IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

SMS SPAM DETECTION USING MACHINE LEARNING & DEEP LEARNING APPROACHES.

K.Suryateja Reddy[1], Dr.R.Suresh[2],

[1]Research Scholar, Department of CSE, Sri Venkateswara College of Engineering, Tirupati
[2]Professor , Department of CSE, Sri Venkateswara College of Engineering, Tirupati

Abstract:

SMS, a widely used and rapidly expanding GSM value-added service globally, has increasingly become a target for unwanted messages, commonly referred to as SMS spam. The impact of SMS spam is considerable, as it undermines user trust and poses significant challenges for service providers. This study evaluates the performance of three models for SMS spam classification: Multinomial Naive Bayes (MultinomialNB), a Custom Vector Embedding bidirectional long short-term memory (BiLSTM) model as well as. The models were evaluated on exactness, accuracy, review, and F1-score. The MultinomialNB model achieved 96.23% accuracy, 100% precision, 72.00% recall, and an F1-score of 83.72%. The Custom Vector Embedding model recorded 98.21% accuracy, 97.79% precision, 88.67% recall, and a 93.01% F1-score. The BiLSTM model showed 98.21% accuracy, 97.10% precision, 89.33% recall, and a 93.06% F1-score. Results indicate that the Custom Vector Embedding and BiLSTM models outperform the MultinomialNB model, highlighting the effectiveness of deep learning approaches for SMS spam detection.

Keywords: Multinomial Naïve Bayes, Bi-LSTM, spam.

1. Introduction:

SMS (Short Message Service) has emerged as one of the most prevalent and rapidly expanding value-added services in the GSM (Global System for Mobile Communications) network. Its widespread adoption and convenience have made it a crucial communication tool for billions of users worldwide. However, this popularity has also led to the proliferation of unwanted SMS, commonly known as SMS spam. These unsolicited messages range from marketing promotions to phishing attempts, posing a significant nuisance and security threat to users [1].

The effects of SMS spam are far-reaching. For users, it results in annoyance, potential exposure to scams, and a loss of trust in the messaging service. For service providers, it leads to increased operational costs, network congestion, and a deterioration of service quality. Consequently, there is a pressing need for effective SMS spam classification systems to mitigate these issues and restore user confidence [2].

SMS spam classification involves the use of AI and normal language handling methods to consequently recognize and filter out spam messages from legitimate ones. Various models and algorithms have been developed to tackle this problem, each with its strengths and limitations. The purpose of this study is to compare three distinct models' performance Multinomial Naive Bayes (MultinomialNB), a Custom Vector Embedding bidirectional long short-term memory (BiLSTM) model as well as —in classifying SMS spam. By evaluating these models on employing critical performance measures including F1-score, recall, accuracy, and precision, aim to identify the most effective approach for SMS spam detection. Figure 1 depicts the simple spam classification model.

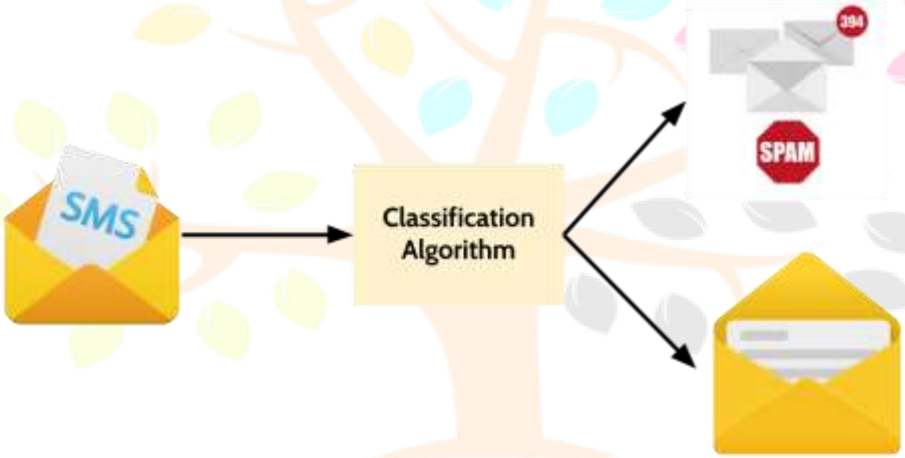


Figure 1. sms spam classification

The findings from this research will provide useful information about the strengths and weaknesses of various SMS spam classification techniques, which contributed to the creation of more durable and effective spam filtering systems[3].

2. Literature Survey:

Delany et al. and other researchers' recent systematic reviews, Abayomi-Alli and others, and Rao and co have highlighted various characterization techniques, highlight extraction, and determination techniques used in the examination and discovery of SMS spam. Sjarif and co. presented an element extraction technique utilizing document frequency is inverse to term frequency (TF-IDF) strategy to recognize pertinent terms. They tested various machine learning models and found that combining TF-IDF with the random forest algorithm surpassed other advanced algorithms, enhancing SMS spam detection[4].

In a similar vein, Cost-touchy tactics were suggested by Lim and Singh [5]. using a stack Various methods for Bayesian networks and multilayer perceptrons (MLPs). Their study demonstrated a notable reduction in misclassification rates and superior performance compared to other machine learning algorithms. Sharaff et al. developed a unique Model for an SMS spam filter that uses the dendritic cell algorithm with krill herd optimization. Theirs experiments revealed to this model was more truthful than different classifiers like Naive Bayes, logistic regression, SVM, and XgBoost[7].

Bosaeed & Co. created a multiple filters system incorporating various machine learning classifiers, including Naive Bayes (NB), SVM, and Naive Bayes Multinomial (NBM). Their study showcased the adaptability to use multiple platforms by implementing their model in full or in part on both mobile and server applications, optimizing computational resources. Alzahrani and Rawat conducted comparing and contrasting various algorithms based on machine learning for detecting SMS spam, finding that the neural network algorithm outperformed other classifiers [7].

Similarly, Theodorus et al. compared eight's performance machine learning classifiers for Text categorization for SMS in Bahasa Indonesia. Other studies have explored the application of machine learning algorithms such as Naive Bayes, neural networks, self-organizing maps, K-nearest neighbors, and the H2O framework. Sisodia and her coworkers used group education., presenting a computerized SMS spam classification framework using various classifiers, including NB, C4.5, SVM, ensemble, KNN, and ID3 methods like Adaboost, random forest, and voting. Their results indicated that ensemble classifiers, particularly those based on random forest, achieved the highest accuracy[8].

In their research, Gadde et al. and Al-Bataineh and Kaur47 16] looked into how deep learning techniques could be used, specifically LSTM, for detecting SMS spam. Gadde et al. employed three different word embedding methods: count vectorizer, TF-IDF, and hashing vectorizer. They compared the performance of LSTM with several state-of-the-art machine learning techniques. Conversely, Al-Bataineh and Kaur focused on demonstrating the a clonal selection algorithm and the robustness of LSTM architectures for text classification. Their evaluation, conducted on three datasets and benchmarked against leading ML classifiers, revealed that their model surpassed others in precision, accuracy, recall, F1 score, and computational time[12]. Similarly, Roy et al.[8] proposed a profound learning approach consolidating both convolutional brain organizations and LSTM for SMS spam classification. They highlighted the superior performance of these deep learning models in three different configurations, and it should be noted that the addition of regularization parameters like dropout made the classification accuracy even better. One more eminent concentrate by Xia and Chen[9] presented a superior Secret Markov Model (Well) that used weighted element and name words. Their findings indicated that this enhanced HMM outperformed LSTM models in terms of accuracy and computational efficiency.

3. Proposed Methodology

The steps involved in the classification process are described in this section, conducted using various machine learning algorithms. Figure 1 provides a blueprint flowchart for the recommended SMS sifting framework. The section is organized into three subsections:

1)Datasets 2) Preprocessing 3) Classification (including experiment and assessment)

3.1. **Dataset**: A total of 5240 SMS messages from the UCI Machine Learning repository are included in the dataset. gathered in 2022. SMS Spam Data Set Collection from UCIMachineLearningRepository,http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection

3.2. **Pre-processing stage:** Removing messages in their native tongues from the SMS database required a thorough search to remove duplicates. The final dataset for this study included 4,420 unique SMS messages, with 2,453 categorized as spam and 1,967 as ham. To enhance the model's effectiveness, we removed all punctuation marks, prepositions, and short words (those with two or fewer letters). Unnecessary elements such as stopwords and exclamations were also eliminated. Given the prevalence of non-standard abbreviations in SMS, our preprocessing approach was designed to be flexible. Figure 4 displays the first few preprocessed training documents along with the token breakdown for each SMS. After pre-processing, figure-2 shows the class distribution.

Figure 2. Class distribution of SPAM and HAM

3.3 Proposed Models

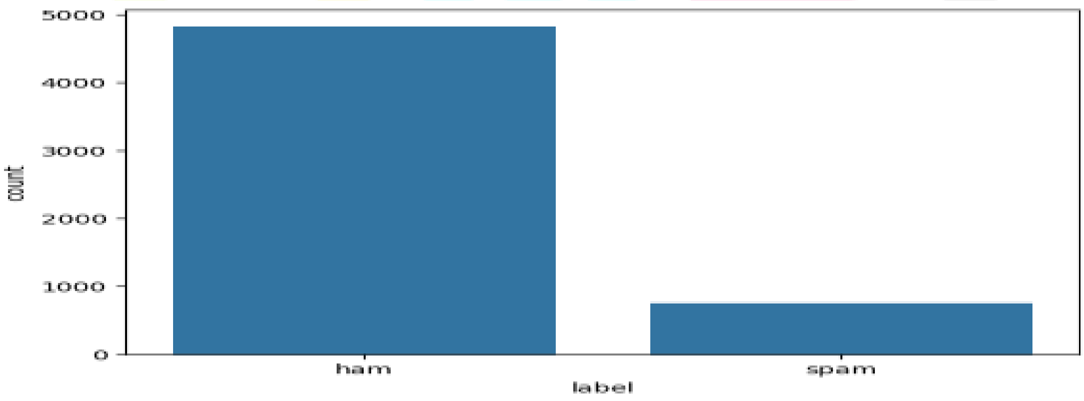
In this work, we have proposed three models

1)MultinomialNB Model

2)Custom-Vec-Embedding Model

3)Bidirectional-LSTM Model

3.3.1. MultinomialNB Model: For text classification tasks, the Multinomial Naive Bayes (MultinomialNB) model is a popular algorithm. It is particularly well-suited for handling data with multiple classes, making it a common choice for document classification and spam filtering. MultinomialNB assumes that features follow a multinomial distribution, which is often the case with text data where features represent word counts or frequencies. Despite its simplicity, MultinomialNB often performs well in practice, especially when dealing with large, high-dimensional datasets such as those found in natural language processing tasks. It's widely used due to its efficiency, scalability, and ease of implementation. Figure 3 shows the classification model using machine learning.



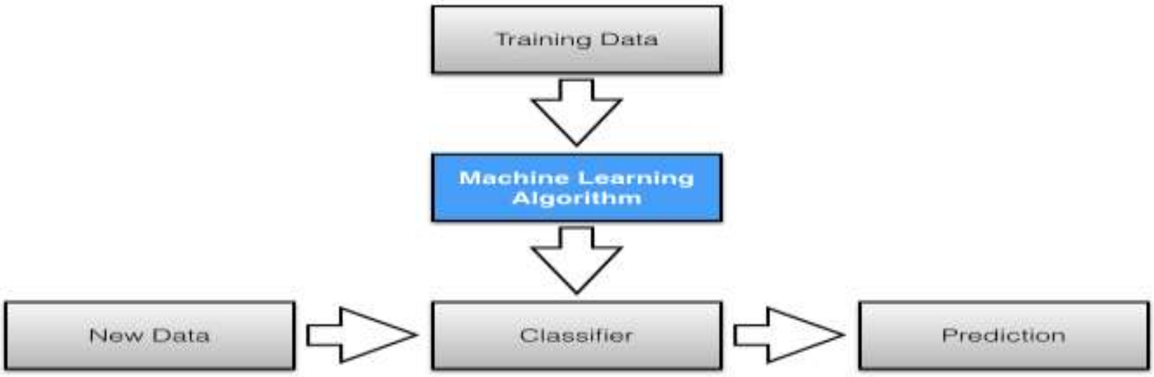


Figure 3. Classification model using MultinomialNB

3.3.2. Custom-Vec-Embedding Model: The Custom-Vec-Embedding model is a tailored embedding technique designed specifically for SMS spam classification tasks. Unlike traditional word embeddings, Custom-Vec-Embedding captures the unique linguistic characteristics and context of SMS messages, enhancing the model's ability to differentiate between spam and non-spam messages effectively. By customizing the embedding process to suit the nuances of SMS language, this model can better represent the semantics and syntax of short text messages, leading to improved classification accuracy. Its specialized approach addresses challenges such as abbreviations, slang, and non-standard grammar commonly found in SMS data, making it well-suited for this classification task.

3.3.3: Bidirectional-LSTM Model: We opted for the BiLSTM model due to its strong track record in various text classification tasks. This model, which utilizes bidirectional long short-term memory networks, excels in learning intricate patterns within sequences and effectively discerning decision boundaries between classes. Previous studies have demonstrated its prowess in tasks like sentiment analysis, where its ability to capture contextual information and dependencies in both forward and backward directions proves highly beneficial. Figure 4 shows the proposed Bi-LSTM Model.

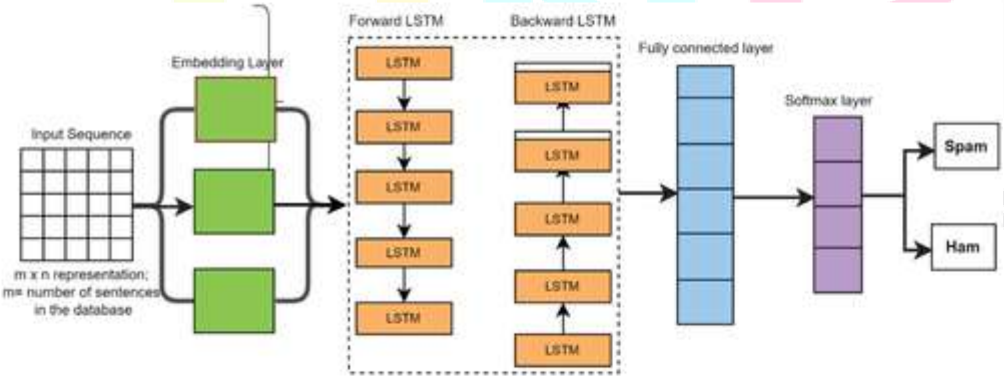


Figure 4. The architecture of our BiLSTM model

4. Results and Discussions

We explore the adequacy of different AI models for SMS spam classification. The primary objective is to compare the performance of three distinct models: Multinomial Naive Bayes (Multinomial NB), a Custom

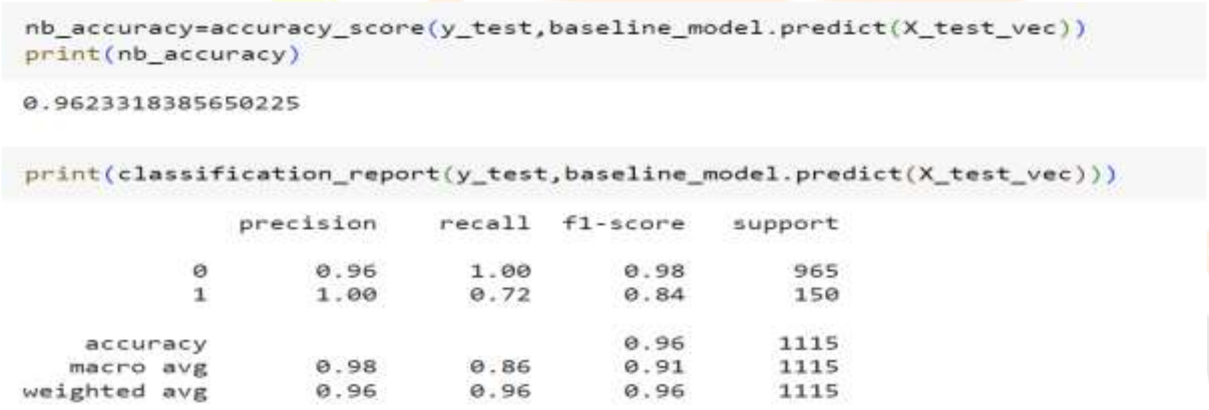
Vector Embedding a different model in addition to the bidirectional Long Short Term Memory (BiLSTM) model. Each model is evaluated according to its F1-score, recall, accuracy, and precision.

The MultinomialNB model accomplished an exactness of 96.23%, a precision of 100%, a memory of 72.00%, and an F1-score of 83.72%. While this model demonstrated perfect precision, its recall rate indicates a relatively higher false negative rate compared to the other models.

The Custom Vector Embedding model significantly improved performance with an accuracy of 98.21%, a precision of 97.79%, a recall of 88.67%, and an F1-score of 93.01%. This model balanced both high precision and recall, showcasing its robustness in identifying spam messages.

The Bidirectional LSTM model exhibited similar performance to the Custom Vector Embedding model, with an accuracy of 98.21%, a precision of 97.10%, a recall of 89.33%, and an F1-score of 93.06%. The BiLSTM model leverages sequential information effectively, resulting in high recall and F1-score.

Our comparative analysis reveals that both the Custom Vector Embedding and Bidirectional LSTM models outperform the MultinomialNB model in terms of balanced precision and recall. These findings suggest that deep learning models, particularly those incorporating sequential processing, offer superior performance in SMS spam detection. This research underscores the importance of model selection in developing effective spam detection systems and provides a benchmark for future studies in this domain. Figure.5 demonstrates how MultinomialNB performs.



```
nb_accuracy=accuracy_score(y_test,baseline_model.predict(X_test_vec))
print(nb_accuracy)
```

```
0.9623318385650225
```

```
print(classification_report(y_test,baseline_model.predict(X_test_vec)))
```

```
              precision    recall  f1-score   support

           0       0.96      1.00      0.98       965
           1       1.00      0.72      0.84       150

    accuracy                           0.96      1115
   macro avg       0.98      0.86      0.91      1115
weighted avg       0.96      0.96      0.96      1115
```

Figure. 5 Performance of MultinomialNB

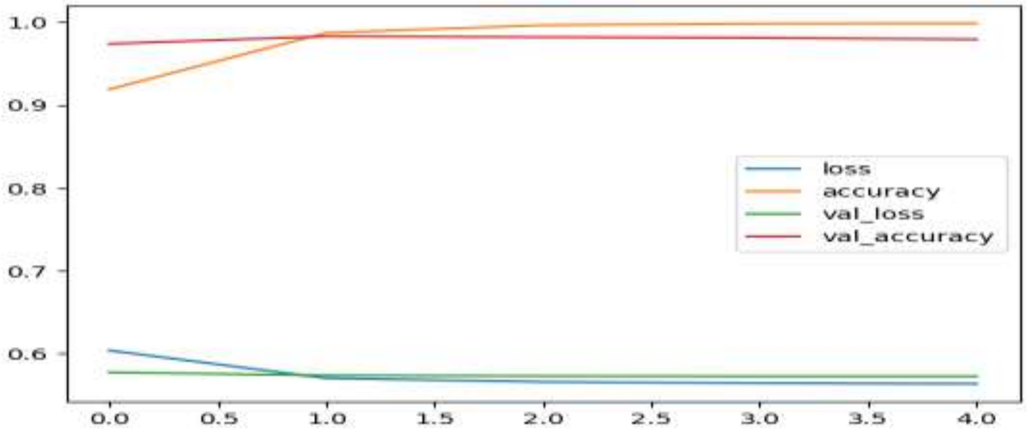


Figure. 6 Shows the performance of Custom-Vec-Embedding Model

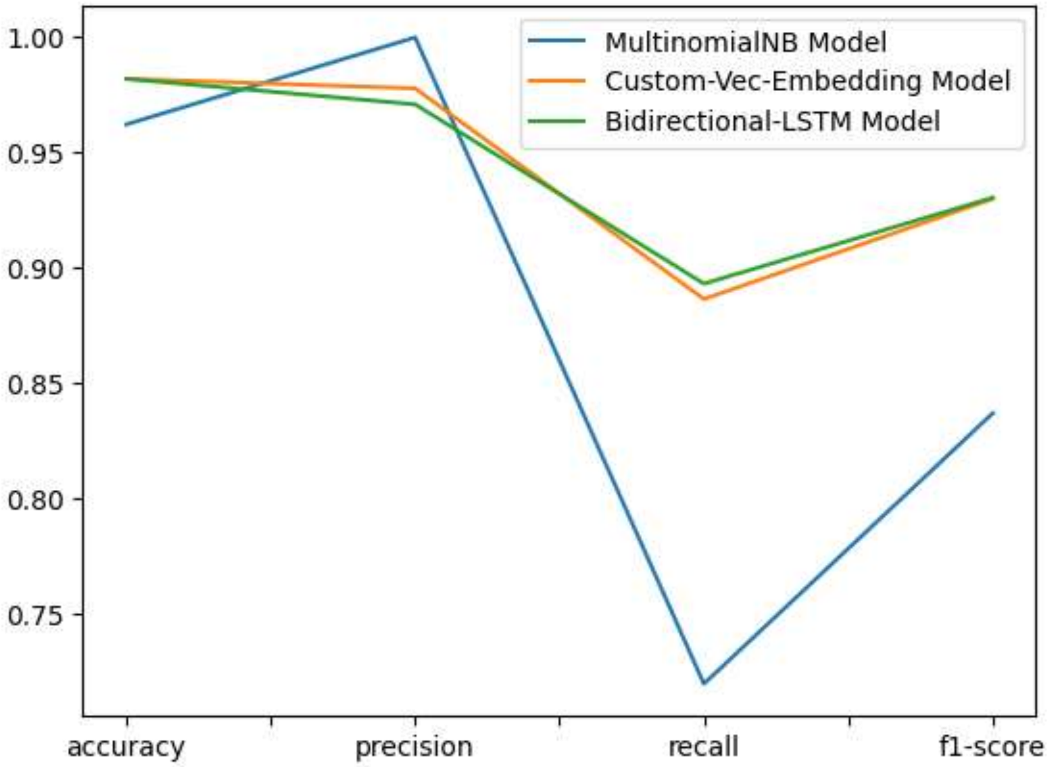
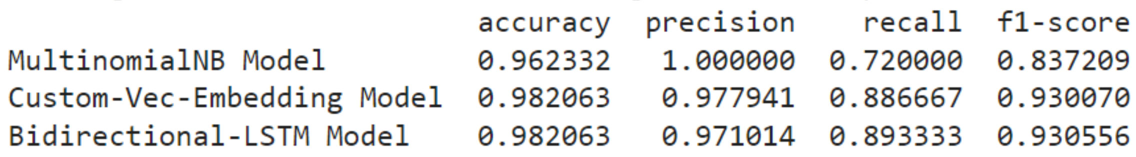


Figure 7 shows the performance of three models



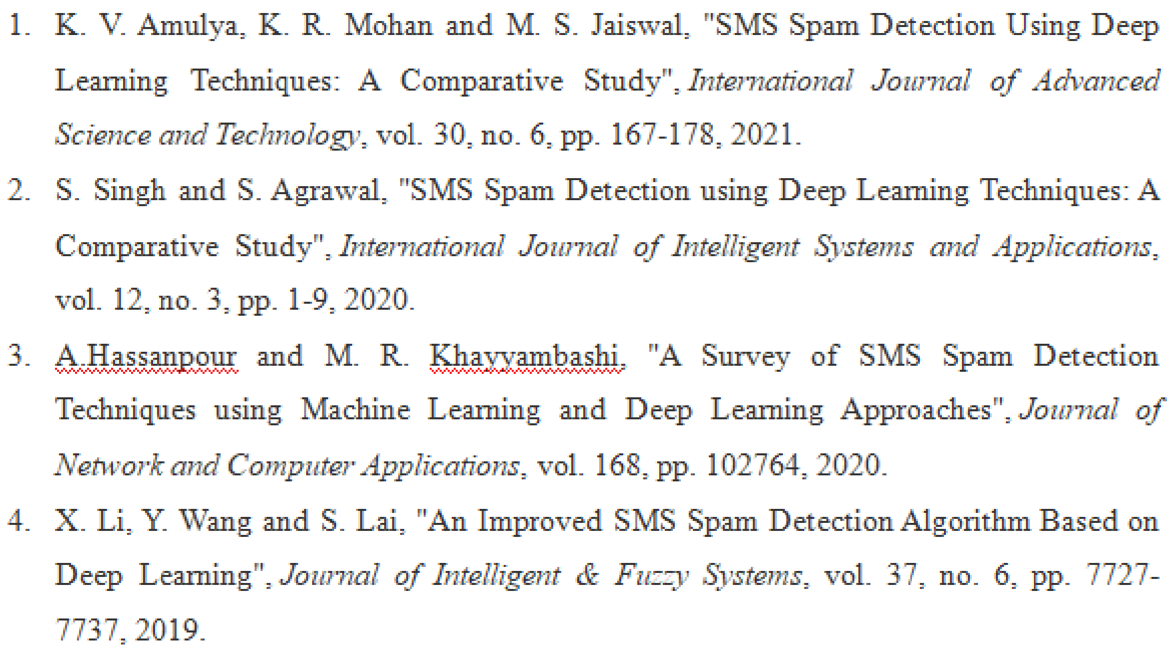
```
                           accuracy  precision    recall  f1-score
MultinomialNB Model        0.962332   1.000000  0.720000  0.837209
Custom-Vec-Embedding Model 0.982063   0.977941  0.886667  0.930070
Bidirectional-LSTM Model   0.982063   0.971014  0.893333  0.930556
```

Figure 8 shows the performance comparison of three models.

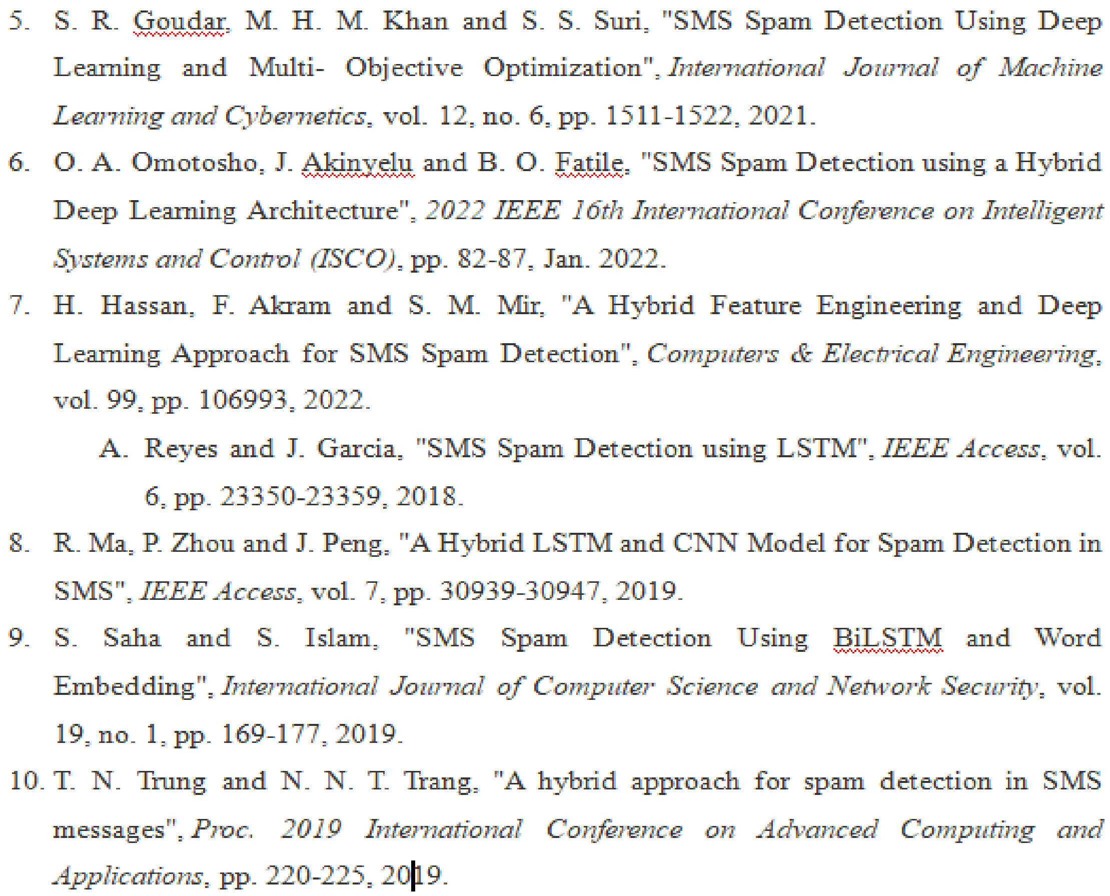
5. Conclusion:

This research provides a comparative analysis of three machine learning models for SMS spam classification: Multinomial Naive Bayes (MultinomialNB), a Custom Vector Embedding Additionally available is a Bidirectional Long Short-Term Memory (BiLSTM) model.The evaluation, contingent upon F1-score, review, exactness, and accuracy, reveals significant differences in performance among these models.The MultinomialNB model, while achieving perfect precision (100%), showed limitations in recall (72.00%), resulting in a higher rate of false negatives. In contrast, the Custom Vector Embedding model demonstrated a substantial improvement has a 98.21% exactness, 97.79% accuracy, 88.67% review, and 93.01% F1-score. This specific model effectively balanced high precision and recall, indicating robust performance in identifying spam messages. The Bidirectional LSTM model performed comparably to the Custom Vector Embedding model, with an accuracy of 98.21%, precision of 97.10%, recall of 89.33%, and an F1-score of 93.06%. The BiLSTM model's ability to leverage sequential information contributed to its high recall and overall F1-score. Our findings suggest that both the Custom Vector Embedding and Bidirectional LSTM models outperform the MultinomialNB model, particularly in achieving a balanced precision and recall. These results highlight the superior performance of deep learning models, especially those incorporating sequential processing, in SMS spam detection. This study underscores the importance of selecting appropriate models for developing effective spam detection systems and provides a benchmark for future research in this field.

6. References:



1. K. V. Amulya, K. R. Mohan and M. S. Jaiswal, "SMS Spam Detection Using Deep Learning Techniques: A Comparative Study", *International Journal of Advanced Science and Technology*, vol. 30, no. 6, pp. 167-178, 2021.
2. S. Singh and S. Agrawal, "SMS Spam Detection using Deep Learning Techniques: A Comparative Study", *International Journal of Intelligent Systems and Applications*, vol. 12, no. 3, pp. 1-9, 2020.
3. A.Hassanpour and M. R. Khayyambashi, "A Survey of SMS Spam Detection Techniques using Machine Learning and Deep Learning Approaches", *Journal of Network and Computer Applications*, vol. 168, pp. 102764, 2020.
4. X. Li, Y. Wang and S. Lai, "An Improved SMS Spam Detection Algorithm Based on Deep Learning", *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 6, pp. 7727-7737, 2019.



5. S. R. Goudar, M. H. M. Khan and S. S. Suri, "SMS Spam Detection Using Deep Learning and Multi- Objective Optimization", *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 6, pp. 1511-1522, 2021.
6. O. A. Omotosho, J. Akinyelu and B. O. Fatile, "SMS Spam Detection using a Hybrid Deep Learning Architecture", *2022 IEEE 16th International Conference on Intelligent Systems and Control (ISCO)*, pp. 82-87, Jan. 2022.
7. H. Hassan, F. Akram and S. M. Mir, "A Hybrid Feature Engineering and Deep Learning Approach for SMS Spam Detection", *Computers & Electrical Engineering*, vol. 99, pp. 106993, 2022.
    A. Reyes and J. Garcia, "SMS Spam Detection using LSTM", *IEEE Access*, vol. 6, pp. 23350-23359, 2018.
8. R. Ma, P. Zhou and J. Peng, "A Hybrid LSTM and CNN Model for Spam Detection in SMS", *IEEE Access*, vol. 7, pp. 30939-30947, 2019.
9. S. Saha and S. Islam, "SMS Spam Detection Using BiLSTM and Word Embedding", *International Journal of Computer Science and Network Security*, vol. 19, no. 1, pp. 169-177, 2019.
10. T. N. Trung and N. N. T. Trang, "A hybrid approach for spam detection in SMS messages", *Proc. 2019 International Conference on Advanced Computing and Applications*, pp. 220-225, 2019.