IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

https://doi.org/10.62647/ijitce.2024.v12.i3.pp31-39

A Robust Approach for Effective Spam Detection Using Supervised Learning Techniques

[1] T Pravallika,[2] G Sreeramulu, Dr.D.William Albert
[1]M.Tech Student, [2]Assistant Professor, [3]Associate Professor
Department of Computer Science and Engineering
Bheema Institute of Technology and Science, Adoni

ABSTRACT

In this age of popular instant messaging applications, Short Message Service or SMS has lost relevance and has turned into the forte of service providers, business houses, and different organizations that use this service to target common users for marketing and spamming. A recent trend in spam messaging is the use of content in regional language typed in English, which makes the detection and filtering of such messages more challenging. In this work, an extended version of a standard SMS corpus containing spam and non-spam messages that is extended by the inclusion of labeled text messages in regional languages like Hindi or Bengali typed in English has been used, as gathered from local mobile users. Monte Carlo approach is utilized for learning and classification in a supervised approach, using a set of features and machine learning algorithms commonly used by researchers. The results illustrate how different algorithms perform in addressing the given challenge effectively.

1. 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 mobile based 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 ticket based 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.

2. LITERATURE SURVEY

SMS spam detection using H2O framework

SMS spams are one of the concerns and many people do not like to receive them since they are annoying. Many SMS spam detection methods already exist and different classifiers were used, such classifiers depended on Support Vector machine, Naïve Bays and many other machine learning algorithms. In this paper, new classifier is proposed which depends mainly on using H2O as platform to make comparisons between different machine learning algorithms. Moreover, Machine learning algorithms that are used for comparisons are random forest, deep learning and naïve bays. In addition to using deep learning and random forest as classifiers, they are also used to determine the most important features that can be used as input to random forest, deep learning and naïve bays classifiers. Experimental results show that the most significant features that can affect the detection of SMS spam are the number of digits and existing of URL in SMS text. The dataset that is used in experiment is the one proposed by UCI Machine Learning Repositories. Therefore, experiments show that the faster algorithm that achieves high performance is naïve bays with runtime 0.6 seconds, however after comparing it with deep learning and random forest it has the lowest precision, recall, f-measure and accuracy. On the other hand, random forest is the best in term of accuracy with 50 trees and 20 maximum depths, where precision, recall, f-measure and accuracy are 96%, 86%, 91% and 0.977% respectively; nevertheless the runtime is high 30.28 seconds.

Spam detection on social media using semantic convolutional neural network

This article describes how spam detection in the social media text is becoming increasing important because of the exponential increase in the spam volume over the network. It is challenging, especially in case of text within the limited number of characters. Effective spam detection requires more number of efficient features to be learned. In the current article, the use of a deep learning technology known as a convolutional neural network CNN is proposed for spam detection with an added semantic layer on the top of it. The resultant model is known as a semantic convolutional neural network SCNN. A semantic layer is composed of training the random word vectors with the help of Word2vec to get the semantically enriched word embedding. WordNet and ConceptNet are used to find the word similar to a given word, in case it is missing in the word2vec. The architecture is evaluated on two corpora: SMS Spam dataset UCI repository and Twitter dataset Tweets scrapped from public live tweets. The authors' approach outperforms the-state-of-the-art results with 98.65% accuracy on SMS spam dataset and 94.40% accuracy on Twitter dataset.

Convolutional neural network based SMS spam detection

SMS spam refers to undesired text message. Machine Learning methods for anti-spam filters have been noticeably effective in categorizing spam messages. Dataset used in this research is known as Tiago's dataset. Crucial step in the experiment was data preprocessing, which involved reducing text to lower case, tokenization, removing stopwords. Convolutional Neural Network was the proposed method for classification. Overall model's accuracy was 98.4%. Obtained model can be used as a tool in many applications.

Spam detection using ensemble learning

In our daily life, we use email and SMS many times to communicate to each other, but due to the increase of spam email and SMS, it becomes a headache for both the sender and receiver. We need spam detection tool to detect the spam, and there are many spam detection tools available in the market but they are not up to the mark because they only emphasize on individual classifier or only one or two combination of classifier. In our research, we present different combinations of four different classifiers, namely “Gaussian Naive Bayes”, “Multinomial Naive Bayes”, “Bernoulli Naive Bayes”, and “Decision Tree”. We have used voting classifier, a type of ensemble learning to calculate the accuracy of different combinations of classifiers. Results show that use of voting classifier produces more accurate prediction than individual classifier. We had also created an android application to serve the purpose. The mobile application works on client–server principle. Basically, the mobile application acts as a client which sends the data clicked by a user from mobile to server. At the server, there is machine learning script which classifies the received data and sends the prediction back to the client.

3. EXISTING 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.

Disadvantages

The system is not implemented Inverse Document Frequency (IDF).
SMS data is to be finally used by the mathematical model–based supervised learning algorithms. These algorithms fail to deal with textual content in the data and are more comfortable with numeric values.

4. PROPOSED SYSTEM

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 is largely ignored in such works.
1. The system introduces the novel context of identifying spam and ham SMS in regional languages that are typed in English, along with the general English corpus of
spam and ham by extending it.
2. The system employs a Monte Carlo approach and ML Classifiers to repeatedly perform classification using different machine learning algorithms on different

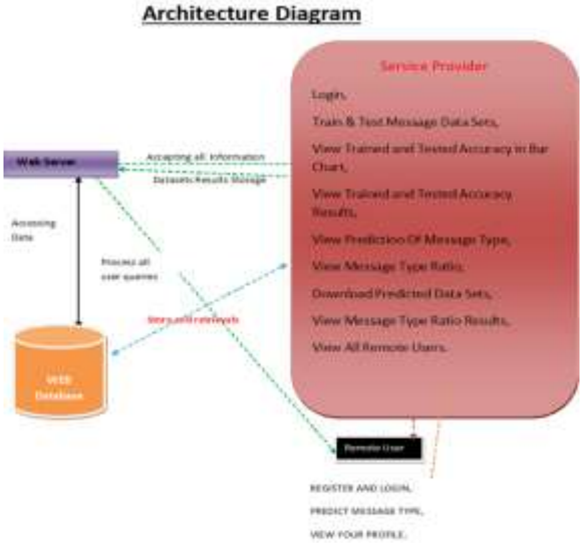
combinations of spam and ham text from the extended corpus (with k-fold cross-validation for a large value of k = 100) in order to determine the efficiency of baseline learning algorithms in comparison to the CNN-based model.

Advantages

The proposed system is more effective due to presence of many ml classifiers.

The proposed system implemented with an accurate prediction for the corresponding dataset.

5. SYSTEM ARCHITECTURE



6. IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train & Test Message Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Message Type, View Message Type Ratio, Download Predicted Data Sets, View Message Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT MESSAGE TYPE, VIEW YOUR PROFILE.

7. SCREEN SHOTS

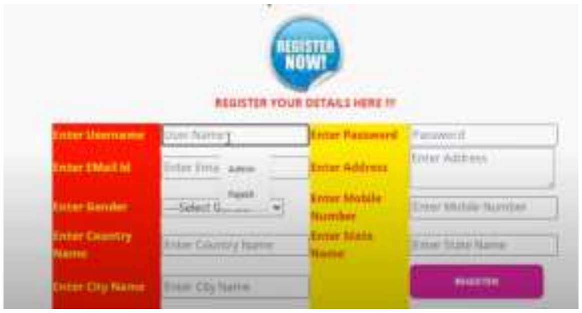
A Robust Approach for Effective Spam Detection Using Supervised Learning Techniques

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

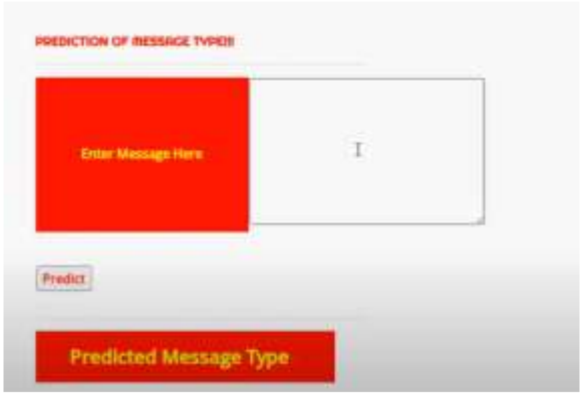
Volume 12, Issue 3, 2024

https://doi.org/10.62647/ijitce.2024.v12.i3.pp31-39



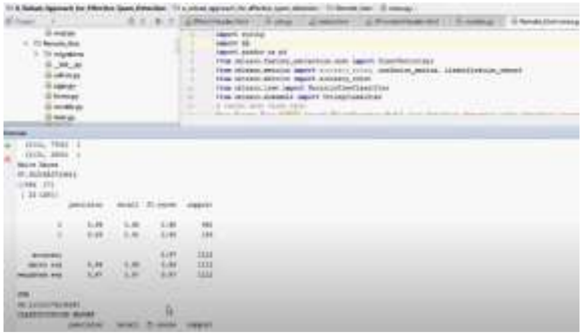
REGISTER YOUR DETAILS HERE !!!

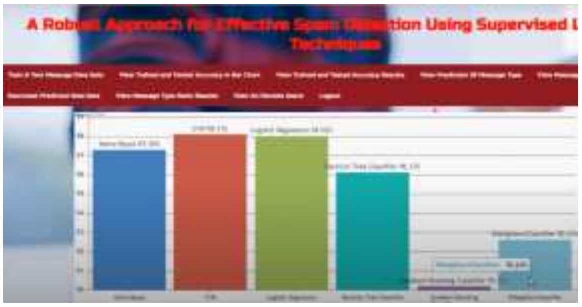
Enter Username	User Name	Enter Password	Password
Enter EMail Id	Enter Ema	Enter Address	Enter Address
Enter Gender	--Select G...	Enter Mobile Number	Enter Mobile Number
Enter Country Name	Enter Country Name	Enter State Name	Enter State Name
Enter City Name	Enter City Name		REGISTER



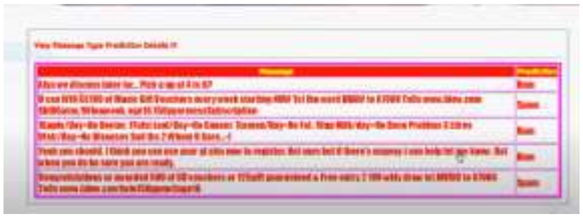
PREDICTION OF MESSAGE TYPE!!!

Enter Message Here



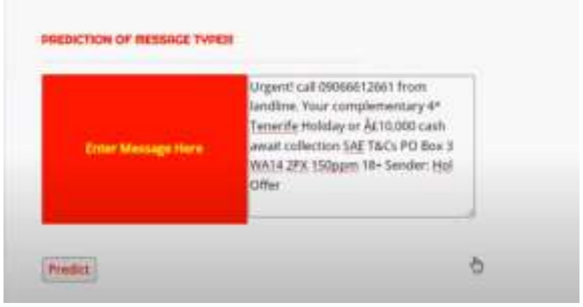


A Robust Approach for Effective Spam Detection Using Supervised L Techniques



View Message Type Prediction Details !!!

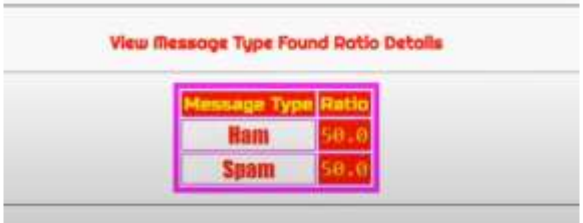
Message	Prediction
[illegible]	Ham
[illegible]	Spam
[illegible]	Ham
[illegible]	Ham
[illegible]	Spam



PREDICTION OF MESSAGE TYPE!!!

Enter Message Here

Urgent! call 09066612661 from landline. Your complementary 4* Tenerife Holiday or Â£10,000 cash await collection SAE T&Cs PO Box 3 WA14 2PX 150ppm 18+ Sender: Hol Offer



View Message Type Found Ratio Details

Message Type	Ratio
Ham	50.0
Spam	50.0

8. 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 Statesof 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.

11. El Hlouli, F.Z., Riffi, J., Mahraz, M.A., El Yahyaouy, A., Tairi, H., Detection of SMS Spam Using Machine-Learning Algorithms, Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco, 1076, 429, Springer Nature, Singapore, 2020.

12. GuangJun, L., Nazir, S., Khan, H.U., Haq, A.U., Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms. Secur. Commun. Netw., Hindawi, 2020, 1–6, 2020.

13. Roy, P.K., Singh, J.P., Banerjee, S., Deep learning to filter SMS spam. Future Gener. Comput. Syst., 102, 524–533, 2020.

14. Ghourabi, A., Mahmood, M.A., Alzubi, Q.M., A Hybrid CNN-LSTM Model for SMS Spam Detection in Arabic and English Messages. Future Internet, 12, 156, 2020.

15. Sammut, C. and Webb, G.I., TF-IDF, in: Encyclopedia of Machine Learning, pp. 986–987, Springer, US, 2010.