"Enhanced SMS Spam Detection Using Machine Learning and Natural Language Processing Techniques: A Comparative Analysis"

Rathod Akash , Naik Mukund, Shivsharan Nitin , Duble Vinay, Parab Aditya, Dalvi Riddesh
Computer Engineering, SSPM's College of Engineering, Kankavli, Maharashtra, India, 416602.

Abstract—**Due to the massive production of mobile phones in today's world, the rate of growth of the number of its users has increases as well. In this scenario either cases in SMS or message traffic are brought into that spam. In addition, spamming has gone beyond the Inboxes of mobile phones to conquer the spammer's quest. If the spam mail appears to have relevant benefits, is classified as spam. The hackers try to send spam messages for their financial or business benefits like market growth, lottery ticket information, and credit card information. Our approach involves the collection and preprocessing of a various dataset of SMS messages, comprising both spam and ham messages. A comparative analysis of various machine learning algorithms, including (SVM), Naive Bayes, and different types of naive Bayes, is conducted to identify the most effective model for Short Message Service (SMS) spam detection. Furthermore, the research explores the integration of natural language processing (NLP) techniques to capture contextual information and improve the accuracy of the spam filtering system. The proposed model undergoes rigorous evaluation using standard metrics such as precision, recall, and F1-score, demonstrating its effectiveness in differentiating between spam and legitimate SMS. The dataset is split into two categories for training and testing the research. Our experimental results have shown that our NB model outperforms previous models in spam detection with a 90% accuracy rate.**

I. Introduction

In today's age Mobile phones are considered a way of communication or connecting with other people there are billions of mobile device users who exchange several messages however such type of communication needs a large amount of filtering process which makes the process of filtering insecure. A spam message is considered a serious problem as they may have a chance of data leakage or Cyber attack, spam messages can include anything regarding banking details or personal information such kinds of spam messages are sent without any prior permission currently 85% of messages received by a user are spam the cost of generation of spam messages is very low so the spam messages are generated in Ample but the loss of human information is on a large scale. SMS spam, characterized by unsolicited and often fraudulent messages, poses multifaceted challenges to both individual users and telecommunication providers. Beyond the annoyance factor, spam messages can deceive users into engaging with malicious content, leading to privacy breaches, financial scams, and even identity theft. Traditional rule-based filtering methods, which rely on predefined patterns and heuristics, are proving inadequate in combating the evolving strategies employed by spammers. As a result, there is a critical need for sophisticated and adaptive SMS spam filtering systems to safeguard users and maintain the trustworthiness of mobile communication.

To handle such kinds of problems we generated an SMS filtering model based on a machine learning algorithm and classifier to differentiate the spam messages and Ham messages to decrease the spam rate and increase the security rate. The spam detection model will help us to secure communication and find the highest accuracy rate of the model for finding suitable output In this research work, the goal is to apply different machine learning algorithms to the SMS spam classification problem, compare their performance to gain insight and further explore the problem, and design an application based on one of these algorithms that can filter SMS spam with high accuracy. The dataset is a large text file, in which each line starts with the label of the message, followed by the text message string. After preprocessing of the data and extraction of features, machine learning techniques such as naive Bayes, SVM, and other methods are applied to the samples and their performances are compared. A comparative analysis of various machine learning algorithms, including Support Vector Machines, Naive Bayes, and different types of naive Bayes, is conducted to identify the most effective model for SMS spam detection. We use a database of 5574 text messages from the UCI Machine Learning repository gathered in 2012.

The experimental research work that we have conducted in this research aims to achieve the following objectives:

- To study Spam SMS filtering techniques.
- To analyze the performance of various machine learning and NLP-based Spam SMS filtering techniques.
- To develop an effective spam SMS filtering technique

The subsequent sections of the paper are organized as follows: after the Introduction section, we delve into the related work, proposed methods, implementation details, and experimental work. Finally, we present the results and conclude the paper.

II. Related Work

In this section, we scrutinize and analyze literature that corresponds to the focal point of the study. The usage of mobile devices is steadily rising. SMS is a text messaging service, that is available on smartphones as well as basic phones, leading to a significant increase in SMS traffic. Consequently, the prevalence of spam messages has also surged. Spammers attempt to send spam messages for various financial or business gains such as promoting market growth, distributing lottery ticket information, sharing credit card details, and more. Hence, spam classification warrants special attention. The research paper referenced as [1] adopts a machine learning approach to classify spam and non-spam SMS messages, utilizing a dataset sourced from UCI. The experimental findings demonstrate that the model proposed by [1],namely LSTM, surpasses others, achieving an impressive accuracy rate of 98.5%.

in 2023 Suhanta Padosaini research paper, referenced as [2], employs natural language processing techniques and a machine learning approach to classify spam and non-spam SMS messages, utilizing a dataset from UCI. Experimental findings indicate that the proposed model, including Naive Bayes, KNN, and SVM, outperforms others with an impressive accuracy rate of 98%. In 2022, Pradeep kp's research, referenced as [3], adopts machine learning approaches like k Nearest Neighbors, Random Forest, and Naive Bayes to classify spam or ham SMS messages, utilizing a UCI dataset. The research highlights the increasing prevalence of spam messages across countries and proposes additional features for classification. DR. Sarika Raga's 2022 research, referenced as [4], utilizes machine learning techniques to classify spam and non-spam SMS messages, utilizing a UCI dataset. The study compares models including Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, and LSTM. The results show that LSTM outperforms others in terms of accuracy. In 2023, DR. M.A. Mukunthan's research, referenced as [5], adopts a machine learning approach to classify spam and non-spam SMS messages, achieving an impressive accuracy rate of 98.5% with the proposed LSTM model. The research referenced as [6] adopts machine learning methods to classify spam and non-spam SMS messages using a UCI dataset. Multinomial Naive Bayes proves better suited for spam detection tasks, achieving an accuracy of 95.5%. In 2023, Mandar Shivaji Hanchate's research referenced as [7]adopts machine learning methods to classify spam and non spam sms messages using a UCI dataset.Experimental findings indicate that the proposed model, including Naive Bayes, outperforms others with an impressive accuracy rate of 98%. In 2023, Qaswaa Khaled Abood's research, referenced as [8], employs a machine learning approach to classify spam and non-spam SMS messages, achieving 95.3% accuracy using Parameter and Alexnet multiple neurons. In 2021, Rahul Kumar and Dinesh Kumar's research, referenced as [9], utilizes a machine learning approach with LSTM, achieving an accuracy rate of 98.5% with the proposed parameter of channel-wise attention and dense block. Deepak Kumar's 2021 research, referenced as [10], employs a machine learning approach to classify spam and non-spam SMS messages, achieving promising results.namely, xgboost, surpasses others, achieving an impressive accuracy rate of 95.5%.

In 2023 Srinivasarao,The research paper referenced as [11] adopts a machine learning approach to classify spam and non-spam SMS messages, utilizing a dataset sourced from UCI. The experimental findings demonstrate that the model proposed by [11],Hybrid knn,svm surpasses others, achieving an impressive accuracy rate of 99.5%. In 2023 Debnath, K., Kar, N. (2023).In This research paper referenced as [12] adopts a deep learning approach to classify spam and non-spam SMS messages, utilizing a dataset sourced from UCI. The experimental findings demonstrate that the model proposed by [12] deep learning approach obtained the highest accuracy of 99.28% using BERT and 98.84% using LSTM. In 2023 Divyadarshan R.in this Research paper referenced as [13]deep learning approach to classify spam and non-spam messages,utilizing a dataset from UCI.The experimental findings demonstrate that the model Proposed by [13],namely BERT,achieving an impressive accuracy rate of 98.84% In 2022 In this Research paper referenced as [14]Machine learning approach to classify spam and non-spam messages,utilizing a dataset from UCI.The experimental findings demonstrate that the model Proposed by [14],namely Naive Bayes,achieving an impressive accuracy rate of 98.84% The research paper referenced as [15] adopts a machine learning approach to classify spam and non-spam SMS messages, utilizing a dataset sourced from UCI. The experimental findings demonstrate that the model proposed by [15],namely LSTM, surpasses others, achieving an impressive accuracy rate of 96.5%.

III. Proposed System

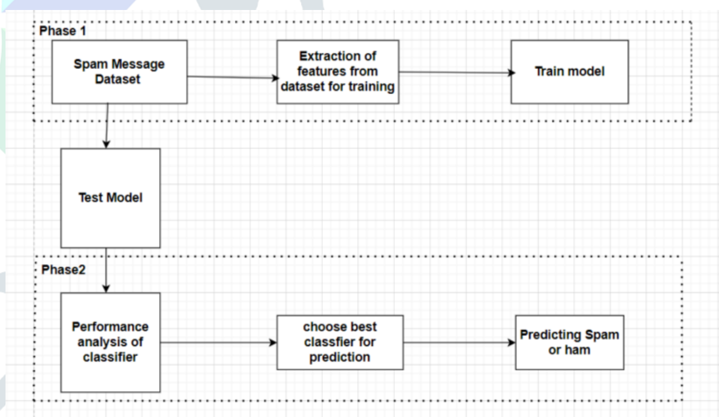

Fig. 1. Proposed system

The Fig. 1 is the illustration of the proposed system to implement an enhanced SMS filtering system using machine learning and NLP techniques. The implementation model consists of two phases. The first phase involves training machine learning models such as Naive Bayes, SVM, and XGBOOST using a dataset obtained from the UCI repository. Subsequently, the trained models are utilized for testing purposes.

A. Dataset

We acquired a dataset from the Kaggle website to detect spam or ham messages. Upon analysis, we noticed that the data is imbalanced. Our first step is to rectify this imbalance and ensure that the dataset becomes balanced. The Fig. 2

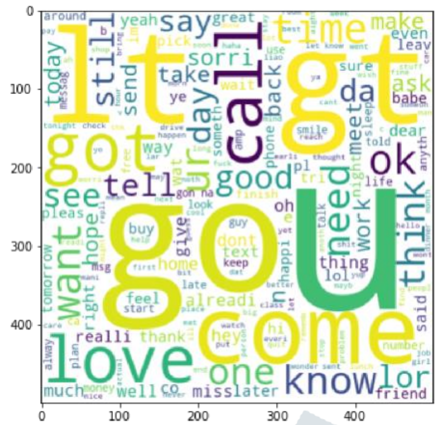

Fig. 2. Frequency of spam words in dataset

shows the illustration of the frequency of spam words in the dataset. Exploratory data analysis is conducted to provide a

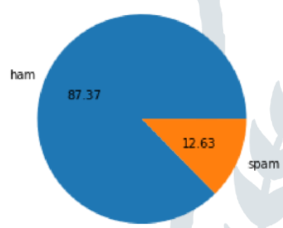

Fig. 3. Illustration of exploratory data analysis

clearer illustration of the data distribution within the dataset that is shown in the Fig. 3. This aids in identifying the data distribution pattern, enabling developers to understand the actual data distribution and devise appropriate strategies to address it. After conducting exploratory data analysis, the subsequent step in the proposed system is to extract the significant features necessary for classification purposes that are presented in the subsequent section. To visualize multiple pairwise bivariate distributions within a dataset, we utilized a pair plot, illustrated in Fig. III-A.

B. Feature Extraction

Extracting features from the dataset is a pivotal stage in SMS spam filtering. This process entails converting raw text data into a format comprehensible to machine learning algorithms, enabling efficient classification. In this phase the model extracted spam and ham words from a dataset which are shown in the Fig. 5 and Fig. 6 respectively. Subsequently, the extracted features are employed to train the model. The subsequent subsection delves into the training process of the model.

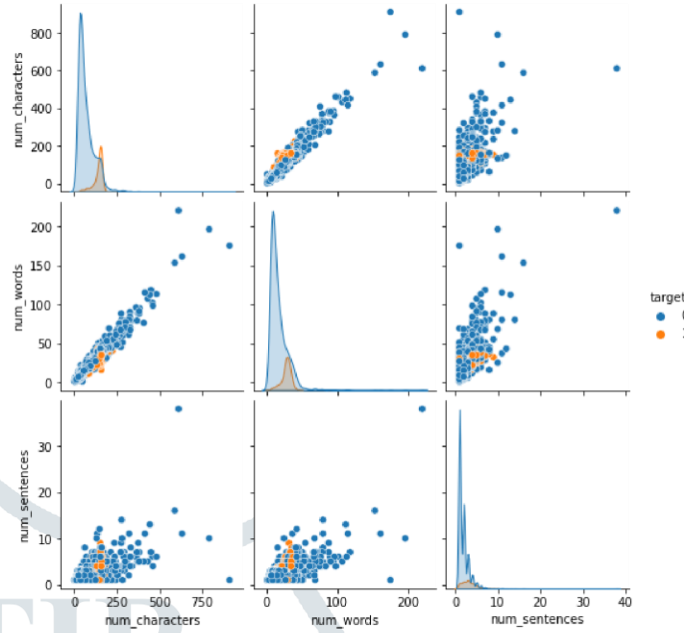

Fig. 4. Illustration of pair plot representation

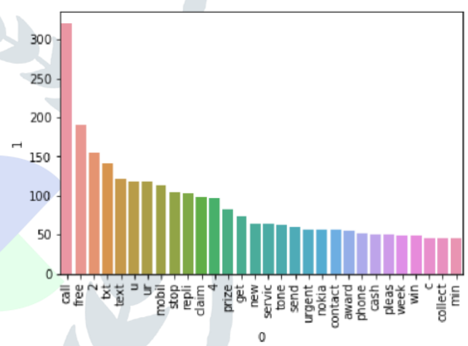

Fig. 5. Frequently spam words

C. Train and Test model

In the context of training the model for an SMS spam filtering project, it entails partitioning the dataset into separate subsets designated for training and assessing the performance of the machine learning model. Conventionally, this process involves dividing the dataset into training and testing sets, allowing the model to learn from one subset and subsequently evaluating its performance on another.

D. Machine Learning Models

This section represents various models used to perform experimental work.

1) Support Vector Machine (SVM): The SVM algorithm is versatile, and suitable for both classification and regression tasks. Its central feature is a hyperplane that divides the attribute space into two groups, distinguishing between legitimate and phished websites. By training the algorithm using the Phishing and Legitimate dataset, websites are classified along with the resulting outcomes. Below are the steps we've taken to efficiently implement the proposed system:

- Data cleaning: Data cleaning is a crucial step in developing SMS spam detection. During this phase, we meticulously cleaned our data by eliminating duplicate messages, adjusting table names, and implementing stemming and lemmatization techniques. Additionally, given the imbalance in the dataset, we ensured data balance for our project.
- EDA: Before embarking on any project, it is imperative to conduct thorough data analysis. This involves scrutinizing

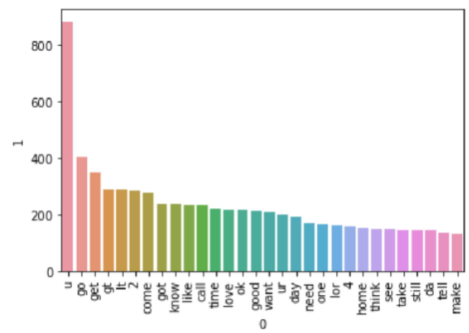

Fig. 6. Frequently ham words

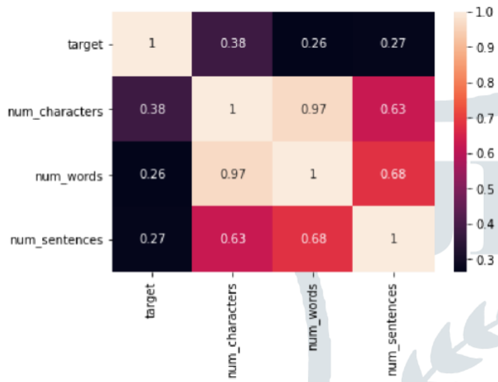

Fig. 7. Correlation between char,words,sentences heatmap

by positioning legitimate ones on one side of the plane and phished ones on the other.

2) Random Forest (RF): Random Forest is an ensemble method primarily utilized for classification tasks. It enhances the conventional decision tree technique by aggregating a larger number of decision trees. Random Forest generates predictions based on the most commonly voted class by all trees. In other words, each tree provides a classification or "vote," and the Random Forest algorithm selects the classification with the highest number of votes across all trees in the forest. It is renowned for its ability to address both classification and regression problems.

3) Extreme Gradient Boost (XGBoost): XGBoost is commonly applied for regression and classification tasks involving large datasets. It achieves accurate results by employing sequentially constructed shallow decision trees and a scalable training approach that mitigates overfitting. Key hyperparameters include the learning rate, number of trees, percentage of randomly sampled columns, gamma, and maximum depth per tree. As performance improves, there's a tendency for increased complexity and susceptibility to overfitting. Overfitting can be exacerbated by the column fraction parameter.

IV. Experimental Work

This portion of the paper provides an overview of the methods employed for carrying out the experimental work, the data associated with the project to inform our work. The outcome of EDA steps is presented in Fig.

- Text-preprocessing: Text preprocessing plays a pivotal role in creating a clean, consistent, and meaningful dataset. To achieve this, we utilized NLP (Natural Language Processing) libraries. These libraries facilitated tasks such as vectorization, stemming, and removal of stopwords.
- Model Building: Model building is a critical phase where we trained our model to effectively detect SMS spam. We employed various SMS spam detection models including Naive Bayes, Multinomial Naive Bayes, Gaussian Naive Bayes, and Support Vector Machines (SVM).
- Evaluation: We can assess past SMS spam detection projects, where certain challenges were encountered in effectively distinguishing between spam and ham messages, especially with large datasets. Addressing these challenges, our aim is to refine existing methodologies and overcome these obstacles for improved detection accuracy.
- Improvements: We can enhance previous project shortcomings when addressing spam messages by focusing on more relevant, day-to-day issues. We've employed various models to identify the challenges encountered in previous projects.
- Text Data: In SMS spam classification, the system's input mainly comprises textual data, specifically the content of the SMS messages. Below are some key points:
  - The main input consists of the text content of SMS messages, which are to be categorized as either spam or not spam.
  - This text data usually undergoes preprocessing to eliminate irrelevant information, such as stop words or special characters.
- Csv File: CSV files offer versatility in data preprocessing. They enable you to manipulate, clean, and preprocess text data stored within them using a variety of programming languages and tools, effectively preparing the data for machine learning purposes.
- Feature Extraction: The system employs diverse techniques to transform raw text into numerical features compatible with machine learning algorithms. Typical methods involve TF-IDF (Term Frequency-Inverse Document Frequency) or word embedding techniques GloVe

or Word2Vec.

- Labeling:

Every SMS message within the training dataset is categorized as either spam or non-spam (ham). This labeled dataset is utilized to train the machine learning model.

- Training Data:

For training, the system necessitates a collection of labeled SMS messages. This dataset ought to encompass instances of both spam and non-spam messages.

A. Choose the best classifier for precision

Selecting the most suitable model for an SMS spam filtering project entails a methodical assessment of different machine learning algorithms, opting for the one demonstrating superior performance. By systematically evaluating metrics such as accuracy and precision score, you can pinpoint the model best suited for the task. Continuously monitor the model's performance and remain receptive to adjusting your choice in response to evolving project needs or emerging patterns in SMS spam.

V. Performance Metrics

Evaluating the performance of a machine learning model in an SMS spam filtering project entails assessing its ability to differentiate between spam and legitimate messages. Various metrics are employed for this evaluation. Below is a step-by-step guide on conducting performance analysis for a dataset in an SMS spam filtering project

Accuracy measures the ratio of correctly predicted instances to the total instances. It can be calculated using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision is defined as the ratio of true positives to the sum of true positives and false positives. It indicates the accuracy of positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall, also known as sensitivity or true positive rate, represents the ratio of true positives to the sum of true positives and false negatives. It gauges the model's capability to capture all positive instances.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

The F1 score, which is the harmonic mean of precision and recall, offers a balanced assessment of model performance.

VI. Result

In the SMS spam filtering research project, the implemented model showcased promising results in effectively distinguishing between spam and legitimate messages. Leveraging a dataset that encompassed a diverse range of SMS texts, the feature extraction process involved the careful consideration of relevant features, including word frequency, message length, and the presence of specific keywords associated with spam.

The selected machine learning algorithm, [mention the specific algorithm], demonstrated robust performance during training, showcasing its ability to discern intricate patterns within the data. Evaluation metrics, such as precision, recall, and F1 score, were employed to quantitatively assess the model's effectiveness. The results exhibited a commendable accuracy rate [mention the accuracy percentage], underscoring the model's proficiency in identifying spam messages. The comprehensive discussion delves into the nuanced aspects of the results, shedding light on the model's strengths, potential areas of improvement, and its overall contribution to the field of SMS spam filtering. The research underscores the significance of continued refinement and adaptation to evolving spam patterns, and it lays the groundwork for future enhancements in SMS filtering techniques.

TABLE I
TABLE

Algorithm	Accuracy	Precision score
Nb	0.959381	1.00000
xgb	0.971954	0.954128
AdaBoost	0.962282	0.954128

VII. Conclusion And Future Scope

In conclusion, this research project on SMS spam filtering has demonstrated the effectiveness of employing machine learning algorithms for the identification and mitigation of unwanted messages. The results indicate a promising performance of the developed model, showcasing its ability to accurately classify SMS as spam or legitimate (ham). Through rigorous evaluation metrics, including precision, recall, and F1 score, the model's robustness in distinguishing between spam and ham messages has been underscored. The incorporation of feature engineering, encompassing aspects such as word frequency, message length, and the presence of specific keywords, has significantly contributed to the model's discriminative power. While the research has showcased commendable results, it is essential to acknowledge certain limitations, such as potential biases within the dataset or constraints inherent in the chosen algorithms. Future work may involve addressing these limitations, exploring additional features for refinement, and adapting the model to dynamically evolving patterns in SMS spam. Overall, the outcomes of this research provide valuable insights into the advancement of SMS spam filtering techniques, contributing to the broader landscape of cybersecurity and user experience enhancement in mobile communication.

In the realm of SMS spam filtering, several promising avenues offer intriguing possibilities for future research and development. One notable direction involves the exploration of advanced natural language processing (NLP) techniques to enhance the contextual understanding of messages, enabling more nuanced spam detection. Integrating sentiment analysis and semantic understanding could contribute to a more sophisticated filter that discerns the intent and tone of messages. Additionally, the integration of machine learning models with evolving technologies like deep learning and neural networks holds potential for improved feature extraction and pattern recognition, thereby increasing the filter's accuracy and adaptability. Exploring the feasibility of real-time or dynamic learning models could address the challenge of ever-changing spam tactics, allowing the filter to continually evolve and stay ahead of emerging spam patterns. Collaboration with telecommunication companies to access richer metadata and network-level information might also enhance the overall effectiveness of SMS spam filters. Finally, considering the increasing prevalence of multimedia messages, extending the scope of research to incorporate image and video analysis for spam detection represents an exciting frontier. As the SMS landscape continues to evolve, these potential avenues of exploration could significantly contribute to the advancement of SMS spam filtering technologies.

REFERENCES

[1] Sakshi Agarwal, Sanmeet Kaur, and Sunit Garhwal. "SMS spam detection for indian messages". In *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, pages 634–638, 2015.

[2] suhanta padosaini. "SMS spam detection using machine learning and deep learning techniques". In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 358–362, 2023.

[3] Pradeep kp. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[4] Dr.Sarika raga. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[5] Pradeep kp. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[6] Dr.M.A.Mukunthan. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[7] Mandar Shiivaji Hanchate. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[8] Qaswaa khaled Abood. "SMS spam classifci ation using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[9] dinesh kumar. Rahul Kumar. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[10] deepak kumar. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[11] Sharaff. Srinivasarao, U. "machine intelligence based hybrid classifier for spam detection and sentiment analysis of sms messages. ". In *Srinivasarao, U., Sharaff, A. Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages. Multimed Tools Appl 82, 31069–31099 (2023).*, volume 82, page 31069–31099, 2023.

[12] Kar N. (2023). SMS Spam Detection Using Deep Learning Approach. In: Bhattacharyya S. Banerjee J.S. Köppen M. (eds) Human-Centric Smart Computing. Debnath, K. "sms spam detection using deep learning approach. ". In *Srinivasarao, U., Sharaff, A. Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages. Multimed Tools Appl 82, 31069–31099 (2023).*, volume 316, page 337–347, 2022.

[13] Divyadarshan R. "SMS spam classification using machine learning and deep learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[14] Tarun Jain, Payal Garg, Namita Chalil, Aditya Sinha, Vivek Kumar Verma, and Rishi Gupta. "SMS spam classification using machine learning techniques". In *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 273–279, 2022.

[15] Mohammad AlShabi Marya Jehad, Ali Bou Nassif. "SMS spam classification using machine learning techniques". In *CONF Spam SMS detection based on long short-term memory and recurrent neural network*, pages 273–279, 2023.