract pertinent features from SMS texts, such as lexical, syntactic, and semantic properties, we first investigate feature engineering and selection strategies [2]. Then, using these features, traditional machine learning models like Random Forests, Naive Bayes, and Support Vector Machines (SVM) are trained [3]. Our suggested approach offers a scalable and effective way to filter SMS spam, giving customers better defense against unsolicited messages while minimizing false positives and maximizing classification accuracy. This work advances the field of spam detection systems and has important ramifications for improving security and user experience in mobile networks.**

Keywords — ubiquity, SMS (Short Message Service), spam, smartphone, state-of-the-art, deep learning, machine learning, unsolicited, false positives, ramifications.

I. Introduction

A rising number of people now consider their mobile phones to be their constant companions [4]. The commercial SMS (Short Message Service) market is now worth millions of dollars. As of early 2013, its value ranged from 11.3 to 24.7 percent of the Gross National Income (GNI) of developing nations [5]. Short messaging service providers (SMS) enable users of mobile phones to send and receive brief text messages. Unsolicited commercial adverts rose in volume as the site gained prominence [6]. The proliferation of mobile devices used for email and messaging has led to an upsurge in spam. Currently, mobile consumers receive 85% of their emails and texts as spam. The cost of sending emails and communications is quite low, but the cost of receiving them is significant. Spam and service provider costs can be quantified by the amount of time lost and essential messages or emails that are lost [7]. Because each user has limited time, memory, and internet access, the value of emails and messages.

Email and text message spam filtering are very different in several key aspects. Unlike emails, which have many large datasets available, there are very few real databases for SMS spam [8]. Furthermore, because text messages are shorter than emails, there is far less information in them that can be used for classification. There isn’t a header here either. Furthermore, text messages are filled with acronyms and significantly less formal language than emails. All these features have the potential to severely reduce the efficacy of widely used email spam filtering algorithms when applied to short text messages [9].

Predicting SMS spam has long been a prominent area of research. The goal is to apply various machine learning algorithms to the problem of SMS spam classification, assess each one’s efficacy to gain insight and further explore the matter, and develop an application built on an algorithm that can reliably filter SMS spam [10]. In the current work, several machine learning and deep learning-based predictive models are proposed to accurately estimate the movement of SMS spam. To further increase the models’ capacity for prediction, the predictive framework incorporates a potent deep learning- based long- and short-term memory (LSTM) network [11].

II. Literature Review

1. Spam Detection Approach for Secure Mobile Message Communication

- Author: Shah Nazir, Habib Ullah Khan and Amin-Ul-Haq
- Year: 2020
- They proposed the applications of the machine learning based spam detection method for accurate detection. They also used Logistic Regression, K-Nearest Neighbor, and Decision Tree for the classification of ham and spam messages. Their LR model had an accuracy of 99% .

2. Spam Filtering in SMS using Recurrent Neural Networks

- Author: Pumrapee Poomka, Wappana Pongsena, Nittaya Kerdprasop, and Kittisak Kerdprasop
- Year: 2019
- They used Natural Language Processing (NLP) and Deep Learning (DL) techniques. The model built using LSTM has an accuracy of 98.18%.

3. SMS Spam Filtering: A Hybrid Approach Using Machine Learning and Lexical Analysis

- Author: R. Vijayalakshmi, Dr. T. Meyyappan
- Year: 2019
- In order to filter SMS spam, this research suggests a hybrid method that combines lexical analysis and machine learning techniques. The research leverages the advantages of both methodologies to improve spam detection's precision and resilience. The efficacy of the hybrid strategy is demonstrated by experimental findings on a large SMS dataset, which demonstrate considerable gains in

classification performance compared to individual techniques.

4. An Effective SMS Spam Filtering Technique Using Machine Learning Approaches

- Author: S. Saranya, R. Saranya, Dr. V. Saravanan
- Year: 2018
- This study presents a machine learning based method for efficiently filtering spam SMS messages. The study evaluates the accuracy of spam message identification using Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbours (KNN) classifiers. The superiority of SVM in terms of accuracy and efficiency is demonstrated by experimental results.

5. SMS Spam Detection Using Deep Learning Techniques

- Author: S. Tiwari, A. Khare
- Year: 2018
- The use of deep learning algorithms for SMS spam detection is investigated in this research. The study contrasts the performance of the deep learning architecture with that of conventional machine learning methods, based on Bidirectional Long Short-Term Memory (Bi-LSTM) networks. The outcomes of the experiments show that the Bi-LSTM model is more effective than other models at correctly classifying spam messages, suggesting that it could be used in practical SMS spam filtering scenarios.

6. SMS Spam Detection Using Machine Learning Techniques: A Case Study

- Author: A. Joshi, V. Sharma
- Year: 2018
- This research uses machine learning approaches to offer a case study of SMS spam identification. Using a sizable SMS dataset, the study evaluates the effectiveness of classifiers including Random Forest, Decision Trees, and Logistic Regression. The outcomes of the experiments show that Random Forest can achieve excellent recall and accuracy rates, which makes it suitable for real-world SMS spam filtering applications.

7. SMS Spam Detection Using Machine Learning Techniques: A Comparative Study

- Author: R. Gupta, S. Shukla, A. K. Gupta
- Year: 2017
- A comparative analysis of machine learning methods for SMS spam detection is presented in this research. Using a sizable SMS dataset, the study assesses the effectiveness of classifiers including SVM, Random Forest, and Gradient Boosting Machines (GBM). Based on experimental results, SVM is a viable method for real-world SMS spam filtering applications since it can achieve high accuracy and F1-score.

8. A Novel Approach for SMS Spam Filtering Using Machine Learning Techniques

- Author: Bhavana N. Mehta, Vijay M. Gulhane
- Year: 2016
- This research provides a novel machine learning-based method for SMS spam filtering. The study applies several classification algorithms, such as Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR), and makes use of feature selection techniques to find pertinent features from SMS texts. The outcomes of the experiments demonstrate how well the suggested method works to differentiate between messages that are spam and those that are not.

9. SMS Spam Filtering: A Review of Machine Learning Techniques

- Author: Y. Osareh, E. Khreich
- Year: 2016
- This research offers a thorough analysis of machine learning-based SMS spam filtering methods. The study looks at several feature extraction techniques, classification algorithms, and assessment criteria used in previous studies. In order to help academics and practitioners create spam detection systems that are more effective, it also addresses the difficulties and potential paths in SMS spam filtering research.

10. An Ensemble Learning Approach for SMS Spam Detection

- *Author: S. Mehmood, U. Qamar, S. Khan*
- *Year: 2015*
- In order to enhance classification performance, this research suggests an ensemble learning method for SMS spam detection that combines many base classifiers. The work builds different classifier ensembles using methods like boosting and bagging, then tests their efficacy on an actual SMS dataset. The ensemble approach outperforms individual classifiers in accurately recognising spam messages while minimising false positives, as demonstrated by the experimental findings.

III. Proposed Methods

A. Classical Machine Learning Classifiers

In this section we discuss briefly the six used machine learning classifiers: Naïve Bayes, Generalized Linear Model (GLM), Fast Large Margin, Decision Tree, Random Forest, Gradient Boosted Trees and Support Vector Machine [12]. A Naive Bayes classifier is a framework used for probabilistic machine learning classification tasks. It's easy to use but computationally cost effective. Naive Bayes' fundamental assumption is that, given the tag (class) value, the value of any attribute is independent of the value of any other attribute. GLM estimates models of regression for results after exponential distributions [13] . These include the Poisson, binomial, and gamma distributions in addition to the Gaussian distribution. Each serves a different purpose and can be used either for prediction or classification depending on the choice of distribution and connection function [27]. GLM is considered a dynamic version of linear regression models. Fast Large Margin is an SVM-Like algorithm which runs in O(N) [14]. Because of its complexity, Fast Large Margin is perfect for classifying big data. A decision tree is a flowchart-like architecture; it can be used to represent decision and decision-making visually and clearly [23]. Each part of decision tree has a role in the classification process; the inner nodes represent checking of attributes, edges represent result of checking and the terminal nodes represent class labels [16]. Random Forest model is developed from decision trees. The primary idea of a random forest is merging a number of decision-making trees into one model. Separately, the findings of decision trees may lead to non-perfect results, but in combination, the findings are enhanced. Gradient Boosted Trees have the same idea of Random Forest models; but the difference is that in Gradient Boosted models the combination task starts at the beginning [26]. If the parameters are carefully adjusted, gradient models may produce better results than random forests. The disadvantage of gradient models is that they suffer from noisy data. SVM is a popular and simple supervised classifier that depends on finding the hyper-plane which makes two given categories somewhat different. SVM is efficient in situations where the number of dimensions exceeds the number of instances [19].

B. Deep Learning Techniques

- Deep Neural Network (DNN): The earliest neural network models, like perceptrons, were tiny and had only one input layer, one output layer, and maybe one hidden layer in between [17]. "Deep" learning is defined as more than three layers (input and output included). It is a notion with a restricted definition, denoting multiple hidden surfaces. In deep learning networks, each layer of nodes trains on a distinct set of features based on how the layers before it executed [15]. As nodes integrate and recombine traits from previous layers, the properties they recognise become more complicated as you move further into the neural net.
- Recurrent Neural Network (RNN): An RNN is a type of artificial neural network that has a "memory" that stores the necessary prior data. Hidden State, which retrieves specific information about any given sequence, such a word set in a sentence, is the essential component of RNNs. Several copies of the same architecture are made with RNN, with each copy sending data to the network after it [20]. It lowers the parameter complexity in contrast to other neural networks. Although RNN has numerous advantages, it has vanishing gradient problem.
- Long Short-Term Memory (LSTM): Several variants have been developed to resolve the problem of Gradients vanishing in RNN.LSTM is considered the best of them. In theory, a repeating LSTM system tries to "remember" all past information that the network has so far been seen and to "forget" irrelevant data [22]. This is accomplished by adding different layers of activation functions called "gates" for different purposes. Each recurrent LSTM unit also preserves a vector called the Internal Cell State that determines the information that the previous recurrent LSTM unit has chosen to maintain conceptually [21]. LSTM contains four different gates: Input, Output, Input Modulation and Forget Gate.
- Gated Recurrent Unit (GRU): The goal of GRU is to address the vanishing gradient issue that arises when using a standard recurrent neural network. GRU is thought of as an LSTM variation [30]. They sometimes produce similarly good effects and have comparable structures. Unlike LSTM, this has just three gates and does not maintain an internal cell state. The hidden state of the Gated Recurrent Unit is filled with the data from an LSTM recurrent unit in the Inner Cell State [25]. The subsequent Gated Recurrent System will receive the shared data. A GRU has three distinct gates: an update gate, a reset gate, and a current memory gate [28].
- Convolutional Neural Network (CNN): The Convolutional Neural Network (CNN), which was first created to perform deep learning for computer vision applications, has shown to be incredibly effective. We used the concept of a "convolution", which is a sliding window or "filter" that moves across the image, identifying and assessing each significant characteristic separately, then distilling them down to their most crucial components and repeating the process [29]. An input sentence is first divided into embedding words or words, which are low-dimensional representations made by models like GloVe or word2vec. Words are input into a convolutional layer based on their properties. Either "pooled" or aggregated to a representative amount are the convolution results.This number is fed to a neural network that is completely connected, making a classification decision based on the weights assigned to each function within the text [18].
- Hierarchical Attention Network (HAN): Based on the same idea as the Attention GRU, HAN was introduced in [24]. Bidirectional GRU is used in the construction of the HAN architecture to obtain the word context. It has two attention levels for each word and sentence.

IV. System Design

A. Architecture Diagram

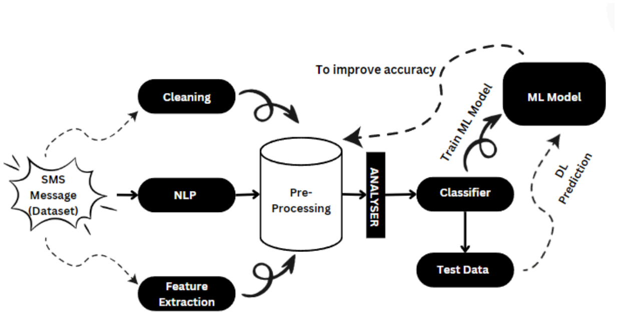

Figure 1 : Architecture Diagram

The above diagram improves the accuracy of an SMS spam classifier using Machine Learning (ML).

- **SMS Messages (Dataset):** It refers to a collection of SMS messages that will be used to train the machine learning model.

- **Cleaning:** The SMS messages go through a cleaning process to remove irrelevant or nonsensical information. This could include things like punctuation, special characters, or extra spaces.
- **NLP (Natural Language Processing):** After the messages are cleaned, they are then processed using natural language processing (NLP) techniques. NLP is an area of study in computer science that addresses the interaction between computers and human language.
- **Feature Extraction:** Once the messages have been cleaned and processed using NLP, features are extracted from them. Features are characteristics of the messages that can be used to identify spam.
- **Train ML Model**: The extracted features are then used to train a machine learning model. The machine learning model is a computer program that learns to identify spam messages based on the features that it has been trained on.
- **Test Data**: After the machine learning model has been trained, it is tested on a separate dataset of SMS messages. The test dataset is used to evaluate the accuracy of the model.
- **ML (Machine Learning)**: This refers to the machine learning model that has been trained to identify spam messages.
- **DL (Deep Learning)**: Based on the extracted features from the SMS message, the deep learning model predicts whether the message is spam or not.

B. Data Flow Diagram

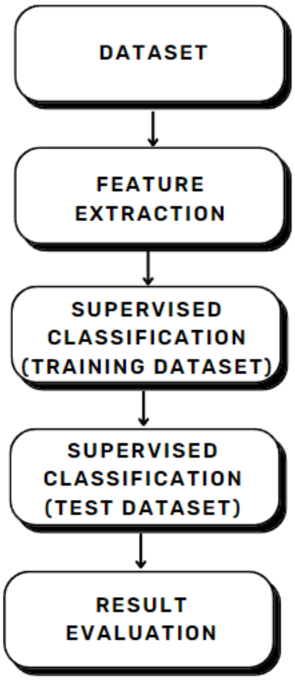
DATASET
FEATURE EXTRACTION
SUPERVISED CLASSIFICATION (TRAINING DATASET)
SUPERVISED CLASSIFICATION (TEST DATASET)
RESULT EVALUATION

Figure 2 : Data Flow Diagram

The above diagram depicts a process of data classification, but it specifically refers to supervised learning for classification.

- **Dataset:** A collection of data that will be used to train the machine learning model. In this case, the data is likely text data.
- **Feature Extraction:** The features are characteristics of the data that can be used to classify it. In text classification, these features might be words, phrases, or other characteristics of the text that can be used to determine its category.
- **Supervised Classification (Test Dataset):** Apply the trained model to a separate dataset to assess how well it generalizes to unseen data. In this stage, feature extraction is performed on the test dataset, and the model makes predictions about the labels of the data points in the test dataset.
- **Result Evaluation:** Evaluate the performance of the machine learning model on the test dataset. This typically involves metrics like accuracy, precision, and recall.

C. Use Case Diagram

Use case diagrams identify the functionalities provided by the use cases, the actors who interact with the system, and the association between the actors and the functionalities.

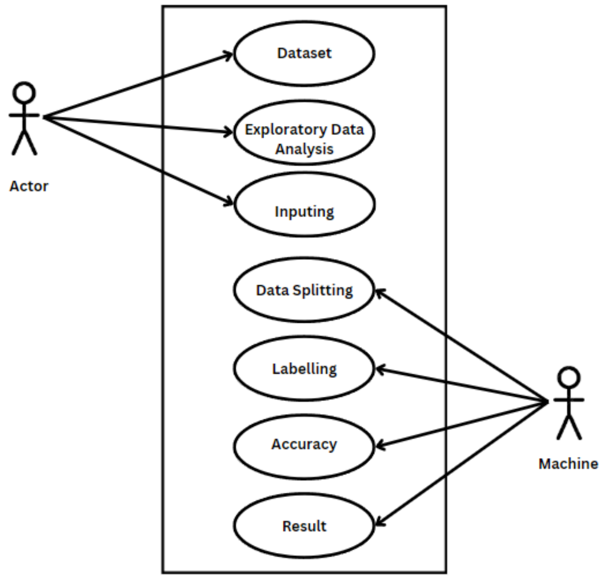

Figure 3 : Use Case Diagram

V. Software Specification

The software requirements document is the specification of the system. It should include both a definition and specification of the requirements. It is a set of guidelines regarding what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification (SRS) document. It is useful in estimating costs, planning team activities, performing tasks, and tracking the teams' progress throughout the development activity.

PYTHON IDE : Anaconda Jupyter Notebook
PROGRAMMING LANGUAGE : Python

VI. Results And Discussions

There are few data sets to predict spam in short message service systems, among which UCI datasets are used in most studies. Thus, in this section of the paper, we first introduce this data set briefly. Then we explain the quality of performing the experiments, where the results on this data set in classification show satisfactory improvement.

A. Data Set Used

The dataset used in this study are of dataset prepared in UCI known as SMS Spam Collection Dataset, and the dataset includes 5573 SMS labelled classified into two groups: 87.37% ham and 12.63% spam. This depicts the proportion between ham and spam drastically increases with inclusion of numbers in the text. Therefore, data is imbalanced.

B. Implementation Details

For implementation, we have used machine learning and deep learning algorithms like Multinomial NB, Decision Tree, K - Nearest Neighbor, Random Forest, AdaBoost, Gradient Boosting, Extra Trees, Bagging, XGB Classifier, LSTM, BI-LSTM. We also created out own custom model.

First we converted the text label to numeric and split the data into training set and testing set. Also, converted label to numpy arrays to fit deep learning models. 80% of data were used for training and 20% for testing purposes. As deep learning models do not understand text, we converted text into numerical representation. For this purpose, the first step is tokenization. The tokenizer API from TensorFlow Keras splits sentences into words and encodes these into integers.

Once tokenization is done, we represented each sentence by sequences of numbers using texts_to_sequences from tokenizer object. Subsequently, we padded the sequence so that we can have same length of each sequence. Sequencing and padding are done for both training and testing data.

All three models (i.e. LSTM, BiLSTM and our custom model) classify the first message ("You have won $100192") and the third message ("You should click the link below...") as spam with very high probabilities (close to 1). This indicates strong agreement on these messages being likely to be spam. All three models assign very low probabilities to the second message ("where are you?") being spam, suggesting they correctly recognize it as ham. The custom model seems to be slightly less certain about the first message compared to the LSTM and BiLSTM models, judging by a slightly lower probability (0.7758 for custom vs. near 1.0 for LSTM/BiLSTM). However, all models still classify it as highly likely to be spam. There are very minor differences for the second and the third messages between the models. All three models appear to perform well on these sample messages, effectively identifying potential spam messages.

VII. Conclusion

Nearly every nation is plagued by the SMS spam message issue, which is growing and shows no signs of abating as the number of mobile users increases in addition to cheap rates of SMS services. Therefore, this paper presents the spam filtering technique using various machine learning algorithms and deep learning techniques. Different algorithms will provide different performances and results based on the features used. For future works, adding more characteristics like message durations could aid the classifiers in training data better and give better performance.

VIII. Future Scope

The future scope of this project will involve adding more feature parameters. The more the parameters are taken into account, more will be the accuracy. The algorithms can also be applied to analyze the content of public comments and thus determine patterns/relationships between the customer and the company. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole.

IX. Applications

- Can be used by companies to prevent users from using fake links.
- Hacking can be prevented.

Acknowledgment

We are pleased to acknowledge our sincere gratitude to Dr. Padmanaban R. of Vellore Institute of Technology, Chennai for his kind encouragement in doing this project and completing it successfully.

We convey our gratitude to VIT Management for providing us with the necessary support and details timely during the progressive reviews.

We wish to express our gratitude to all Teaching and Non- Teaching staff members of the Department of Computer Science and Engineering (SCOPE), who were helpful in many ways for the completion of the project.

References

[1] SMS Spam Detection using Machine Learning and Deep Learning, Sridevi Gadde, 2021.

[2] Short Message Service (SMS) Spam Filtering using Machine Learning in Bahasa Indonesia, Agustinus Theodorus, Tio Kristian Prasetyo, Reynaldi Hartono, Derwin Suhartono, 2021.

[3] Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms, Shah Nazir, Habib Ullah Khan and Amin Ul Haq, 2020.

[4] The Impact of Deep Learning Techniques on SMS Spam Filtering, Wael Hassan Gomaa, 2020.

[5] SMS Spam Detection using Machine Learning Approach, Houshmand Shirani-Mehr, 2019.

[6] Comparative Study of Machine Learning Algorithms for SMS Spam Detection, Amani Alzahrani, Danda B. Rawat, 2019.

[7] SMS Spam Detection Based on Long Short-Term Memory and Gated Recurrent Unit, Pumrapee Poomka, Wattana Pongsena, Nittaya Kerdprasop, and Kittisak Kerdprasop, 2019.

[8] A review of soft techniques for SMS spam classification: Methods, approaches and applications, Olusola Abayomi-Alli, Sanjay Misra, Adebayo Abayomi-Alli, Modupe Odusami, 2019.

[9] SMS Spam Filtering using Supervised Machine Learninng Algorithms, Pavas Navaney, Gaurav Dubey, Ajay Rana, 2018

[10] A Review on Mobile SMS Spam Filtering Techniques, Shafi'I Muhammad Abdulhamid, Muhammad Shafie Abd Latiff, Haruna Chiroma, Oluwasfemi Osho, Gaddafi Adbul-Salaam, Adamu I. Abub, 2017.

[11] Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique, Neelam Choudhary, Ankit Kumar Jain, 2017.

[12] Spam Filtering in SMS using Recurrent Neural Networks, Rahim Taheri, Reza Javidan, 2017.

[13] SMS Spam Filtering For Modern Mobile Devices, N. A Azeez, O. Mbaike, 2017.

[14] SMS Spam filtering and thread identificatiob using bi-level text classification and clustering techniques, Naresh Kumar Nagwani, Aakanksha Sharaff, 2016.

[15] Factorial design analysis applied to rhe performance of SMS anti-spam filtering systems, Marcelo V. C. Argao, Edielson Prevato Frigieri, Carlos A. Ynoguti, Anderson P. Paiva, 2016.

[16] Text Normalization and Semantic Indexing to Enhance Instant Messaging and SMS Spam Filtering, Tiago A. Almeida, Tiago P. Silva, Igor Santos, Jose M. Gomez Hidalgo, 2016.

[17] Intelligennt SMS Spam Filtering Using Topic Model, Jialin Ma, Yongjun Zhang, Jinling Liu, Kun Yu, XuAn Wang, 2016.

[18] A Method of SMS Spam Filtering Based on AdaBoost Algorithm, Xipeng Zhang, Gang Xiong*, Yuexiang Hu, Fenghua Zhu, Xisong Dong, Timo R. Nyberg, 2016.

[19] Metin Sınıflandırma ve Uzman Sistem Tabanlı ˙Istenmeyen Kısa Mesajların Filtrelenmesi SMS Spam Filtering based on Text Classification and Expert System, Yavuz Selim Bozan, Önder Çoban, Gül¸sah Tümüklü Özyer, Barı¸s Özyer, 2015.

[20] SMS Spam Filtering Application Using Android, Gaurav Sethi, Vijender Bhootna, 2014.

[21] The Impact of Feature Extraction and Selection on SMS Spam Filtering, A. K. Uysal, S.Gunal, S.Ergin, E. Sora Gunal, 2013.

[22] Content based Hybrid SMS Spam Filtering System, T. Chaminda, T. T. Dayaratne, H. K. N. Amarasinghe, J. M. R. S. Jayakody, 2013.

[23] Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector Machine, Tej Bahadur Shahi, Abhimanu Yadav, 2013.

[24] SMS Spam Filtering: Methods and Data, Sarah Jane Delany, Mark Buckley, Derek Greene, 2012.

[25] SMS Spam Filtering Technique Based on Artificial Immune System, Tarek M Mahmoud, Ahmed M Mahfouz, 2012.

[26] Simple SMS spam filtering on independent mobile phone, M.Taufiq Nuruzzaman, Chanmoo Lee, Mohd. Fikri Azli bin Abdullah, Deokjai Choi, 2012.

[27] A novel framework for SMS spam filtering, Alper Kursat Uysal, Serkan Gunal, Semih Ergin, Efnan Sora Gunal, 2012.

[28] Independent and Personal SMS Spam Filtering, M. Taufiq Nuruzzaman, Changmoo Lee, Deokjai Choi, 2011.

[29] SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering, 2011.

[30] Spam Filtering for Short Messages, Gordon V. Cormack, Jose Maria Gomez, Enrique Puertas Sanz, 2007.