Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

A RELIABLE METHOD FOR EFFICIENT SPAM DETECTION EMPLOYING MACHINE LEARNING ALGORITHMS

[1]KRISHNA VENI POTTA,[2]H.MADHUSUDHANA RAO
[1]Student,[2]Assistant professor MCA,M.Phill,(Ph.D)
Department of CSE

ABSTRACT:

Short Message Service (SMS) has fallen out of favour in this era of ubiquitous instant messaging apps, becoming instead the domain of service providers, corporations, and spammers looking to reach the masses. Spam communications are becoming more complex to identify and filter as they increasingly include material written in regional languages and typed in English. Text messages in regional languages like Hindi or Bengali written in English were collected from local mobile users and labelled for use in this study, expanding on a standard SMS corpus that includes spam and non-spam texts. In a supervised learning and classification setting, the Monte Carlo method is employed with a standard set of features and machine learning algorithms. The results demonstrate the effectiveness with which various algorithms tackle the problem statement.

Keywords: spam, supervised learning, regional spam, Monte Carlo method, deep learning, convolutional neural networks, short message service spam, term frequency-inverse document frequency (TF-IDF) vectorization.

I.INTRODUCTION:

Man is a social animal, and the very essence of this socializing nature lies in their ability to effectively communicate. From the cave drawings in early ages to the blazingly fast instant messaging applications prevalent in these times, the need for effective and timely communication has always been a priority in human life. The basic components of a typical communication are as shown in Figure 9.1 where a communication medium is used by sender(s) to communicate with the receiver(s). This medium of communication has taken several forms over the many decades of human civilization. For instance, cave walls, letters (pages), and text messages are all different forms of communication medium that man has used With the onset of mobile technology in human lives, the concept of hand-written letters was replaced by a new form of communication, referred to as the Short Message Service or SMS. The first instance of sending a mobile device–based text message was recorded in the year 1992 [1], and it has come a long way since then. This service gained popularity at a very rapid rate, and became an integral part of technology enriched human life in the last two decades. Using the SMS, each mobile device user can compose a textual message of length up to 160 characters including alphabets, numeric values, and special symbols [2]. This constitutes the “short message” that can be sent to

a recipient (another mobile device user). This mode of communication has utility especially in cases where short pieces of information need to be urgently conveyed or where attending calls is not plausible. However, the last decade has witnessed the meteoric rise in the use of internet-based messaging services which are faster and cheaper than SMS in most cases. Also, such services are made more attractive with no message length limit, inclusion of stickers, GIFs, and other application specific enhancements to make them the primary choice of mobilebased communication. This has pushed the erstwhile default communication medium to a secondary position, and nowadays it is seldom used in day-to-day communication by general mobile users. Instead, this service has become a handy tool for different service and/or product-based companies, who use it to implement their strategy of direct marketing. The SMS-based marketing strategy adapted by different companies provides a unique opportunity to identify and incite their potential clients by providing them attractive incentives and offers on chosen products or services. A recent survey revealed that 96% of the participants from India admitted they receive unwanted spam message every day, of which 42% receive almost 7 such SMS per day [3]. Despite the regulatory and preventive norms put in place by the Telecom Regulatory Authority of India (TRAI) on the broadcast of unwanted messages, only about 6% of Indian mobile users find the Do Not Disturb (DND) service useful [4]. A general understanding of spam as unwanted or unsolicited messages is essential in order to effectively prevent or detect and filter such messages at the user end. Oblivious mobile users are highly prone to signing up for such irritating SMS automatically when they are availing a service or purchasing a product of their choice. Online marketing, banking, telecom service, etc., constitute a bulk of the unwanted or spam messages that Indian users usually receive. Yet more harmful is the set of fraudulent spam messages that target innocent users and aim to lure them and extract crucial information regarding their personal details, banking passwords, etc., as shown in Figure 9.2. On the other hand, the desired electronic texts that a mobile user expects to receive are called ham messages. Such SMS could be bank account related updates or travel ticketbased information, etc. So, it is essential to accurately distinguish between these two types of SMS. Typically, the SMS-based communication including spam filtering may be illustrated as represented as shown in Figure 9.3.

Over the years, there has been extensive research on different spam detection and filtering techniques, though not all of them have resulted in efficient and productive end user applications. The current work deals with the determination of robustness of the commonly used classification algorithms consisting of conventional machine learning classifier models as well as contemporary Deep Neural Network architecture–based models. This is undertaken by utilizing the Monte Carlo approach by performing the training and classification tasks on different combinations of both spam and ham data for up to 100 times. As a result, the definitive performance statistics for each classification model can be realized and the best performing

model may be chosen as the ideal one. The state of the art of research on spam identification has been discussed in the following Literature Review.

II.LITERATURE SURVEY:

In this section, the authors have discussed some recent state of the art research works in the field of spam detection on SMS messages, going up to the last 5 years. The discussed works have proposed and implemented novel features, effective processing techniques and different advanced machine learning algorithms toward developing an efficient SMS spam recognition system. Back in 2015, Agarwal et al. [5] utilized the comprehensive data corpus consolidated by [6] and extended it by adding a set of spam and ham SMS collected from Indian mobile users. They demonstrated how different learning algorithms like Support Vector Machine (SVM) and Multinomial Naïve Bayes (MNB) performed on the Term Frequency–Inverse Document Frequency (TF-IDF)–based features extracted from the corpora. Starting at around this time, a plethora of research works have used the same corpus and similar set of features and learning algorithms for designing spam detection systems. In the following set of similar works, it is observed that a set of learning and classification algorithms are used for a performance comparison study. Also, there is a paradigm shift toward neural network-based learning algorithms in more recent times. In such a work in 2017, Suleiman et al. [7] demonstrated a comparative study of the performance of MNB, Random Forest, and Deep Learning algorithm–based models by using the H2O framework and a self-determined set of novel features on the same SMS corpus. Using word embedding features, Jain et al. [8] showed in 2018 how Convolutional Neural Network (CNN) can be utilized to achieve a better performance than a number of other baseline machine learning models in determining the spam messages from the corpus of [6]. In the same year, Popovac et al. [9] illustrated how CNN algorithm performs on the same SMS corpus using TD-IDF features. In 2019, Gupta et al. [10] proposed a voting ensemble technique on different learning algorithms, namely, MNB, Gaussian Naïve Bayes (GNB), Bernoulli Naïve Bayes (BNB), and Decision Tree (DT) for spam identification using the same corpus. The trend of classifier performance comparison continues till recent times in 2020, where the work by Hlouli et al. [11], illustrated how Multi-Layer Perceptron (MLP), SVM, k-Nearest Neighbors (kNN), and Random Forest algorithms perform on the same SMS corpus for detecting spam and ham using Bag of Words and TF-IDF–based features. In a similar contemporary work, GuangJun et al. [12] highlighted the performance of kNN, DT, and Logistic Regression (LR) models on SMS spam corpus, though the feature extraction techniques were not discussed. A recent but different type of work by Roy et al. [13] shows how the same SMS corpus by Hidalgo et al. [6] is classified using Long Short Term Memory (LSTM) and CNN-based machine learning models with a high accuracy. The authors also noted that dependence on manual feature selection and extraction results often influences the efficacy of the spam detection system and consequently utilized the inherent features determined by the LSTM and CNN algorithms. Another interesting observation stems from the inclusion of SMS content in

languages other than English for spam and ham identification, as undertaken by Ghourabi et al. [14] in their recent work. The authors used TF-IDF and word embedding–based features for the conventional machine learning models (such as SVM, kNN, DT, and MNB) and proposed CNN-LSTM hybrid model, respectively. This is the only recent research work that intends to identify the spam content in non-English language from a multi-lingual corpus.

It is observed that in spite of the comparative study of classification performance undertaken by the aforementioned state-of-the-art works, none of them have attempted to determine and establish the robustness of the classification techniques in spam identification. Also, the abundance of spam messages in regional language (typed in English) is largely ignored in such works.

III.CONCLUSION:

Effective spam detection and filtering is a very well visited field of research, and there is a wide variety of feasible solutions that have been proposed. It is obvious from a review of relevant, recent state-of-the-art literature that the most distinct progress is in the use of newer, advanced algorithms that are capable of learning more about the inherent patterns of different spam and ham messages in a text corpus. Such algorithms are mostly based on Neural Networks and variants of Deep Neural Networks, such as CNN and LSTM. In the current work, a spam detection system that takes as input a comprehensive and well tested SMS corpus, which has been extended by including the context of regional messages typed in English, has been designed and evaluated. The system employs a Monte Carlo approach to determine which of the supervised classification algorithms among CNN and other conventional machine learning algorithms like SVM, kNN, and DT and is the most robust in detecting the spam messages accurately. For this purpose, k-fold cross-validation has been utilized with a high value of k = 100, at intervals of 10 folds. It has been determined experimentally that the proposed approach results in consistent performance in case of all the classifiers and that CNN emerges as the most robust classification technique with an accuracy and F1 score about 99.5%. Also, among the conventional learning algorithms, SVM is the most robust with standard evaluation metric values of above 98%. Thus, the given novel text corpus has been effectively classified by the designed system and CNN can be utilized as a robust learning and classification technique. A cloud-based framework for implementing the proposed classifier is also discussed. In future, this work can be used as a reference for building robust, real-time spam detection and filtering systems that need to work on SMS corpora that is challenging and contains novel contexts.

REFERENCES

1. Hppy bthdy txt!, BBC, BBC News World Edition, UK, 3 December 2002, [Online]. Available: http://news.bbc.co.uk/2/hi/uk_news/2538083.stm. [Accessed October 2020]. 2. Short Message Service (SMS) Message Format, Sustainability of Digital Formats, United States of

America, September 2002, [Online]. Available: https://www.loc.gov/preservation/digital/formats/fdd/fdd000431.shtml. [Accessed, October 2020]. 3. India's Spam SMS Problem: Are These Smart SMS Blocking Apps the Solution?, Dazeinfo, India, August 2020, [Online]. Available: https://dazeinfo.com/2020/08/24/indias-spam-sms-problemare-these-smart-sms-blocking-apps-the-solution/. [Accessed October 2020]. 4. The SMS inbox on Indian smartphones is now just a spam bin, Quartz India, India, March 2019, [Online]. Available: https://qz.com/india/1573148/telecom-realty-firms-banks-sendmost-sms-spam-in-india/. [Accessed October 2020]. 5. Agarwal, S., Kaur, S., Garhwal, S., SMS spam detection for Indian messages, in: 1st International Conference on Next Generation Computing Technologies (NGCT) 2015, UCI Machine Learning Repository, United States of America, IEEE, pp. 634–638, 2015. 6. Almeida, T.A. and Gómez, J.M., SMS Spam Collection v. 1, UCI Machine Learning Repository, United States of America, 2012. [Online]. Available: http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/, [Accessed October 2020]. 7. Suleiman, D. and Al-Naymat, G., SMS spam detection using H2O framework. Proc. Comput. Sci., 113, 154–161, 2017. 8. Jain, G., Sharma, M., Agarwal, B., Spam detection on social media using semantic convolutional neural network. Int. J. Knowl. Discovery Bioinf. (IJKDB), IGI Global, 8, 12–26, 2018. 9. Popovac, M., Karanovic, M., Sladojevic, S., Arsenovic, M., Anderla, A., Convolutional neural network based SMS spam detection, in: 2018 26th Telecommunications Forum (TELFOR), Serbia, 2018. 10. Gupta, V., Mehta, A., Goel, A., Dixit, U., Pandey, A.C., Spam detection using ensemble learning, in: Harmony Search and Nature Inspired Optimization Algorithms, pp. 661–668, 2019.