Proceedings of 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS 2024)
IEEE Xplore Part Number: CFP24UD1-ART; ISBN: 979-8-3503-7274-8

SMS Spam Detection using NLP and Deep Learning Recurrent Neural Network Variants

2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS) | 979-8-3503-7274-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICC-ROBINS60238.2024.10533895

Vangapandu Venkata Kalyani
Assistant Professor, Department of CSE
Raghu Engineering College
Dakamarri, Visakhapatnam,A.P,India
kalyani.mith@gmail.com

Dr.M.V.Rama Sundari
Professor in AIML
Gokaraju Rangaraju Institute of Engineering and Technology,Hyderabad
mvramasundari@gmail.com

S. Neelima
Assistant Professor, Department of IT
Aditya Engineering College
Surampalem,India
neelima.sadineni@gmail.com

Pennada Siva Satya Prasad
Assistant Professor,Department of IT
Aditya Engineering College
Surampalem,India
sivasatyaprasadp@aec.edu.in

P.PattabhıRama Mohan
Assistant Professor, Department of IT
Aditya Engineering College
Surampalem,India
ramspatnala@gmail.com

A.Lakshmanarao
Associate Professor, Department of IT
Aditya Engineering Collee
Surampalem,India
laxman1216@gmail.com

Abstract— **The ubiquity of Short Messaging Services (SMS) has been accompanied by an upsurge in spam messages, necessitating effective detection mechanisms. This paper delves into the intricate task of SMS spam detection using a Kaggle dataset of 5570 samples. As initial step, the challenge of class imbalance was addressed by applying oversampling technique ADASYN. Later, TF-IDF embedding technique was employed to capture semantic meaning and contextual information within SMS messages. A conventional ML classifier, Random Forest has achieved a good accuracy of 92.3%. Later, various RNN architectures, including Simple RNN, LSTM, Bi-LSTM and Gated Recurrent Unit (GRU), applied for SMS spam detection. LSTM variants demonstrated superior performance with 93% accuracy for RNN, 98% for LSTM and 98.2%, 98.3% for BI-LSTM and GRU. To further enhance the accuracy, a hybrid architecture combining Bi-LSTM and GRU layers was explored, resulting in improved accuracy of 99%. The proposed model outperformed traditional ML approaches for SMS spam detection. This paper sheds the efficacy of LSTM variants in addressing the SMS spam detection challenge. Emphasizing the significance of handling class imbalances and employing effective embedding techniques, the achieved accuracies hold promising implications.**

Keywords— SMS Spam, Deep Learning, Random Forest, LSTM.

I. Introduction

In the contemporary realm of modern communication, Short Messaging Services (SMS) have established themselves as a pervasive medium for swift and direct information exchange. The convenience and immediacy offered by SMS make it an integral part of daily communication. However, the widespread adoption of SMS has brought forth a significant challenge – the proliferation of SMS spam. The pervasive use of SMS as a communication channel has led to a surge in the frequency of unwanted and unsolicited messages. This influx of SMS spam not only disrupts the smooth flow of communication but also presents inherent threats to user privacy and security. The relentless barrage of unwanted messages can result in an overwhelming and intrusive experience for users. Effectively addressing the issue of SMS spam requires the development of robust and adaptive detection mechanisms. These mechanisms need to be capable of identifying and filtering out spam messages with a high degree of accuracy. Given the dynamic nature of spam, where tactics employed by spammers constantly evolve, adaptive detection systems become crucial in staying ahead of emerging spam techniques. Developing effective SMS spam detection mechanisms involves leveraging advanced machine learning algorithms and artificial intelligence techniques. These technologies can analyse patterns, content, and sender behaviour to distinguish between legitimate messages and spam. Moreover, the system should continually learn and adapt to new spam patterns, ensuring a proactive defence against evolving spam tactics.

While SMS remains a ubiquitous and valuable communication tool, the challenge of SMS spam requires proactive measures. The development of sophisticated and adaptive detection mechanisms is essential to maintain the integrity and security of SMS communication, preserving its effectiveness as a rapid and direct information exchange medium. This research embarks on a comprehensive exploration into the intricate realm of SMS spam detection, leveraging advanced ML to construct models that are not only effective but also adaptable to the evolving nature of spam messages.

II. Literature Survey

In [1], a dataset was collected from Kaggle, and diverse ML techniques were applied for message detection. The Random Forest classifier proved as good model, achieving the accuracy of 97% in SMS spam detection. The utilization of this classifier underscored its robust performance in accurately identifying and classifying spam messages within the dataset. The proposed technique in [2] incorporated a text embedding strategy that makes upon the latest advancements of the GPT-3 Transformer. The goal of the advanced approach is to provide high-quality representations that will improve the detection results. Additionally, an ensemble learning technique was used, which included combining four distinct machine learning models into a cohesive model that outperformed each of its individual components. The experiments evaluations of model were conducted using the SMS Spam Collection Dataset and yielded good results.

The authors of [3] introduced a ML model-based SMS spam categorization and detection system that makes use of

BERT, NLP models. In addition to feature extraction, tokenization and stop word removal were used as data preparation approaches. It was easier to distinguish between spam and ham signals when BERT was used in conjunction with ML models including SVM, RF, Gradient Boosting, NB, and Logistic Regression. The results of the evaluation showed a precision of 97% with the NB classifier. The Naïve Bayes approach and TF IDF used to build the model in [4]. The dataset, acquired from Kaggle, is used to train the model. Remarkably, the model is included on a locally hosted webpage made using the PyCharm IDE. The model's efficacy in correctly recognizing situations is shown by the analysis, which indicates an astounding 95% model accuracy, proving its ability to precisely identify and classify relevant data. The authors of [5] discussed the various techniques, compared each to the dataset, and selected the one that was most effective in categorizing SMS messages as spam. Comparisons of accuracy and precision showed that GloVe was the best feature extractor, whereas the top non-pretrained machine learning model, GRU, with 91% accuracy and 91% precision.

In [6], the NB classifier excelled in precision and effectiveness, categorizing spam to shield users from harm messages. A user-friendly interface simplifies interaction with the spam detection system, allowing input of messages and language for real-time feedback on spam likelihood. The authors highlighted the importance of considering linguistic diversity in spam detection algorithms, paving the way for robust language-aware spam detection systems. İn [7], the authors applied NLP techniques followed by conventional ML models and given good results. They applied four ML models and achieved good accuracy with SVC. The authors of [8] used the Kaggle SMS Spam Collection dataset and the Bow and TF-IDF weighing feature selection algorithms. For further feature selection, the chi-square matrix was used, and a thorough comparison with cutting-edge techniques was carried out. The results demonstrated the best performance, with our suggested Naïve Bayes model achieving the maximum accuracy. Comparing the study to previous research in the area, it demonstrates how effective our method is in achieving high detection rates and accuracy. In [9], diverse ML & DL techniques applied to address SMS spam detection, utilizing a dataset from UCI LSTM model given good results. Several ML & DL classifiers were used in [10] to identify spam in emails and SMS datasets, improving accuracy. The SVM classifier proved to be the most successful model, and the they found that the suggested model done well compared to earlier models after running tests on the actual dataset. These results imply that SVM is the best option for detection.

In order to solve SMS spam filtration methods, the authors in [11] introduced a hybrid system that made use of both supervised and unsupervised ML models. The hybrid approach was created with better F-measures and spam filtering accuracy in mind.

Most of the previous work used traditional ML classification techniques for SMS spam detection. Class imbalance issue also not resolved in many works. İn this paper, the authors proposing DL techniques with resolution of class imbalance techniques.

III. METHODOLOGY

The proposed method is depicted in Figure 1. The proposed methodology encompasses several key steps aimed at effectively detecting SMS spam using a combination of traditional ML and DL. Firstly, the dataset is collected from Kaggle website [12] consisting of 5570 SMS samples. This dataset served as the basis for training and evaluating the models. The dataset is imbalanced. To address the challenge of class imbalance within the dataset, oversampling technique ADASYN is employed. These techniques help mitigate the imbalance between spam and non-spam samples, ensuring a more balanced representation in the training data. Text representation plays a crucial role in transforming the raw SMS messages into numerical vectors that can be processed by machine learning algorithms. To achieve this, TF-IDF embedding technique is utilized. TF-IDF captures the semantic meaning and contextual information within the SMS messages, enabling more effective analysis and classification. In the realm of machine learning classifiers, a Random Forest classifier is initially trained using the TF-IDF features to establish a baseline performance. This baseline model provides a benchmark for further evaluation. Later several deep learning methods applied for SMS spam detection. Deep learning architectures, specifically RNNs are explored to enhance SMS spam detection. Various RNN architectures, including Simple RNN, LSTM, BI-LSTM, and GRU, are investigated for their effectiveness in identifying spam messages. Later a combined Bi-LSTM + GRU applied for SMS spam detection. After applying all the algorithms, the model's performance is evaluated and best model is selected for SMS spam detection.

A. Dataset and preprocessing

The dataset was obtained from Kaggle [12]. It comprises of 747 spam messages and 4,825 non-spam messages. The total 5,570 messages are in the dataset. The data is a sequential data. After collecting data, it is pre-processed by removing stop words and applying stemming.

B. ML and DL Algorithms

Algorithms for ML and DL are essential for spam identification because they allow automated systems to recognize patterns and traits in spam communications. The proposed model incorporates ML algorithm named Random Forest. Furthermore, DL algorithms such as RNN, LSTM, and GRU are employed in the model. These are discussed below.

C. Random Forest

It is effective for classification and regression tasks. It builds multiple decision trees, introducing randomness for diversity, and combines their predictions for robust and accurate results. Known for handling large datasets, it is widely used in various machine learning applications.

D. RNN

Five diverse classification algorithms RF, DT, SVC, KNN and NB are employed for training models. The pseudo-labels generated from K-means clustering guide the supervised learning process, enabling the models to learn from both labelled and unsupervised data. This integration ensures a comprehensive understanding of fraud patterns.

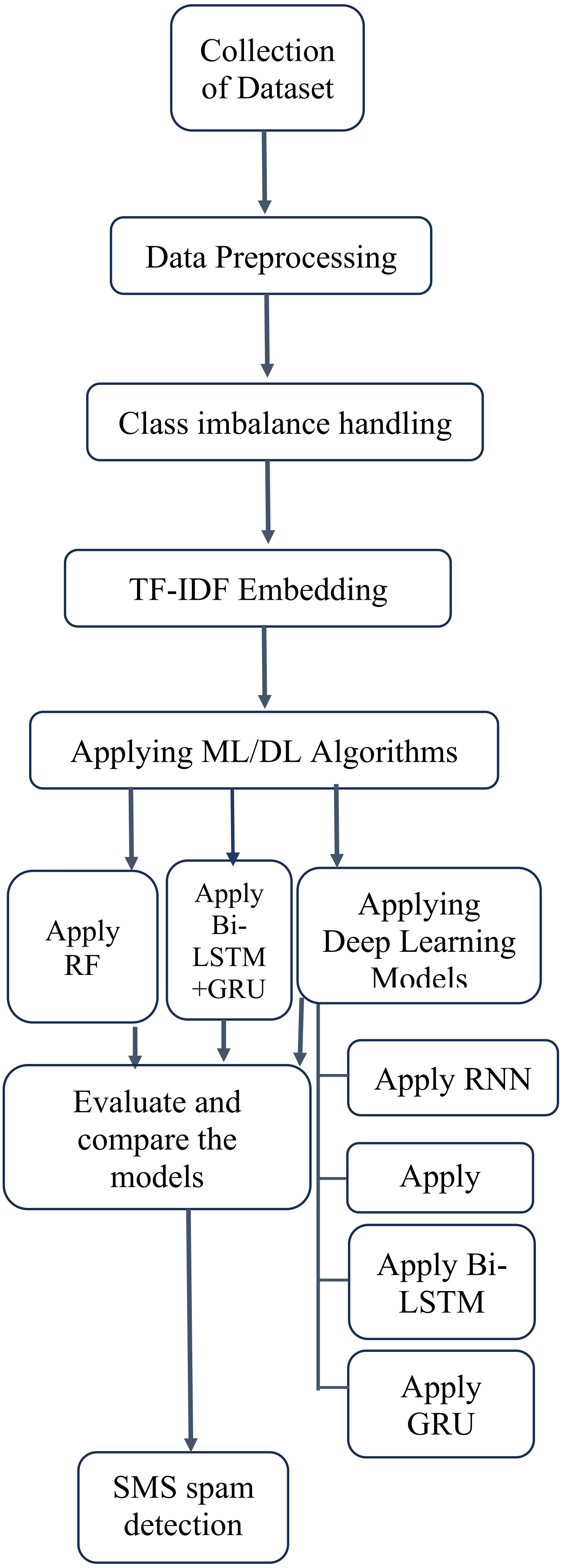

Fig. 1. Proposed Method

E. LSTM, BiLSTM and GRU

LSTM and Bi-LSTM, advanced variations of RNNs, address limitations by incorporating memory cells and bidirectional processing. LSTMs capture long-term dependencies, while Bi-LSTM processes input in both directions, enhancing context understanding. These architectures are applied in tasks like language modelling and sentiment analysis for their effectiveness in handling sequential data. GRU, a variant of RNNs, efficiently tackles long-term dependencies in sequential data with simplified structures and gating mechanisms. Applied in tasks like natural language processing, GRUs offer computational efficiency for sequential modelling.

IV. Experiments and Results

All the expriments are condiced using google Collaboratory. Ther language used for experimentation is python.

A. Handling class imbalance

As the dataset contains imbalanced classes, the first step in the experimentation is handling imbalance issue. For this, we applied sampling technique namely ADASYN. ADASYN adapts to dataset complexity and generates synthetic samples with varying densities in regions with obvious class imbalance. Through more equitable representation of the majority and minority classes, these solutions aim to enhance model performance and ultimately increase classification accuracy in imbalanced datasets. With this technique, the number of samples of spam and non spam messages are equal.

B. Applying TF-IDF

In this work, the feature engineering approach for SMS spam identification includes the application of TF-IDF. By using TF-IDF, the model is better able to differentiate between authentic and spam messages since it is able to extract contextual information and semantic meaning from SMS messages. The suggested strategy is more successful when TF-IDF is used, and the spam detection system's accuracy increases as a result.

C. Applying Random Forest

Later, a conventional ML classifier random forest applied. The accuracy achieved is 92.3%. So, it is observed that RF performed well for from SMS spam detection.

D. Applying Simple RNN

Next, several DL techniques applied for SMS spam detection. Initially, a simple RNN is applied. The proposed RNN architecture contains 3 hidden layers with 180, 100 and 80 neurons in each layer. The number of epochs used in the model is 50. The accuracy achieved with Simple RNN is 93%.

E. Applying LSTM

After applying RNN, we applied LSTM model. The architecture of LSTM model contains 3 hidden layers with 100, 80 and 20 neurons and it consists of 50 epochs. The accuracy achieved with LSTM is 98%.

F. Applying Bi-LSTM

After applying LSTM, we applied Bi-LSTM model. The architecture of Bi-LSTM model contains 4 hidden layers with

100, 80, 40 and 20 neurons and it consists of 50 epochs. The accuracy achieved with Bi-LSTM is 98.2%.

G. Applying GRU

After applying Bi-LSTM, we applied GRU model. The architecture of GRU model contains 3 hidden layers with 120, 60 and 40 neurons and it consists of 50 epochs. The accuracy achieved with GRU is 98.2%.

H. Comparison of models used in the work

After applying all the ML and DL models, a performance comparison among all the models done. The comparison is shown in Table 1 and Figure 2.

TABLE I. RESULTS COMPARISON

Technique	Precision	Recall	Accuracy
RF	98%	87%	92.3%
RNN	96%	95%	93%
LSTM	97%	98%	98%
Bi-LSTM	98%	98.5%	98.2%
GRU	98.3%	97%	98.3%

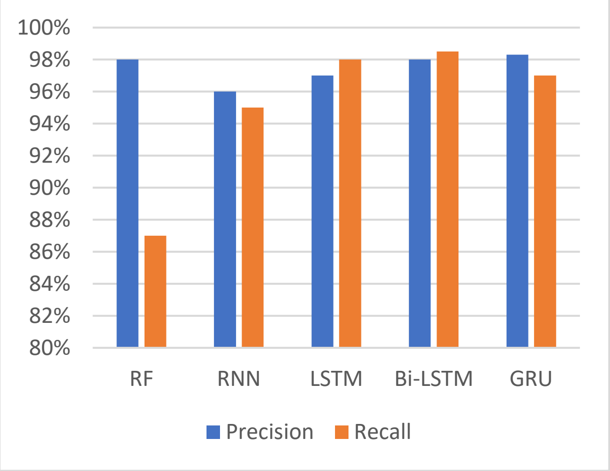

Fig. 2. Precision, Recall comparison

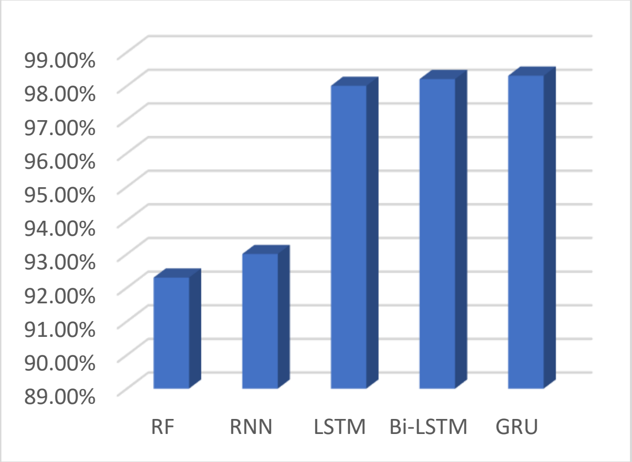

Fig. 3. Accuracy comparison

From figure-2, it is observed that all the applied models given good performance of more than 90%. But, Bi-LSTM given better precision and recall values of 98% and 98.5%. Later, GRU and LSTM given good values for recall and precision.

From figure-3, it is observed that BI-LSTM and GRU given goos accuracy of 98.2%, 98.3% for sms spam detection. Later, LSTM given good accuracy of 98%. Among all, RF given less accuracy of 92.3%.

I. Combined Bi-LSTM and GRU

To further increase the accuracy, the hybrid Gru and Bi-LSTM applied. The reason for selecting these two is that, these models performed well with more than 98% accuracy. A hybrid model combined the strengths of Bi-LSTM and GRU and enhanced the performance. The accuracy achieved with hybrid Bi-LSTM and GRU is 99%.

J. Comparison with previous work

The proposed DL technique is compared with existing works. The comparison is shown in the figure 4 and Table II.

TABLE II. COMPARISON WITH EXISTING WORK

Technique	Accuracy
LSTM [9]	98%
Deep Learning [10]	98.7%
Proposed Model	99%

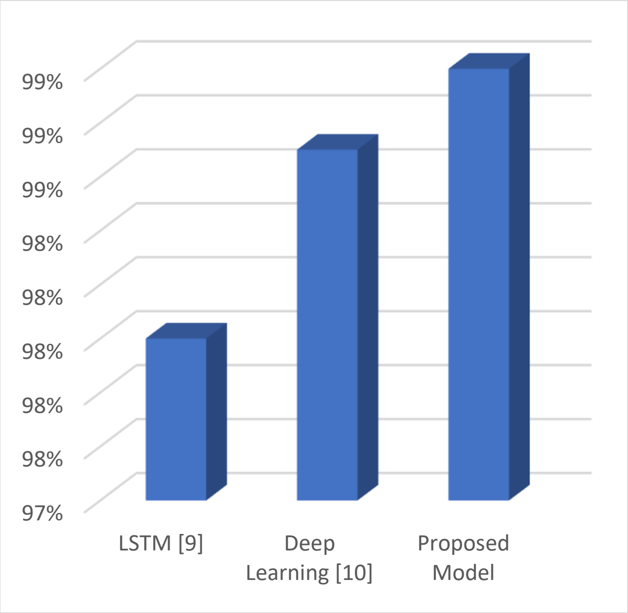

Fig. 4. Comparison with existing work

V. CONCLUSION

This study has addressed the growing concern of SMS spam through effective detection mechanisms. By employing oversampling techniques to tackle class imbalance and utilizing TF-IDF embedding, traditional machine learning with Random Forest has achieved a commendable accuracy of 92.3%. However, the exploration of various recurrent neural network architectures highlighted the superior performance of LSTM variants, with BI-LSTM and GRU reaching an impressive 98.2%, 98.3% accuracy. To further increase accuracy a hybrid Bi-LSTM and GRU applied and achieved accuracy of 99%. This underscores the efficacy of hybrid models in enhancing SMS spam detection, emphasizing the importance of addressing class imbalances and employing advanced embedding techniques for robust results. The findings hold promising implications for combating SMS spam in real-world scenarios. In future, more ensemble methods can tried to further increase accuracy for SMS spam detection.

REFERENCES

[1] A Sireesha et al.,"SMS Spam Detection Using Machine Learning", SJIS, vol. 35, no. 1, pp. 749–754, Apr. 2023.

[2] A. Ghourabi et al., "Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning," Sensors, vol. 23, no. 8. MDPI AG, p. 3861, Apr. 10, 2023.

[3] D. A. Oyeyemi et al., "SMS Spam Detection and Classification to Combat Abuse in Telephone Networks Using Natural Language Processing," Journal of Advances in Mathematics and Computer Science, vol. 38, no. 10. Sciencedomain International, pp. 144–156, Oct. 31, 2023. .

[4] V. Dharani et al., "Spam SMS (or) Email Detection and Classification using Machine Learning," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, Jan. 23, 2023. .

[5] W. Siagian et al., "Improving SMS Spam Detection Through Machine Learning: An Investigation of Feature Extraction and Model Selection Techniques," 2023 International Conference on Information Management and Technology (ICIMTech). IEEE, Aug. 24, 2023.

[6] A. K and S. Halder, "Detection of Multilingual Spam SMS Using NaïveBayes Classifier," 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA). IEEE, Oct. 07, 2023.

[7] T. Jain et al., "SMS Spam Classification Using Machine Learning Techniques," 2022 12th International Conference on Cloud Computing, Data Science &amp; Engineering (Confluence). IEEE, Jan. 27, 2022.

[8] H. Jain et al.,"An Analysis of SMS Spam Detection using Machine Learning Model," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT). IEEE, Jul. 2022.

[9] Gadde, S. et al.: SMS Spam Detection using Machine Learning and Deep Learning Techniques. 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021..

[10] U. Maqsood et al., "An Intelligent Framework Based on Deep Learning for SMS and e-mail Spam Detection," Applied Computational Intelligence and Soft Computing, vol. 2023. Hindawi Limited, pp. 1–16, Sep. 20, 2023.

[11] H. Baaqeel et al., "Hybrid SMS Spam Filtering System Using Machine Learning Techniques," 2020 21st International Arab Conference on Information Technology (ACIT). IEEE, Nov. 28, 2020.

[12] https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset