IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10403

A Model for Filtering Spam SMS Using Deep Machine Learning Technique

J. Palimote[1], V.I.E Anireh [2], N.D Nwiabu [3]

Department of Computer Science, Rivers State University, Port Harcourt, Nigeria[1,2,3]

Abstract: In recent years, a substantial growth has been experienced in the mobile phone market. A cumulative of 432.1 million mobile gadgets has delivered in the second quarter of 2013 with an increment of 6.0% year over year. As the acquisition of cellphone gadgets has become common, Short Message Service (SMS) has developed into a multi-billion-dollar business. A rush in the quantity of unwanted business notices sent to cell phones utilizing text messages has additionally expanded due to the increased popularity of mobile platforms. This rise attracted attackers, which have resulted in SMS Spam problem. This study presents model for SMS spam filtering classification using Deep Machine Learning Techniques. The system uses the deep machine learning model(MLPNM) in tensorflow and keras framework to classify SMS Message dataset containing 5574 messages. The dataset was read from directory using the pandas.read_csv function. The dataset was cleaned to make sure there are no null values present. The Deep learning model was built with a total of three dense layer with takes in 8672 inputs and 1 output, a batch size (batch size equals the total dataset thus making the iteration and epochs values equivalents) of 32 and epoch value of 50. This trained model was saved and exported into web for easy access and testing with the help of python flask, so that users make various input SMS message. Bootstrap framework (HTML and CSS) was use to design the Front End, while for the Backend, python programming language was use. The results of the test showed accuracy of 99.82% of all input message classified as either ham(legimate) or spam to verify if it's actually a Spam SMS message or a Ham (Legitimate) SMS spam messages.

Keyword: SMS, Spam, Deep Learning, Tensorflow, Keras.

1. INTRODUCTION

Generally, Short Message Service (SMS) is one of the trendy communication services in which a message is sent electronically. The reduction in the cost of SMS services by telecom companies has led to the increased use of SMS. This rise attracted attackers which have resulted in SMS Spam problem. A spam message is generally any unsolicited message that is sent to user's mobile phone. Spam messages include advertisements, free services, promotions, awards, etc. Short Messaging Service (SMS) is a fast growing GSM value added service that is supported by all GSM Phones and by wide range of network standards worldwide [1]. Spam is generally defined as an unsolicited or unwanted messages sent indiscriminately by a sender with no prior relationship to the user mostly for commercial reasons [2]. Spam can be depicted as undesirable or spontaneous electronic messages sent in mass to a gathering of beneficiaries. The messages are portrayed as electronic, spontaneous, business, mass comprises a developing danger primarily because of the accompanying variables: Accessibility of minimal effort mass SMS plans; Dependability (since the message arrives at the cell phone client); Low possibility of accepting reactions from some clueless beneficiaries; and the message can be customized. Versatile SMS spam location and anticipation is definitely not a paltry issue. It has taken on a ton of issues and arrangements acquired from generally more seasoned situations of email spam recognition and separating.

In Nigeria, an average client/user get at least 6 to 10 spam SMS daily either from Mobile network operators advertising their products, banks or the popular "419", Business Company advising their product scammers giving out free gift, job offer and raffle draws etc. [3]. A huge number of SMS are going around the globe over mobile networks per seconds about 33.3% of these SMS are spam [4]. The SMS spam threat is very clear, because users feel the mobile phone is a personal piece of technology that should be kept useful, personal and free from invasions such as spam and viruses [5]. The opposite of spam is called "Ham" which is referred to as legitimate, genuine or desirable message. SMS Spam is not only annoying, it is also frustrating and time wasting because the end users are helpless in controlling the number of SMS spam they receive [6] and it announces its arrival once it is received by the mobile phone [7].

Several strategies have been utilized for SMS spam recognition, like Naïve Bayes (NB), support vector machine (SVM), artificial neural network, decision tree, k-nearest neighbor (KNN) and random forest, in adding up with hybrid methods. However different techniques using different datasets are correlated and analyzed to show that the highest accuracy is achieved using SVM and NB classifiers, though methods like Bayesian classification, logistic regression and decision tree still experience time consuming problem. Recently, Deep learning neural networks an advanced class of machine learning algorithms have obtained significance, which can learn features dynamically and act as a feature extractor and on the other hand it can represent a classifier that classifies the data based on the features learned autonomously from the data. Hence, this study focuses on designing a framework for filtering spam SMS using deep learning algorithm.

2. RELATED WORKS

The paper “SMS Spam Detection using H2O Framework” by [8] proposed a new classifier 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. Their 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 their experiment is the one proposed by UCI Machine Learning Repositories. Therefore, their experiments show that the fastest 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.

The paper “Dendritic Cell Algorithm for Mobile Phone Spam Filtering” by [9] explores a number of content-based feature sets to enhance the mobile phone text messaging services in filtering unwanted messages (spam). Moreover, it develops a more effective spam filtering model using a combination of most relevant features and by fusing decisions of two machine learning algorithms with the Dendritic Cell Algorithm (DCA). The performance has been evaluated empirically on two SMS spam datasets. The results showed that significant improvements can be achieved in the overall accuracy, recall and precision of spam and legitimate messages due to the application of the proposed DCA-based model. Their result on the first and seconddatasets are as follows: Support Vector Machine 96.45%, 96.02% and Naïve Bayes 94.74%, 95.86%.

The paper “Content-Based SMS Spam Filtering Using Machine Learning Technique” by [10] Present Study focuses on Spam SMS identification using Machine Learning technique which is implemented using open source software Python. The dataset they used consist of a collection of 5568 SMS and 2 attributes. The first attribute is class attribute whereas the second attribute is text attribute i.e. SMS. Class attribute has two possible values namely Span and Ham. Among 5568 SMS, 746 SMS are of type Spam and 4822 SMS are of Ham type. For gaining more accuracy; they cleaned their data. Present dataset consists of SMS text which is not useful for text processing. Also, SMS consists of stop words i.e. some words are frequently used such as conjunctions, numbers, prepositions, names, base verbs, etc. The experimental result shows that approximately 98 % Spam SMS’s are identified.

The paper “Using Classification Techniques to SMS Spam Filter” by [11] developed a spam filter for Arabic and English languages by using two filter to be able to detect spam SMS efficiently. Content based method was used to build spam filter for English and Arabic languages. based on this method, there are a number of steps should be taken which are Read English and Arabic dataset, Preprocessing phase, Feature Extraction and Classification. The first step after reading the dataset for Arabic and English languages is preprocessing phase which is important step to get more accurate results. The next step is extracting the features from the body of each message. Eight features have been extracted from English messages and six features from Arabic messages. Then features of messages for English and Arabic languages are splitted into two set: training set and testing set. Training set are used to train the algorithms while the test set are used evaluate the performance of proposed Spam filter for the English and Arabic language. They used two classifiers in detecting SMS spam messages. The classifiers are Naive Bayes is and an Artificial Neural Network as second classifier. The incoming messages are passed through Naive Bayes classifier. If it is classified as ham then passes to second classifier to make sure if it is spam, otherwise it doesn’t pass to second classifier. The results of their proposed system were acceptable with 97% accuracy is obtained for English language when using eight features and 80% from dataset for training. And 95% accuracy is obtained for Arabic language with six features and 70% from dataset for training.

The paper “SMS Spam Filtering for Modern Mobile Devices” by [12] examines solutions to the growing problem of spam and fraudulent messages that are prevalent in the mobile phone industry today. It begins with an examination of some common methods for detecting spam messages such as: The Rule-based method and the Statistical Learning method using Naive Bayes approach. This work specifically explores Naive-Bayes classifier for categorizing messages based on their resemblance with words that feature in other spam and non-spam messages in the training set, thereby reducing the number of spams that get through to the end user and completely eliminate false positives (messages that are misclassified as spam). The dataset they used is the SMS Spam Corpus v.0.1 Big. It has 1,002 SMS ham (legitimate) messages and 322 spam messages. They concluded that using a spam threshold of 0.7 along with adjustments to the Naive Bayes algorithm, gave them some desirable results of 88% on a threshold of 0.9.

The Paper “A Critical Analysis of Existing SMS Spam Filtering Approaches” by [13] critically reviewed the existing SMS spam filters by identifying and analyzing their problems. Some of these problems are adaptability to spammers’

DOI 10.17148/IJARCCE.2021.10403

concept drift, SMS flooding on the network, overhead during training and testing; memory and computational robustness. Furthermore, a taxonomy for existing SMS Spam filtering techniques was constructed. They finally conclude by recommending the use of an adaptive and collaborative SMS spam filtering system.

3. DESIGN METHODOLOGY

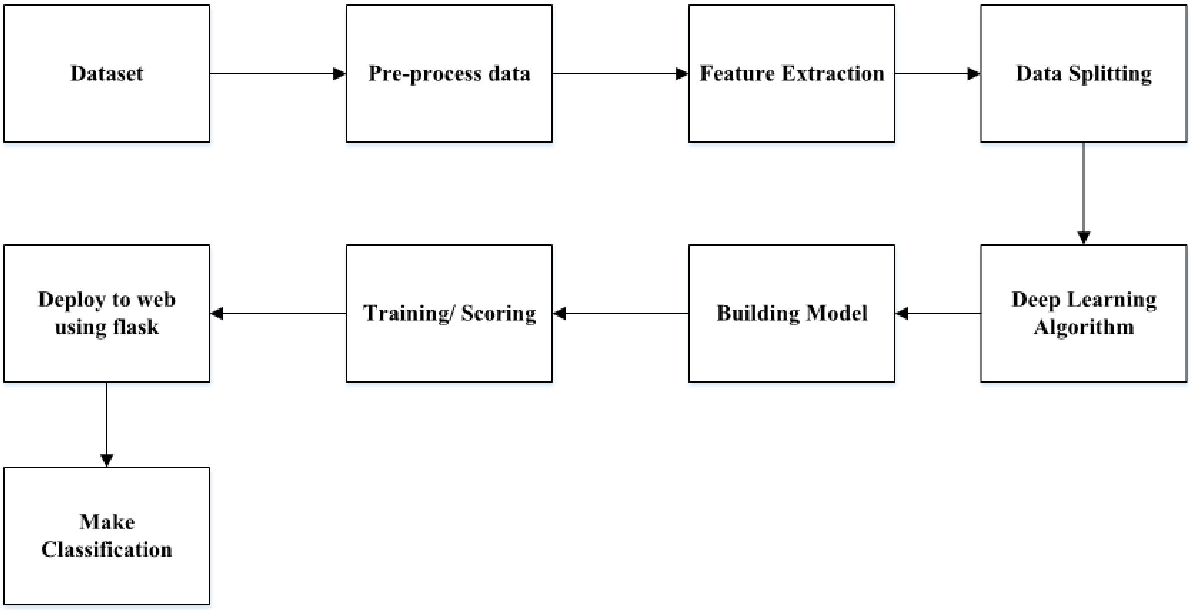
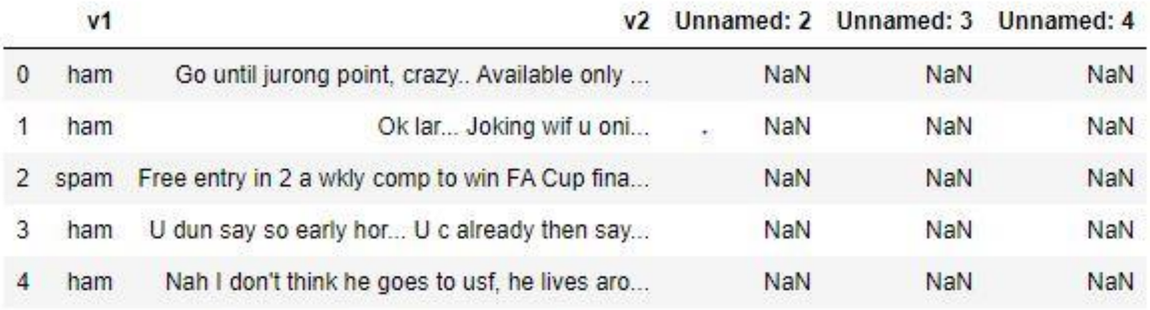


Figure 1: Architecture of the proposed system design

The system uses a SMS Spam dataset which was downloaded from kaggle.com. SMS spam collection is a set of SMS tagged messages that have been collected for SMS spam research. It contains one set of SMS message in English of 5,574 Messages, tagged according being ham (legitimate) or Spam. This dataset was created by (Tiago et al., 2016). The dataset was cleaned and preprocessed making sure that there are no null values present. Feature extraction was used in reducing the dataset dimension and also converting the SMS Spam dataset into array using CountVectorizer function. The dataset was then split into a training and testing set. A Deep Neural Network Algorithm was used into building and training the model in other to detect both spam and ham messages.

4. EXPERIMENT

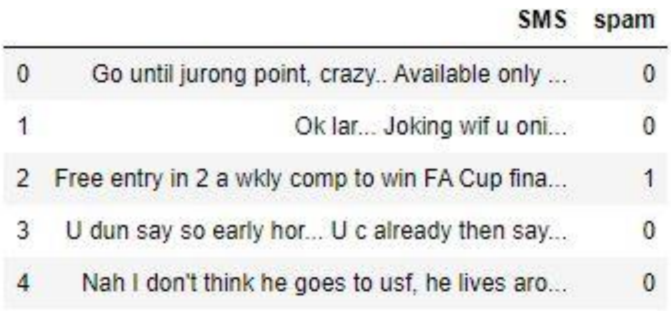
The system uses a Tensorflow and Keras framework in building a Deep Machine Learning Algorithm in classifying SMS spam messages. This Deep Machine Learning Algorithms uses dataset which contains 5574 messages which can be sub divided into 4,650 Ham (Legitimate) messages and 924 Spam Messages. The dataset was read from directory using the pandas.read_csv function. The dataset was cleaned making sure that there are no null values, duplicate value present. CountVectorizer was used in transforming the SMS text to a vector of term/token counts for a better use as input to the Deep Learning Algorithm. This dataset was further divided into x and y variables where x contains the input (which are the SMS messages) and y contains the output (which will display either a Ham or Spam message) by using the train_test_split module from sklearn.model_selction library. The Deep learning model was built with a total of three dense layer with takes in 8672 inputs and 1 output, a batch size (batch size equals the total dataset thus making the iteration and epochs values equivalents) of 32 and epoch (The training steps) value of 50. Figure 1 shows the first five rows of the original SMS spam dataset, which was unprocessed (containing some finite and Nan values) and it contains some unwanted features, there affecting the performance of the model during training. Figure 2 shows the first five columns of the processed (all Nan, duplicate values and unwanted columns have been removed) dataset. Figure. Figure 3 shows the number of Ham and spam messages, figure 4 shows the training processes of the deep machine learning model. Figure 5 shows the accuracy got during training of model, figure 6 shows the loss values at each training process.



	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

Figure 1: Dataset downloaded from kaggle.com

This dataset was created by Tiago et.al. (2016) with a total number of 5574 SMS messages. This figure displays the first 5 messages of the dataset where v1 represents the Ham or Spam messages and v2 represents the SMS messages itself.



	SMS	spam
0	Go until jurong point, crazy.. Available only ...	0
1	Ok lar... Joking wif u oni...	0
2	Free entry in 2 a wkly comp to win FA Cup fina...	1
3	U dun say so early hor... U c already then say...	0
4	Nah I don't think he goes to usf, he lives aro...	0

Figure 2: The Training dataset

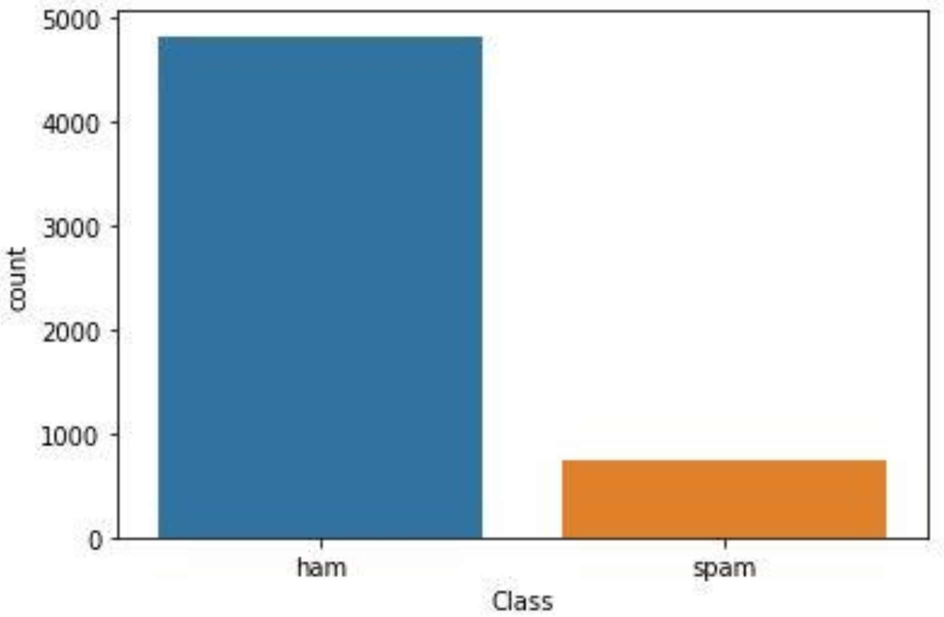
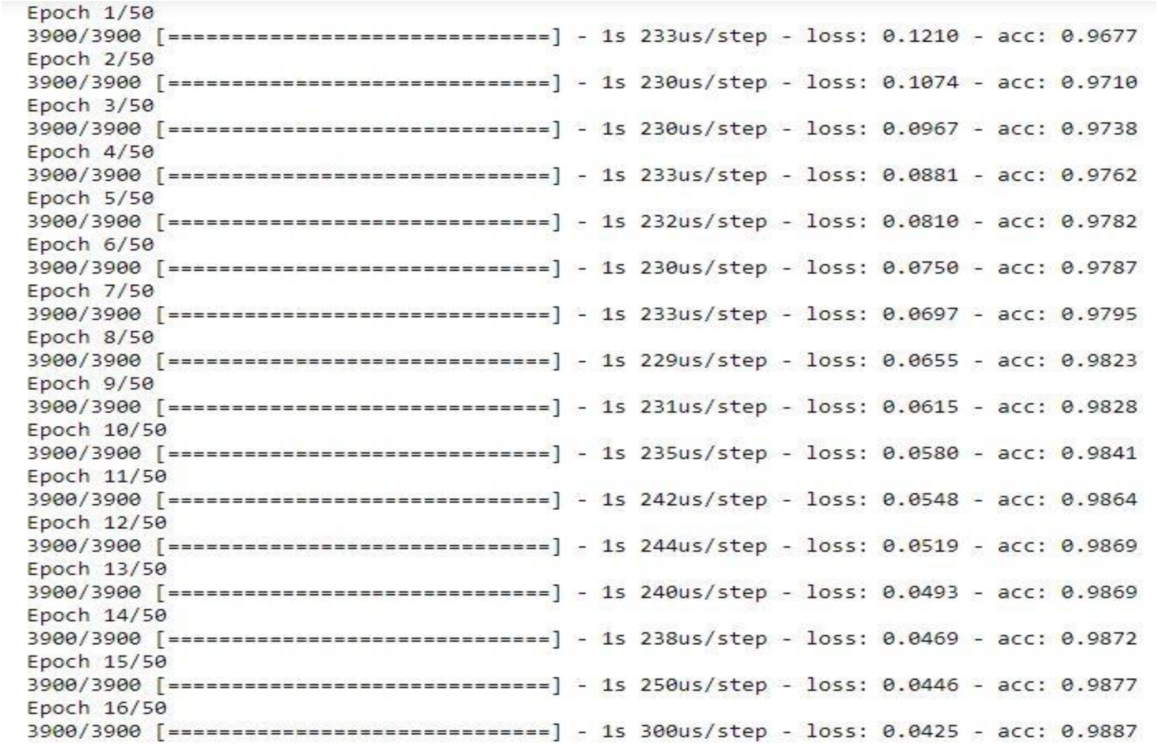


Figure 3: A count plot of the dataset which contains 4,650 Ham messages and 924 Spam messages.

DOI 10.17148/IJARCCE.2021.10403



```
Epoch 1/50
3900/3900 [==============================] - 1s 233us/step - loss: 0.1210 - acc: 0.9677
Epoch 2/50
3900/3900 [==============================] - 1s 230us/step - loss: 0.1074 - acc: 0.9710
Epoch 3/50
3900/3900 [==============================] - 1s 230us/step - loss: 0.0967 - acc: 0.9738
Epoch 4/50
3900/3900 [==============================] - 1s 233us/step - loss: 0.0881 - acc: 0.9762
Epoch 5/50
3900/3900 [==============================] - 1s 232us/step - loss: 0.0810 - acc: 0.9782
Epoch 6/50
3900/3900 [==============================] - 1s 230us/step - loss: 0.0750 - acc: 0.9787
Epoch 7/50
3900/3900 [==============================] - 1s 233us/step - loss: 0.0697 - acc: 0.9795
Epoch 8/50
3900/3900 [==============================] - 1s 229us/step - loss: 0.0655 - acc: 0.9823
Epoch 9/50
3900/3900 [==============================] - 1s 231us/step - loss: 0.0615 - acc: 0.9828
Epoch 10/50
3900/3900 [==============================] - 1s 235us/step - loss: 0.0580 - acc: 0.9841
Epoch 11/50
3900/3900 [==============================] - 1s 242us/step - loss: 0.0548 - acc: 0.9864
Epoch 12/50
3900/3900 [==============================] - 1s 244us/step - loss: 0.0519 - acc: 0.9869
Epoch 13/50
3900/3900 [==============================] - 1s 240us/step - loss: 0.0493 - acc: 0.9869
Epoch 14/50
3900/3900 [==============================] - 1s 238us/step - loss: 0.0469 - acc: 0.9872
Epoch 15/50
3900/3900 [==============================] - 1s 250us/step - loss: 0.0446 - acc: 0.9877
Epoch 16/50
3900/3900 [==============================] - 1s 300us/step - loss: 0.0425 - acc: 0.9887
```

Figure 4: Training process of the Deep Learning Model which displays the training steps, loss values and accuracy for 1-16 epochs

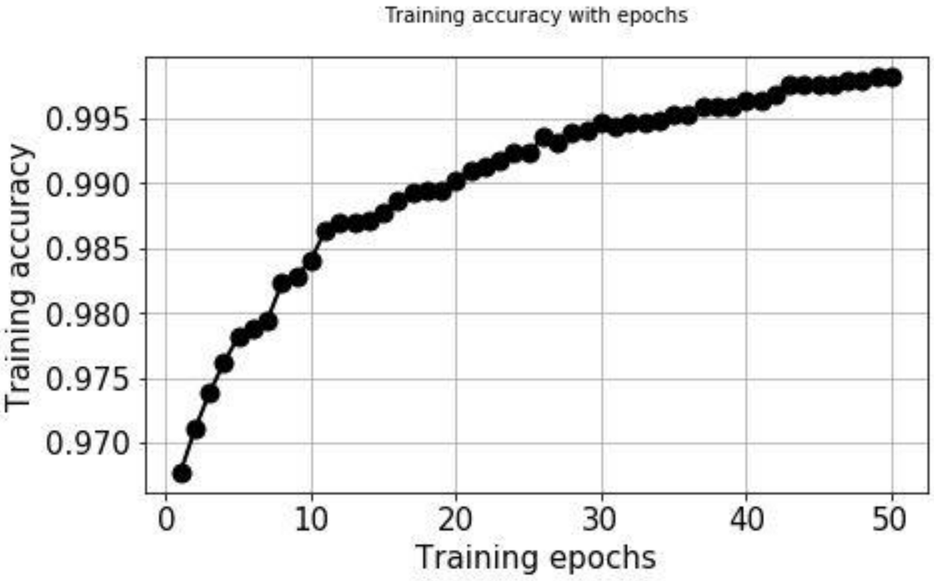


Figure 5: A graphical representation of Training Accuracy Vs Training Epochs.

DOI 10.17148/IJARCCE.2021.10403

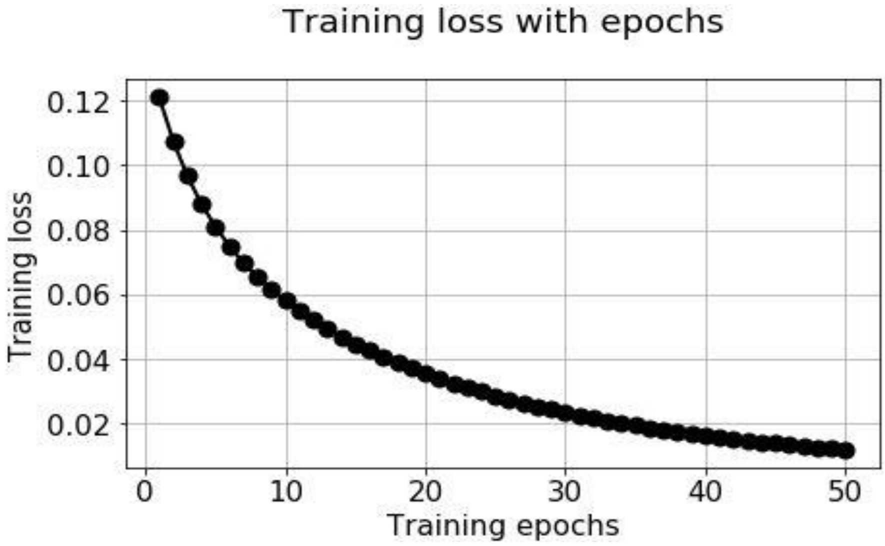


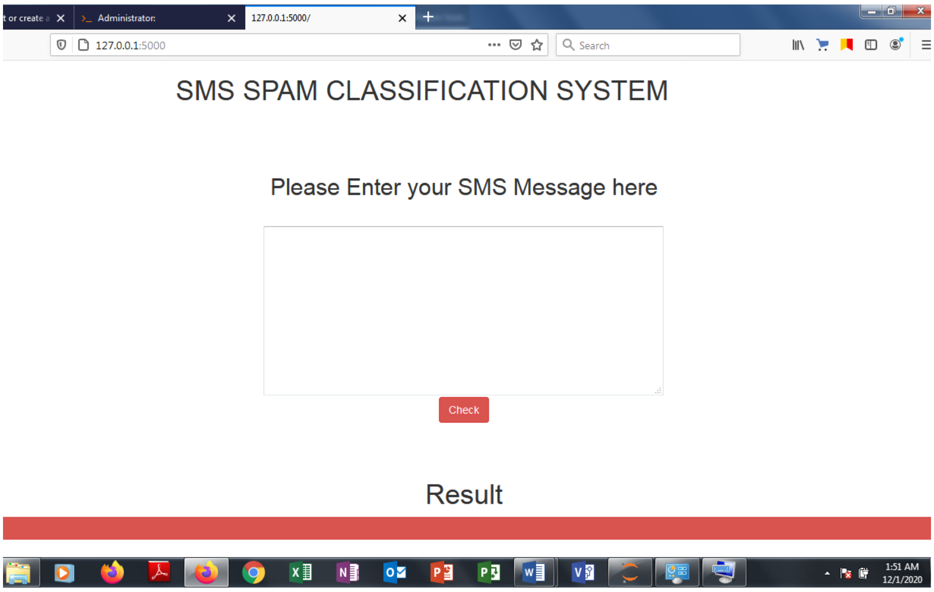
Figure 6: A graphical representation of Training Loss Values Vs Training Epochs.

5. RESULT

After successful training of the model, an accuracy of 99.82% was achieved as show in table 4.1. This trained model was being saved and exported into web for easy execution and testing. The graphical user interface of Filtering Spam SMS is presented in this Paper. Figure B.1 depicts Home page interface where the user inputs his/her SMS. Figure B.2 depicts blank message; no message was inputted. Figure B.3 depicts the classification results been carryout as spam. Figure B.4 depicts the classification results been carryout as Ham.

Table 4.2 1 classification report of Deep Machine Learning techniques

	Precision	Recall	f1-score
0	0.98	0.99	0.99
1	0.97	0.89	0.92
Accuracy			99



SMS SPAM CLASSIFICATION SYSTEM

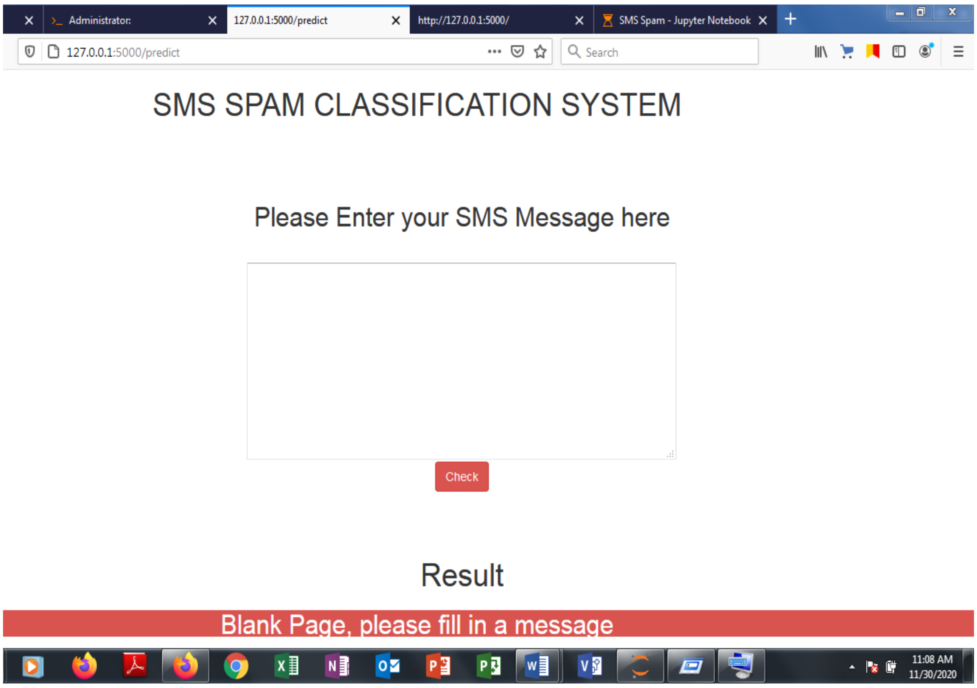
Please Enter your SMS Message here

Check

Result

Figure B.1: Home Module

DOI 10.17148/IJARCCE.2021.10403



SMS SPAM CLASSIFICATION SYSTEM

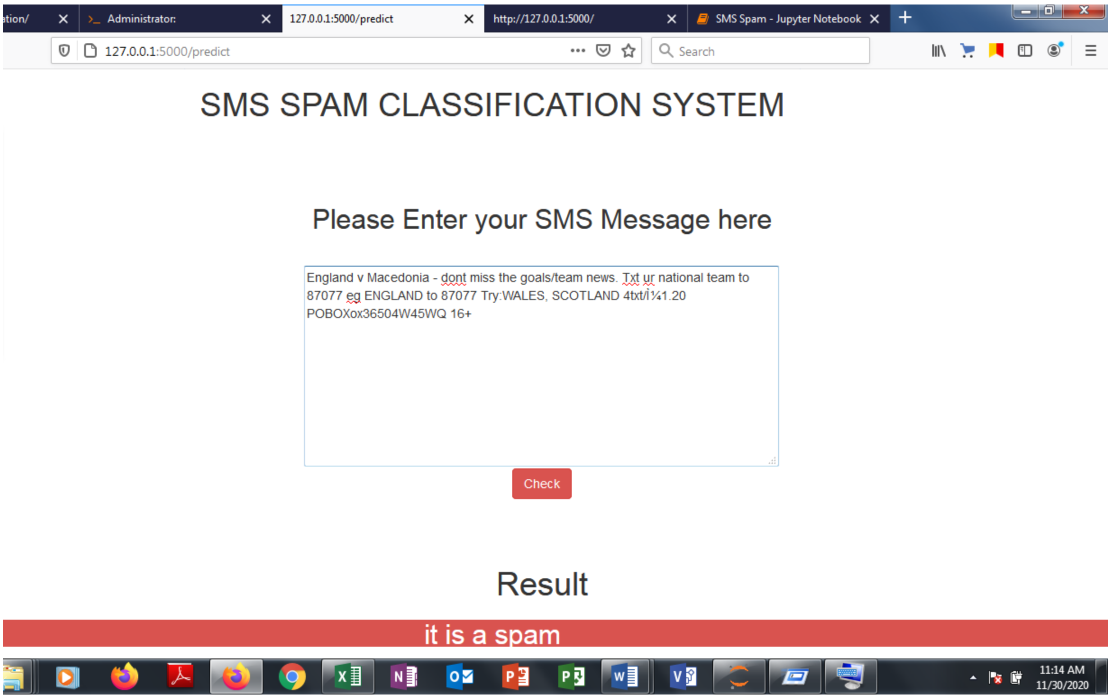
Please Enter your SMS Message here

Check

Result

Blank Page, please fill in a message

Figure B.2: Bank Message



SMS SPAM CLASSIFICATION SYSTEM

Please Enter your SMS Message here

England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt/ì¼1.20 POBOXox36504W45WQ 16+

Check

Result

it is a spam

Figure B.3: Spam Result

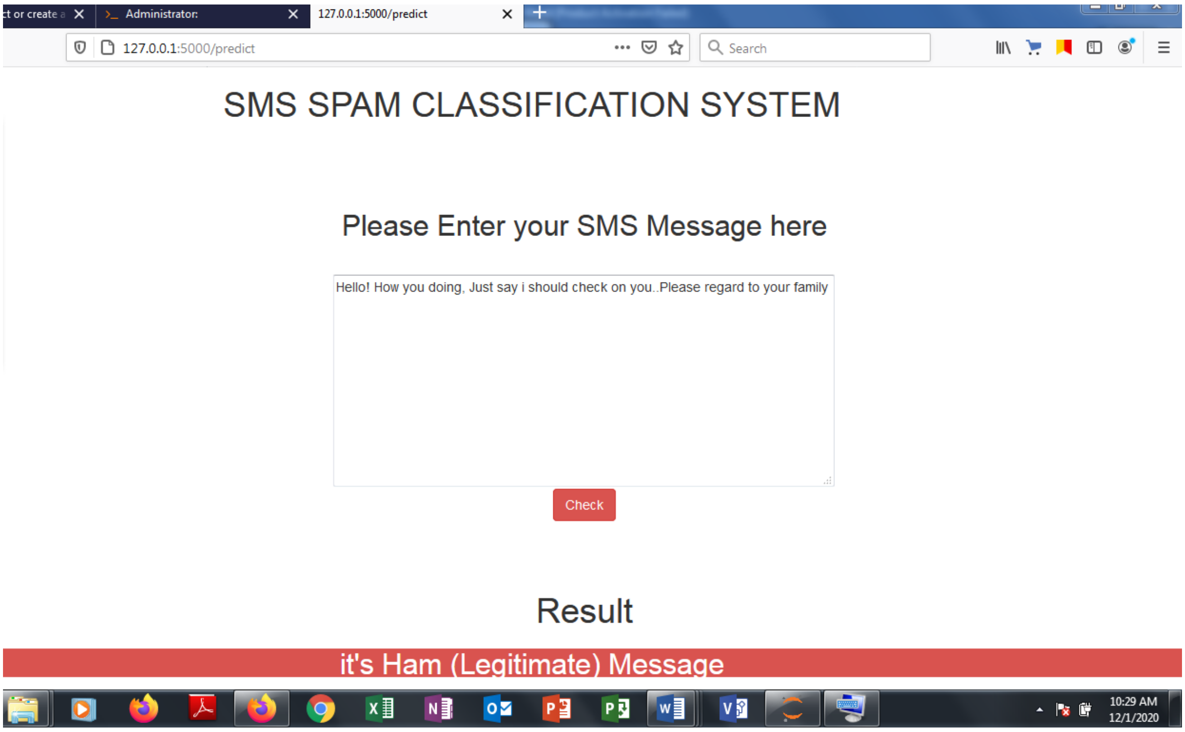
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10403



SMS SPAM CLASSIFICATION SYSTEM

Please Enter your SMS Message here

Hello! How you doing, Just say i should check on you..Please regard to your family

Result

it's Ham (Legitimate) Message

Figure B.4: Ham Result

6. CONCLUSION AND RECOMMENDATION

This paper presents a Tensorflow and Keras framework in building a Deep Learning Algorithm in classifying SMS spam messages. This Deep Learning Algorithms uses dataset which contains 5574 messages which can be sub divided into 4,650 Ham (Legitimate) messages and 924 Spam Messages. The model was built and trained with a total number of 8672 input neurons, 1 output and one dense layer, a batch size (batch size equals the total dataset thus making the iteration and epochs values equivalents) of 32 and epoch (The training steps) value of 50. After successful building of the model, an accuracy of 99.82% was achieved. This paper can further be extended by deploying the trained model into an android application where users can detect an SMS sent to their phone to be either Spam or Ham message.

REFERENCES

[1]. MobileCommonsBlog(2016). https://www.mobilecommons.com/blog/2016/01/how-textmessaging-will-change-for-the-better-in-2016/

[2]. Zeltsan, Z. (2004). General Overview of Spam and Technical Measures to Mitigate the Problem. *ITU-T SG 17. Interim rapporteur meeting, Available online at http://www.docstoc.com/docs/3731634/businessproposal-letters*

[3]. Abayomi-Alli, A. (2009). Content Analysis of Fraudulent Nigeria Electronic Mails to Enhance E-Mail Classification using E-SCAT. MSc *Thesis, Department of Computer Science University of Ibadan, Nigeria*,1(5)30-50.

[4]. Hidalgo, J. M. G., Bringas, G. C., Sánz, E. P. and García, F.C. (2006). Content based SMS spam filtering. In *ACM Symposium on Document Engineering*, 107–114.

[5]. Chaminda, T., Dayaratne T., Amarasinghe H., Jayakody, J. (2013). Content-Based Hybrid SMS Spam Filtering System. In *Proceedings of ITRU Research Symposium, University of Moratuwa, Sri Lanka*, 31-35.

[6]. Nuruzzaman, M. T., Lee, C., Abdullah, M. and Choi, D. (2012). Simple SMS spam filtering on independent mobile phone. *Security Communication Network*, 1209–1220.

[7]. Rafique, M. Z., Farooq, M. **(**2010). SMS Spam Detection by Operating on Byte-Level Distributions Using Hidden Markov Models (HMMs). *In Proceeding of the 20th Virus Bulletin International Conference,* 100-121.

[8]. D. Suleiman and G. Al-Naymat(2017). SMS Spam Detection using H2O Framework. The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks, 133(2017)154-161.

[9]. A. A. Al-Hasan and E. M. El-Alfy (2015). Dendritic Cell Algorithm for Mobile Phone Spam Filtering. Proceeding from the 6th International Conference on Ambient Systems, Networks and Technologies, 244-251.

[10]. D. R. Kawade and K. S. Oza (2018).Content-Based Sms Spam Filtering Using Machine Learning Technique. International Journal of Computer Engineering and Applications, 2321-3469.

[11]. H. H. Mansoor and S. H. Shaker **(**2019**)**. Using Classification Techniques to SMS Spam Filter. International Journal of Innovative Technology and Exploring Engineering, 2278-3075.

[12]. N.A Azeez and O. Mbaike(2017). SMS Spam Filtering for Modern Mobile Devices. Futa Journal of Research in Sciences, 2315-8239.

[13]. O. O. Abayomi-Alli, S. A. Onashoga, A. S. Sodiya, and D. A. Ojo (2015). A Critical Analysis Of Existing Sms Spam Filtering Approaches. Proceeding of the 1st International Conference on Applied Information Technology, 5(9) 211-220.