International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 5, Sep-Oct 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Spam Message Classification Using LSTM

Mr.M.Priyadharshan*, Sakthivel**, Sanjay A***, Sanjeevi krishnaa A****
Siva Balaji*****
*(CSE, Hindusthan College of Eng & Tech, and Coimbatore
Email: priyadharshan.cse@hicet.ac.in)
**(CSE, Hindusthan College of Eng & Tech, and Coimbatore
Email: sakthivel12357@gmail.com)
*** (CSE, Hindusthan College of Eng & Tech, and Coimbatore
Email: stephensanjay196 @gmail.com)
****(CSE, Hindusthan College of Eng & Tech, and Coimbatore
Email: Sanjeevikrishnaa34 @gmail.com)
*****(CSE, Hindusthan College of Eng & Tech, and Coimbatore
Email:thiyas14300@gmail.com)

----------**********----------

Abstract :
Spam message classification has become a crucial task in ensuring secure and reliable digital communication, especially with the exponential growth of mobile messaging and email services. Traditional machine learning models such as K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Random Forest have been widely applied; however, these models often rely heavily on feature engineering and may not efficiently capture sequential dependencies in text. To overcome these limitations, this study employs a Long Short-Term Memory (LSTM) based deep learning approach for spam detection. LSTM networks are well-suited for natural language processing tasks due to their ability to retain long-term dependencies and context within sequential data. By leveraging word embeddings and an LSTM architecture, the system automatically learns meaningful text representations and classifies messages as spam or ham with high accuracy. This method reduces reliance on handcrafted features and demonstrates superior adaptability to diverse linguistic patterns.

Keywords—Deep Neural Networks; Long Short-Term Memory; Machine Learning; Support Vector Machine and Word Embedding.

----------**********----------

1. INTRODUCTION

Running days, only mobile phones are there in absolutely

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions incredible high rate usage. (SMS) short message service is also known as "text message" are the medium of communication among the peoples. Spamming also was known as "bulk message" is a process of commercializing the product or sending a suspicious message with advertising content to the arbitrary recipient, which is a cost-effectiveness source of business marketing among worldwide cell numbers. It is a true challenge for Business Corporation and organization how to make the text beneficial along with tiny size. A spam message is an emerging and rapidly growing problem for the

internet world. However, the message receiver feel awkward while getting undesirable advertisements in text from an obscure sender. The mobile phone is likewise in danger in light of spam assaults. These assaults keep growing and turned into a difficult issue in numerous nations. Plenty techniques and mechanism have been proposed to distinguish these attacks in sites, email, and SMS. In spite of that, attack frequency still increases [1].

Machine learning method comes to be the most popular choice for Spam detection, a number of researchers have utilized supervised machine learning methods for comparative results of spam classification. A plenty of research has been carried out in this direction make use of machine learning techniques such as Navïe Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). Using traditional machine learning methods mentioned above, the feature engineering is a time-consuming process with the extra computational expense. It is also difficult to extract all the information in short length of the text. In this paper, we will perceive the SMS spam detection using Long Short-Term Memory (LSTMs) method.

2. RELATED WORK

2.1 Machine Learning

Machine learning as its name indicates that to use machines in the sense that it can learn by itself as a human does from experience. More precisely, the whole process of learning and predicting information from its experience (data) is known as machine learning. Herein examples of machine learning, which are the particularly common situation, like, online customer support, social media services, virtual personal assistants, E-mail and Malware filtering and many more, perhaps very few people might have no idea that they are driven by the wave of the machine learning.

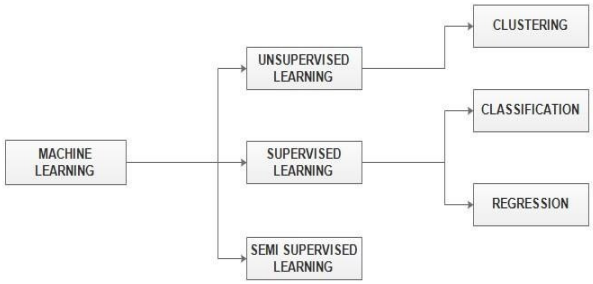
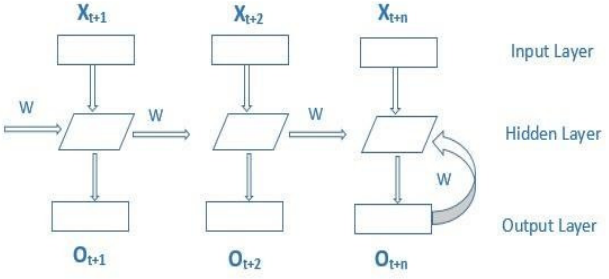


Figure 1. Machine Learning

At present, supervised learning, semi-supervised learning and unsupervised leaning are sub-parts of machine learning. Machine learning is the utilization of existing information as an experience for learning in order to use this information to make better decision in future. As a NLP (natural language processing), mathematics in science, and computer technology which was governed by machine learning, demonstrates another exploration of research and techniques. Additionally, got noteworthy achievements, like neural system accomplished a great outcome in the field of word embedding, which is region where NLP has dependably been exceptionally hot. These exciting achievements are indivisible from machine learning. And impact of two fields will turn into a problem area.

Existing system:

Existing systems for spam message classification largely use traditional machine learning and text processing techniques. Initially, rule-based filters were applied, focusing on keywords, blacklisted senders, and suspicious patterns. Later, machine learning methods such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees,



Random Forests, and K-Nearest Neighbors (KNN) improved accuracy by analyzing features like

word frequency, n-grams, and message length. With advancements in Natural Language Processing (NLP), models began using Bag-of-Words, TF-IDF, and Word2Vec for better feature extraction. Recently, deep learning approaches such as CNN and LSTM have been integrated, offering higher precision by capturing semantic and contextual meaning in messages.

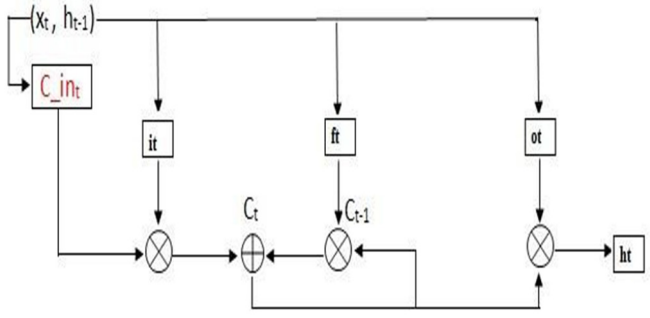
What is new in my proposed system:-

Most existing spam classification systems rely on a single approach, such as Naïve Bayes, SVM, or LSTM, which have strengths but also limitations. Your system introduces a hybrid approach that combines KNN, Random Forest, and LSTM, which is new because it merges the strengths of multiple models. KNN adds a similarity-based classification layer, Random Forest contributes robust pattern learning, and LSTM captures deep contextual meaning in messages. This combination improves accuracy, adaptability, and resilience against evolving spam tactics, making the system more efficient and reliable compared to existing single-method systems.

2.2 Recurrent Neural Network

Recurrent Neural Network (RNN) are well-known models that proved itself in many NLP tasks. RNN perform the repetition of the same job for every element of the input sequence and output is calculated based on previously calculations. In other words RNN has a memory

$$i_t = g(W_{xi}x_t + W_{hi}h_{t-1} + bi) \quad (1$$



which stores previously calculated information. In conceptual, RNN can make use of hidden state's information in unlimited sequences but in practice it's more complicated to move back up from many time steps .The structure of RNN could be shown in Fig. 2, RNN network is compound of three layers, input layer, hidden layer, and an output layer. There is a one-sided information transmit from the input layer to the hidden layer. Another one-sided information transmitted is from the hidden layer to the output layer and output information is returned to the hidden layer from the output layer. Only this makes RNN totally different from others traditional neural networks. The memory of the RNN network memorizes the previously processed information for the utilization in calculation of the current output, that is, the nodes between the hidden states are no longer connected but in reality they are connected each other, and the inputs to the hidden layer include not only the output of the current input layer but also the output of the hidden layer at the previous time.

3. PROPOSED METHOD (LSTM)

For improvement of spam classification in the form of accuracy and feature engineering, we used long short Term Memory (LSTM), the usage of LSTM is crucial to this success. It is an improved version of the recurrent neural network which employees for many tasks, it works better than traditional networks. In our proposed work, we applied this method to baseline text message (SMS) dataset for spam classification. Further discussion is shown below.

Recurrent Neural Network (RNN) is well known as critical thinking of time arrangement, in spite of the fact that, RNN tended to an expansive assortment of solutions including speech recognition, language translation, and natural language processing, it failed to deal with long-term dependencies. LSTMs (Long Short-Term Memory) is capable of storing (remembering) information in the memory cell for huge space of time, it is a behavior by default. The physical structure of LSTM is the same as RNN has, but according to the functionality, LSTMs model has a bit change in hidden layer for keep tracking and calculating the state vector. The structure which makes it different, that are four main elements: three gates and one is cell unit. That is, an input gate, a forget gate, output gate, and last one is a selfrecurrent connection along with neuron respectively

Where it is used:- In spam message classification, the LSTM (Long Short-Term Memory) model is specifically used in the deep learning stage where the actual prediction is performed. After the text data is preprocessed by cleaning, tokenizing, and padding the sequences, an embedding layer is applied to convert words into dense vectors. Following this, the LSTM layer is added to the model architecture, where it learns sequential dependencies and contextual meaning from the message text, something that traditional models cannot capture effectively. The LSTM helps in identifying patterns like repeated spam phrases, contextual word usage, and long-term relationships between words. Finally, a dense output layer with a sigmoid activation is used to classify the message as spam or ham. Thus, LSTM is mainly used in the model-building section of the code, where it serves as the core layer responsible for capturing the temporal and semantic patterns in messages.

3.1 Word Embedding

Word embedding transforms the contents of the textual data into real number vectors. Word representation is a vector related to every word that contains information about the semantics of the word . This alteration is essential in light of the fact that, machine learning and deep learning algorithms are most familiar with a numerical representation of data instead of text data. The word embedding technique which is a subfield of natural language modeling, utilized to graph words or phrases from vocabulary to a relating vector of a real number.

In the word embedding One-Hot Representation and Distributed Representation are two main forms of the representation. To represent the word in One-Hot representation, a long vector is utilized, and in all components of the vectors only a single component of the vector is 1, and others are all 0s. The drawback of this representation is that there will be a word gap, that is, the semantic relations cannot be described between two-word vectors. The second form is Distributed Representation, which was first proposed by Hinton in 1986 to defeat the deficiencies of one-hot representation [2]. The skip-gram technique is one of the types of word embedding, the working process of the skip gram model is opposite to others models. It predicts the sequence of related words to given input word. In our proposed work we use the skip gram model, furthermore, description is given below.

3.2 Skip-gram Model

Skip-Gram model is operated to ascertain word representations that are worthwhile for predicting the nearby words in a sentence or a document. In more rigorous way, aim of the Skip-Gram model is to optimize the average log probability of a given sequence of training words w1, w2, w3, . . . ,w. This method was first introduced by Mikolov et. al. [6] which is considered as an efficient method for learning high-quality vector representations of words mined from disorganized text data. Training of the Skip gram model (as shown in figure 4) does not involve dense matrix multiplications as has been done by most of the previously used neural network architectures for learning word vectors. An optimized single-machine implementation can train more than 100 billion words on routine basis, making words training more efficient. The word representations computed using Skip Gram model seem very interesting because the learned vectors unambiguously encrypt many linguistic regularities and patterns. Somewhat unpredictably, large number of patterns can be represented as linear translations, Such as, the output of a vector calculation vec (“Madrid”) – vec (“Spain”) + vec (“France”) is closer to vec (“Paris”) than to any other word vector. Skip gram model has ability to produce many more training instances and better performance than CBOW on limited amount of training dataset. Based on this fact, we are able to extract more information and essentially have more accurate results.

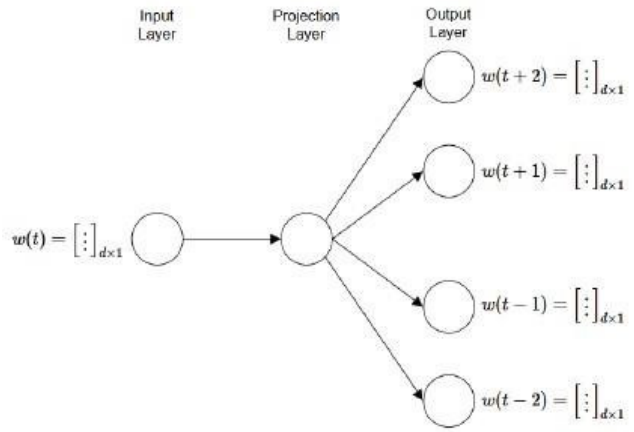


Figure 4. The Skip-gram Model Architecture.

3.3 Dataset:-

We used the publicly available dataset of SMS labeled messages - SMS Spam Collection v.1 – that have been composed for cellphone spam research. It has one of the assemblage of dataset composed by 5,574 English, real and non-encoded messages, labeled according to being appropriate (ham) or spam. This corpus has been collected from free for research sources on the Internet. A group of 425 SMS spam messages was manually extracted from the Grumble text Website is http://www.grumbletext.co.uk/. And a collection of 3,375 SMS of the NUS SMS Corpus (NSC), randomly chosen as a ham messages for research at the National University of Singapore. The NUS SMS Corpus is availablehttp://www.comp.nus.edu.sg/~rpnlpir/ downloads/corpora/ smsCorpus/. List of 450 SMS ham messages collected from Caroline Tag's Ph.D. Thesis available at http://etheses.bham.ac.uk/253/1/Tagg09PhD.pdf, and incorporated the Spam SMS Corpus v.0.1 Big. It has 1,002 ham messages and 322 spam messages and it is publicly available at http://www.esp.uem.es/jmgomez/smsspamcorpu s/ [8]. Table 1, shows the description of the dataset.

Dataset	Training Set	Testing Set	Total
Ham	2773	924	3697
Spam	1408	469	1877

We distributed our dataset into two parts, one is 75% of the dataset used for training and rest 25% used for testing of the model.

3.4 Baseline Methods

In order to evaluate the effectiveness of the proposed framework, the LSTM method was compared with five existing machine learning algorithms as having been previously used in [8] for Spam Classification:

3.4.1 SVM (Support Vector Machine)

Support Vector Machine SVM, is the best linear structure to guide surface classification, is utilize for classification of statistical study, targeting at two categories of classification.
SVM is known as one of the algorithm for very high classification accuracy in the text classification. However, when SVM needs to deal with large datasets, the convergence rate is comparatively become slowly, so it has to use enormous memory resources for massive calculations [9]. SVM transformed data using kernel trick technique and used it to find an appropriate hyperplane for separation of output classes. It is also can transform complex data. SVM is categorized into linear support vector machine and RBF support vector machine [8].

3.4.2 Decision Tree

Decision trees learning is one in which a decision tree is used as a predictive model which maps observations of an object to the conclusions of the target object. When the target variables in a tree model is a finite set, then it is called a classification tree. The leaves of the tree represent the labels and branches denote conjunctions of features that lead to class labels .

3.4.3 KNN (K-Nearest Neighbors)

The KNN algorithm is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether

KNN is used for classification or regression [8]. The classification speed of KNN is slightly lower and slow than other mainstream classification algorithms.

3.4.4 Random Forest

Random forests or random decision forests are an ensemble learning method for tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the made of the classes (classification) or means prediction (regression) of the individual trees.

3.4.5 NB (Naive Bayes)

In machine learning, naive Bayes classifiers are a group of simple probabilistic classifiers in view of applying Bayes' theorem with strong independence expectations between the features. Naive Bayes classifiers are highly scalable and require several parameters linear in the number of variables in a learning problem. Maximum likelihood training can be done by evaluating a closedform expression, which takes linear time, as opposed to by iterative estimation as utilized some different sorts of classifiers.

Where it is Used :-

In the spam message classification system, the machine learning models are applied after converting the raw text messages into numerical features using the TF-IDF vectorizer. This transformation captures the importance of words across the dataset and prepares the data for traditional classifiers. Once the feature vectors are generated, three classical machine learning algorithms are employed. The K-Nearest Neighbors (KNN) algorithm is trained on the TF-IDF vectors to classify a new message by comparing it with the nearest neighbors in the feature space. Similarly, the Random Forest classifier is applied, which constructs multiple decision trees and aggregates their predictions to improve robustness and accuracy. Additionally, a Naïve Bayes classifier (commonly used in text classification due to its probabilistic approach) can also be introduced on the same TF-IDF vectors, leveraging word occurrence probabilities to distinguish between spam and non-spam messages. These machine learning methods serve as a comparative baseline against the deep learning approach, the LSTM model, which uses sequence-based embeddings to capture contextual dependencies in the text.

Why it is used?

KNN (K-Nearest Neighbors)

Purpose: KNN is used to classify a message based on its similarity to other messages in the dataset.

How it works: It compares a new message's features (like word frequency or TF-IDF vectors) to all other messages and finds the "K" closest ones. The majority label among those neighbors determines the classification (spam or ham).

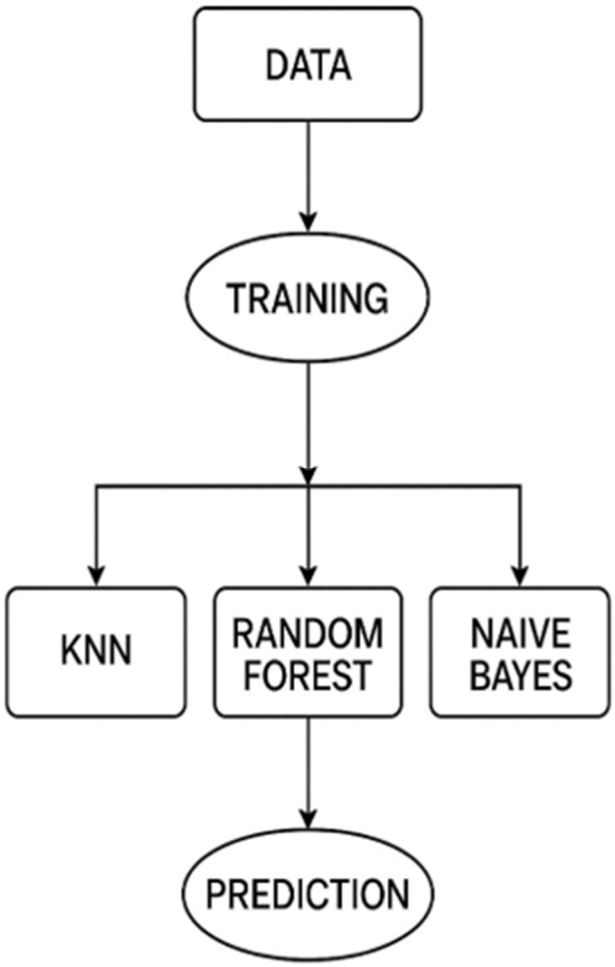
Why used: KNN is simple, effective for small datasets, and doesn't require heavy model training. It helps in identifying spam messages based on similarity patterns.

2. Random Forest

Purpose: Random Forest is used for robust classification by combining multiple decision trees.

How it works: It builds many decision trees on random subsets of the data and averages their results.

Why used: It handles large datasets well, reduces overfitting, and improves accuracy. Random Forest is effective at learning complex



4. RESULTS AND DISCUSSIONS

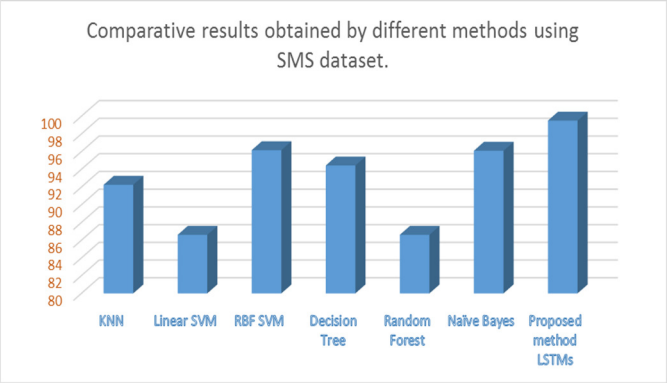
LSTM has addressed critical issues of time series problem, along with it keeps tracking memory of previous information by gating mechanism associated with the memory cell. This mechanism makes the LSTM a best-suited method for the task of SMS spam classification.

LSTM provides better accuracy and performance compared to other common machine learning classifiers, such as SVM, Navïe Bayes (NB), KNN, Decision Tree, and Random Forest used for spam filtering technique, the comparison results obtained by abovementioned algorithms as below shown in table 2.

Table 2. Experimental Results

We evaluated the accuracy of SMS spam classification by using LSTM and skip gram model. We trained our model with the same dataset as baseline models have been trained to facilitate the experimental results. We have been successful in achieving 97.5% accuracy, which is encouraging result.

RBF SVM	96.17
Decision Tree	94.44
Random Forest	86.61
Naïve Bayes	96.10
Proposed Method LSTM	97.5



4.1 Data Preprocessing

The following data preprocessing steps are needed in SMS short text message dataset for training the model:

We removed unnecessary information, punctuation marks like, full stop, comma and some other inoperable marks which are impact the training model of spam filtering.

The skip gram model is well known for word to vector transformation. To carry out the present work, skip gram model is used to transform each word into a vector for the process of the training, Softmax along with skip gram model can also be used for more appropriate word vector.

5. CONCLUSION

This paper deals with the critical issues faced by the other spam filtering techniques, such as accuracy, feature selection. LSTM technique preceded by skip gram model has been implied to mitigate these issues. A comparison of different methods is also presented using the same dataset has been used in the previous methods (SVM, NB, KNN, random forest and

decision tree) which shows that LSTM method seems to be a better alternative for the spam filtration. LSTM provides 97.5% accuracy as measured by performing the experiments.

Though the results are satisfactory, still a deep understanding of different activation functions is needed in LSTM. As skip gram model generate better results for short length of dataset. So, in future, for more improvement we will use LSTM with different word embedding techniques with huge amount of dataset, which will be the future work of the authors.

REFERENCES:

1.Oyeyemi, D. A., & Ojo, A. K. (2024). SMS Spam Detection and Classification to Combat Abuse in Telephone Networks Using Natural Language Processing. arXiv:2406.06578. — SMS spam detection study that compares NLP models; includes experiments with LSTM and modern transformer baselines.

2. Busyra, R. F. (2024). Applying Long Short-Term Memory Algorithm for Spam Detection (Journal of SMS Management Research). — LSTM applied to multilingual web-form spam (English + Indonesian), includes preprocessing and embedding details.

3. Manasa, P. (2024). Detection of Twitter Spam Using GLoVe Vocabulary and LSTM. (Published / indexed in 2024) — Uses GLoVe embeddingsLSTM for detecting spam tweets; reports comparative metrics against other RNNs.

4. Ghourabi, A., et al. (2023). Enhancing Spam Message Classification and Detection (Peer-reviewed / PMC). — Proposes improved SMS classification pipeline; evaluates Bi-LSTM and hybrid CNN+Bi-LSTM variants on public SMS corpora.

