SMS Spam Detection using Machine Learning and Deep Learning Techniques

Sridevi Gadde
Department of Computer Science& Engineering
Raghu Engineering College
Visakhapatnam, A.P,India
sridevi.gadde@raghuenggcollege.in

A.Lakshmanarao
Department of Information Technology
Aditya Engineering College
Surampalem, A.P, India
laxman1216@gmail.com

S.Satyanarayana
Department of Computer Science& Engineering
Raghu Engineering College
Visakhapatnam, A.P,India
satyanarayana.sivakoti@raghuenggcollege.in

2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) | 978-1-6654-0521-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACCS51430.2021.9441783

***Abstract*—The number of people using mobile devices increasing day by day.SMS (short message service) is a text message service available in smartphones as well as basic phones. So, the traffic of SMS increased drastically. The spam messages also increased. The spammers try to send spam messages for their financial or business benefits like market growth, lottery ticket information, credit card information, etc. So, spam classification has special attention. In this paper, we applied various machine learning and deep learning techniques for SMS spam detection. we used a dataset from UCI and build a spam detection model. Our experimental results have shown that our LSTM model outperforms previous models in spam detection with an accuracy of 98.5%. We used python for all implementations.**

***Keywords*—Short Message Service, Spam, Machine Learning, Deep Learning,LSTM,UCI.**

I. Introduction

The number of mobile phone (smartphone) users increases from 1 billion to 3.8 billion in five years [1]. The top three countries using more mobiles are China, India, US. Short Message Service or SMS is a text messaging service available for the last several years. SMS service can be availed without internet also. So, SMS service is available in smartphones and basic mobiles also. Although smart phones bring several apps like WhatsApp for text messaging, this service can be availed with the help of the internet only. But SMS can be availed at any time. So, the traffic for SMS service increasing day by day. A spammer is a person/company which is responsible for unsolicited messages. For their organization benefits or personal benefits, spammers sending a vast number of messages to the users. These messages are called spam messages. Although there are various SMS spam filtering techniques available[2], still there is a need to handle this problem with advanced techniques. Mobile users may get annoyed because of spam messages. Spam messages can be two types, SMS spam or email spam. The purpose of email spam or SMS spam is the same. Generally, these spam messages are spent by spammers for the promotion of their utilities or business. Sometimes, the users may also undergo financial loss due to these spam messages. Machine Learning is a technology, where machines learn from previous data and made a prediction on future data. Nowadays, machine learning and deep learning can be applied to solve most of the real-world problems in all sectors like health, security, market analysis, etc. There are various techniques available in machine learning like supervised learning, unsupervised, semi-supervised learning, etc. In supervised learning, the dataset is having output labels, whereas unsupervised learning deals with datasets with no labels. We used a dataset from UCI with labels, So we applied various supervised learning algorithms for SMS spam detection.

II. Literature Survey

Applying ML and DL techniques for spam detection is not a new era. Previously, various researchers applied ML techniques for classification SMS spam. Nilam Nur Amir Sjarif[3] et.al applied the TF-IDF technique in combination with a random forest classifier and achieved an accuracy of 97.5%.TF-IDF is a method used to quantify the words in a document by using two measures Term Frequency and Inverse Document Frequency.A.Lakshmanarao[4] et.al applied four machine learning classifiers Decision Trees, Naive Bayes, Logistic Regression, Random Forest for email spam filtering, and achieved an accuracy of 97% with random forest classifier. Pavas Navaney[5] et.al proposed various machine learning algorithms and achieved an accuracy of 97.4% with support vector machines. Luo GuangJun [6] et.al applied various shallow machine learning algorithms and achieved a good accuracy rate with logistic regression classifier. Tian Xia[7] et.al proposed the Hidden Markov Model for the detection of SMS spam. Their model used the information about the order of words thereby solving issues with low term frequency. They achieved an accuracy of 98% with their proposed HMM

978-1-6654-0521-8/21/$31.00 ©2021 IEEE

model. M. Nivaashini [8] et.al applied a deep neural network for SMS spam detection and achieved an accuracy of 98%. They also compared DNN performance with NB, Random Forest, SVM, and KNN. Mehul Gupta[9] et.al compared various spam detection machine learning models with deep learning models and shown that deep learning models achieved a high accuracy rate in SMS spam detection. Gomatham Sai Sravya[10] et.al compared various machine learning algorithms for SMS spam detection and achieved the best accuracy with the Naive Bayes classification model. M.Rubin Julis[11] et.al applied various machine learning classifiers and achieved an accuracy of 97% with a support vector machine. K. Sree Ram Murthy [12] et.al proposed Recurrent Neural Networks for SMS spam detection and achieved a good accuracy rate. S. Sheikh[13] proposed SMS spam detection using feature selection and the Neural Network model and achieved a good accuracy rate. Adem Tekerek[14] et.al applied various machine learning classification models for SMS spam detection and achieved an accuracy of 97% with a support vector machine classifier.

III. RESEARCH METHODOLOGY

First, we collected an SMS spam dataset from the UCI ML repository [15]. Later, we applied different text preprocessing techniques to clean the dataset. Then we applied, we applied various machine learning algorithms and LSTM. The proposed model was depicted in Figure-1.

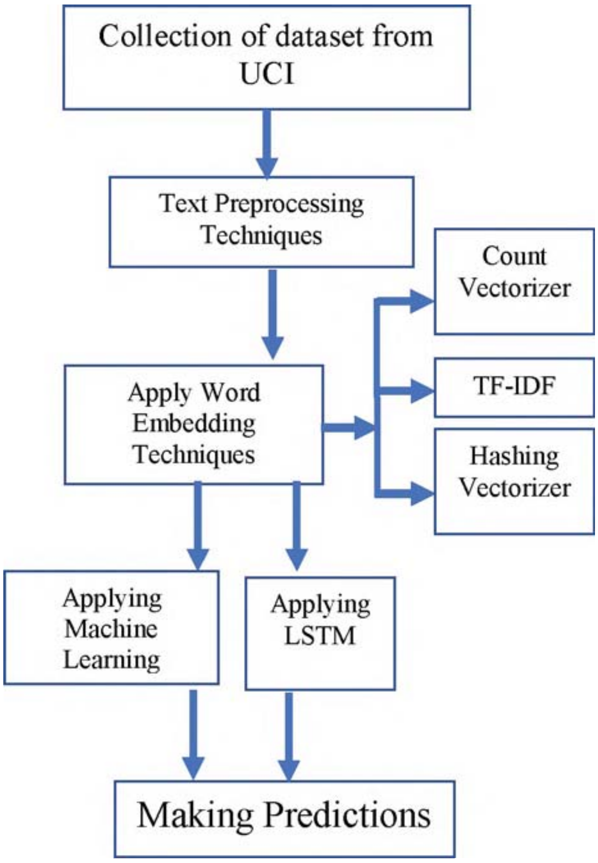
Collection of dataset from UCI
Text Preprocessing Techniques
Count Vectorizer
TF-IDF
Hashing Vectorizer
Apply Word Embedding Techniques
Applying Machine Learning
Applying LSTM
Making Predictions

Fig. 1. Proposed Framework for SMS Spam Detection

A. *Dataset*

We downloaded a dataset from the UCI repository. The dataset contains 5572 rows with two columns. The first column specifies whether the message is "spam" or "ham". Here "spam" means unsolicited message and "ham" means normal message. The second column contains the actual message.

TABLE I. DATASET DETAILS

Dataset	Number of Training samples	Number of Testing Samples	Total
Dataset-(UCI)	4746	986	5572

B. *Text Preprocessing*

Text data can be represented in vectorized format. Text preprocessing can be done by various NLP(Natural Language Processing) techniques. Machine Learning algorithms work with numbers only. So, there is a need to encode text data into the numeric format. Tokenization is a process of dividing text data into different parts. In-Text preprocessing stop words are removed. stop words are the words that are not useful for analyzing the text data. For example, the words like is, was, that are stop words. After removing, stemming can also be applied. Stemming is a process of reducing the word to its stem. For example, the word "playing" can be changed as "play". After that word embeddings can be done, where the words are changed as vectors of real values.

We applied three different word embedding techniques.

- Count Vectorizer
- TF-IDF Vectorizer
- Hashing Vectorizer.

Count Vectorizer

In this, first, all the preprocessing is done like removing special symbols, converting to lower case, etc. Later, it identifies the unique words in the whole text and creates an array of zeros for every sentence. Next, it adds word count to each sentence. The resultant vector is the vector representation of the given text.

TF-IDF Vectorizer

This model uses two measures Term Frequency and Inverse Document Frequency.TF is number of times the sentence appears.IDF is inverse document frequncies. IDF(i)=log2(No.of documents/No.of docs with sente

Hashing Vectorizer

Hashing Vectorizer uses a hashing algorithm for converting text to vector.The algorithm can be applied to all sentences in the document.

C. Machine Learning techniques

After converting text into real valued vectors,we applied various machine learning classifiers like naive bayes,random forest decision tree etc.We also applied a deep learning model "LSTM".

Naive Bayes Classification

Naive Bayes classification algorithm is based on bayes theorem.This theorem is based on probability theory.

Logistic Regression

Logistic Regression uses logit function and sigmoid function for classification tasks.The output variable is predcited based on s-shaped curve.

K-Nearest Neighbors

K-NN is a simple but efficient machine learning classifier,which is based on distance calculation.It identifies the k-nearest neighbors and based on the count of neighbors class,the new data point is classified.

Decision Tree Classification

Decision Tree classifier builds a tree based on which classification can be done.The tree is built recursively until a fixed number of minimum nodes.

Random Forest classification

Random Forest classifier takes the opinion of several decision trees to decide the class of a new datapoints.It is an ensemble approach.

Support Vector Machine

In SVM,a hyper plane is constructed based on which classification is done.Some datapoints are used as support vectors to find the hyperplane.

LSTM

Recurrent Neural Networks are a type of Artificial Neural Networks. In RNNs, the current state input is obtained from the previous state output. But traditional Recurrent Neural Networks suffer from the problem of vanishing gradient descent.LSTM(Long Short Term Memory) is introduced to solve the problems with traditional RNNs.LSTMs are better suited for text mining problems.

IV. Experimentation and Results

A. Evaluation Metrics

Precision, Recall, Accuracy are the three measures used for comparing performance evaluation of classifiers.

Precision=True Positives/(True Positives + False Positives)

Recall=True Positives/(True Positives + False Negatives)

Accuracy= (TP + TN) /(TP+TN+FP+FN).

B. Results of Experiments with CountVectorizer

We applied six machine learning classification algorithms namely Logistic Regression,K-NN,DT,SVC,Naive Bayes, Random Forest.We achieved best accuracy(94%) with Logistic Regression.Table-2 shows the results of experiments.

TABLE II. Results of experiments with ML algorithms

Algorithm	Precision	Recall	Accuracy
Logistic Regression	92%	97%	94%
Naïve Bayes	45%	90%	82%
Decision Tree	90%	98%	94%
SVM	95%	94%	93%
K-NN	100%	44%	91%
Random Forest	93%	88%	92%

The dataset is unbalanced. Out of 5572 samples, only 747 samples are with the class label 1(spam). So, we applied a sampling technique to balance the dataset. We applied SMOTE Synthetic Minority Oversampling). It is a heuristic method to generate samples. Table-3 shows the results of experiments with a sampling technique SMOTE. We achieved good accuracy (95%) with Logistic Regression.

TABLE III. Results of experiments with count vectorizer & SMOTE sampling

Algorithm	Precision	Recall	Accuracy
LG	93%	99%	95%
SVM	96%	99%	94%
Decision Tree	88%	99%	92%
Naïve Bayes	82%	99%	88.8%
K-NN	67%	100%	75%
Random Forest	89%	99%	93%

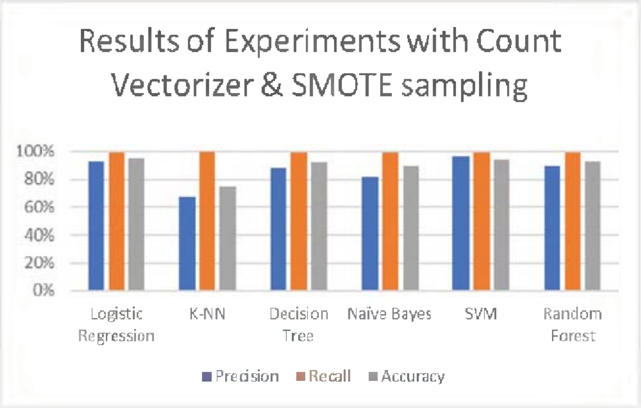

Fig. 2. Results of ML algorithms with Count Vectorizer & SMOTE Sampling

C. Results of Experiments with TF-IDF

We applied TF-IDF word embedding technique and then applied six classification algorithms. The results of the experiments are shown in table-4.

TABLE IV.RESULTS OF EXPERIMENTS WITH TF-IDF

Algorithm	Precision	Recall	Accuracy
Logistic Regression	99%	72%	95%
K-NN	85%	82%	83%
Decision Tree	88%	86%	96%
Naïve Bayes	51%	81%	85%
SVM	99%	82%	97%
Random Forest	99%	78%	96.5%

Figure-3 shows the results of experiments with TF-IDF.We achieved best accuracy with SVM(97%).Decision Tree and Random Forest also given accuracy above 96%.

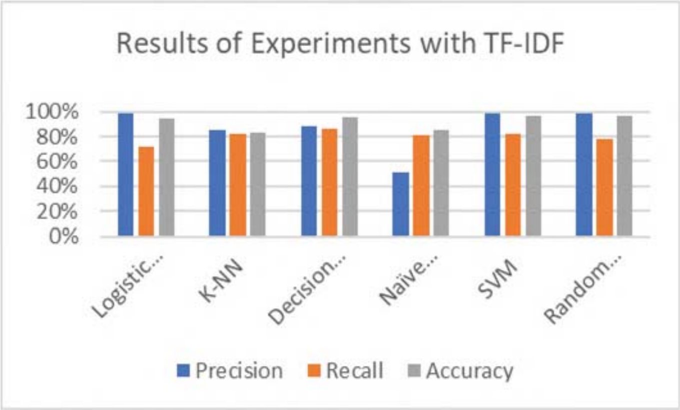

Fig. 3. Results of ML algorithms with TF-IDF

D. Results of Experiments with Hashing Vectorizer

TABLE V. RESULTS OF EXPERIMENTS WITH HASHING VECTORIZER

Algorithm	Precision	Recall	Accuracy
LG	98%	73%	95%
Naïve Bayes	30%	86%	67%
Decision Tree	82%	84%	94%
K-NN	100%	47%	92%
SVM	99%	83%	97%
Random Forest	98%	77%	96%

We applied Hashing Vectorizer word embedding technique and then applied six classification algorithms.The results of the experiments are shown in table-5. We achieved best accuracy with SVM (97%).

Random Forest also given accuracy above 96%.

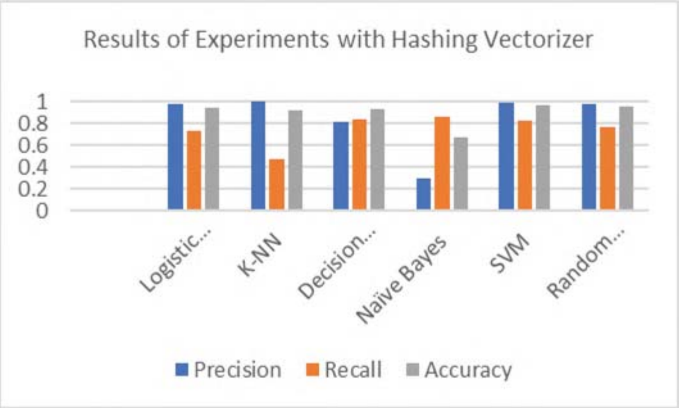

Fig. 4. Results with Hashing Vectorizer

E. Results of Experiments with LSTM

After applying all ML classifiers,we applied a deep learning model.We applied LSTM for the same dataset and achieved an acuracy of 98.5%.

F. Comparison with Previous Work

Table 6 shows the accuracy comparison of the proposed model with previous work. In [3], the authors achieved an accuracy of 97.5% with a ensemble model random forest classifier. In [5], they achieved an accuracy of 97% with an SVM classifier. Our proposed LSTM model achieved an accuracy of 98.5%.

TABLE VI. COMPARISON WITH PREVIOUS WORK

Model	Accuracy
Random Forest[3]	97.5%
SVM[5]	97%
proposed model(LSTM)	98.5%

V. CONCLUSION & FUTURE WORK

In this paper, we proposed a deep learning model for SMS spam detection. We used a UCI dataset for our experiments. We applied three different word embedding techniques count vectorizer, TF-IDF, Hashing Vectorizer. Later, we applied various classification algorithms. We achieved an accuracy of 98.5% with the LSTM model. Experimental results showed that our model outperforms previous models for spam detection.We tested this model on only this dataset.In Future,we will test our model on several datasets.

REFERENCES

[1] Online: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[2] S. M. Abdulhamid, M.S.Abd Latif, Haruna Chiroma, "Robust Heart Disease Prediction A Review on Mobile SMS Spam Filtering Techniques", EEE Access, vol. 5, pp. 15650-15666, 2017, doi: 10.1109/ACCESS.2017.2666785

[3] Nilam Nur Amir Sjarif, N F Mohd Azmi, Suriayati Chuprat, "SMS Spam Message Detection using Term Frequenct-Inverse Document Frequency and Random Forest Algorithm," in The Fifth Information Systems International Conference 2019, Procedia Computer Science 161 (2019) 509–515,ScienceDirect.

[4] A.Lakshmanarao,K.Chandra Sekhar, Y.Swathi, "An Efficient Spam Classification System Using Ensemble Machine Learning Algorithm," in Journal of Applied Science and Computations, Volume 5, Issue 9, September/2018.

[5] Pavas Navaney, Gaurav Dubey, Ajay Rana, "SMS Spam Filtering using Supervised Machine Learning Algorithms.," in 8th International Conference on Cloud Computing, Data Science & Engineering, 978-1-5386-1719-9/18/ 2018 IEEE.

[6] Luo GuangJun,, Shah Nazir, Habib Ullah Khan, Amin Ul Haq, "Spam Detection Approach for Secure Mobile Messgae Communication using Machine Learning Algorithms.," in Hindawi,Security and Communication Netwroks,Volume 2020,Article id:8873639.July-2020.

[7] Tian Xia, Xuemin Chen, "A Discrete Hidden Markov Model for SMS Spam Detection.," in Applied Science,MDPI, Appl. Sci. 2020, 10, 5011; doi:10.3390/app10145011.

[8] M. Nivaashini, R.S.Soundariya, A.Kodieswari, P.Thangaraj, ": SMS Spam Detection using Deep Neural Network.," in International Journal of Pure and Applied Mathematics, Volume 119 No. 18 2018, 2425-2436.

[9] Mehul Gupta, Aditya Bakliwal, Shubhangi Agarwal,Pulkit Mehndiratta, ": A Comparative Study of Spam SMS Detection using Machine Learning Classifiers.," in 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018.

[10] Gomatham Sai Sravya, G Pradeepini, Vaddeswaram, ": Mobile Sms Spam Filter Techniques Using Machine Learning Techniques.," International Journal Of Scientific & Technology Research Volume 9, Issue 03, March 2020.

[11] M.Rubin Julis, S.Alagesan:, "Spam Detection In Sms Using Machine Learning through Textmining", International Journal Of Scientific & Technology Research Volume 9, Issue 02, February 2020.

[12] K. Sree Ram Murthy,K.Kranthi Kumar, K.Srikar, CH.Nithya, S.Alagesan:, "SMS Spam Detection using RNN", : International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 05,May 2020.

[13] S. Sheikhi,M.T.Kheirabadi,A.Bazzazi, "An Effective Model for SMS Spam Detection Using Content-based Features and Neural Network", : International Journal of Engineering, IJE TRANSACTIONS B: Applications Vol. 33, No. 2, (February 2020) 221-228.

[14] Adem Tekerek "Support Vector Machine Based Spam SMS Detection", Journal of Polytechnic, 2019; 22 (3): 779-784.

[15] https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection