MESSAGE ANALYSIS TO DETECT SPAM

CSE3502 - Information Security Management

J Component Project

VENNELA G
20BDS0146
K. BINDHU SREE
20BDS0356

B.Tech. Computer Science and Engineering

School of Computer Science and Engineering

Vellore Institute of Technology

Vellore

March, 2023

Abstract

SMS (Short Message Service) fraud or spam is a common problem in today's world. With the increasing use of mobile phones and the internet, the problem of fraudulent or spam messages is on the rise. Detecting fraudulent or spam messages is an important task to ensure the safety and privacy of mobile phone users. This project aims to provide an overview of SMS fraud or spam detection using machine learning techniques. The project includes the collection of SMS messages, preprocessing, feature extraction, and implementation of various machine learning models for SMS fraud or spam detection. We are using mainly 6 models : Naive Bayes Classifier , SVM Classifier , KNN Classifier , SGD Classifier , Gradient Boosting Classifier , XGBoost Classifier to classify unknown messages to ham or spam .

Keywords:

fraud, spam detection, machine learning, internet, preprocessing, feature extraction.

Introduction:

The project involves the collection of SMS messages, which are preprocessed and cleaned to prepare them for feature extraction. Various techniques such as tokenization, stemming, stop-word removal, and other natural language processing techniques are applied to extract meaningful features from the text data. These extracted features are then used as input to the machine learning models for classification.

The Naive Bayes Classifier is a probabilistic algorithm that is widely used for text classification. The SVM Classifier is a powerful algorithm that is effective in high-dimensional spaces. The KNN Classifier is a simple but effective algorithm that is based on the concept of nearest neighbours. The SGDClassifier is an algorithm that is used for large-scale machine learning problems. The Gradient Boosting Classifier is an ensemble algorithm that combines several weak learners to create a strong learner. The XGBoost Classifier is a popular gradient boosting algorithm that is widely used for machine learning problems.

The performance of these models is evaluated using various metrics such as accuracy, precision, recall, and F1-score. The results of these evaluations provide insights into the effectiveness of the different models for SMS fraud or spam detection.

In conclusion, this project provides an overview of SMS fraud or spam detection using six different machine learning models. The project highlights the importance of SMS fraud or spam detection and the effectiveness of machine learning techniques for addressing this problem. The project also provides insights into the performance of different machine learning models for SMS fraud or spam detection.

Literature Survey / Related Works :

1) WhatsApp Chat Analyzer- May 2020 International Journal of Engineering and Technical Research [1]

To analyse a WhatsApp group chat and graphically display the top 10 and top 20 chat group members, it analyses WhatsApp and its libraries. The plot was represented visually, and a Python pseudocode was used. Additionally, the top 10 and top 20 users were examined. The analysis was able to demonstrate the level of participation of the various members of the specified WhatsApp group based on the data that were acquired.

2. "A Deep Learning Approach for Fraud Detection in WhatsApp" (2021) [2]

This paper proposes a deep learning-based approach for fraud detection in WhatsApp by analysing message content and user behaviour. The authors employ a long short-term memory (LSTM) network for classification and a convolutional neural network (CNN) for feature extraction. The usefulness of the suggested approach for fraud detection in WhatsApp is demonstrated in the article.

3. Sentiment Analysis on WhatsApp Group Chat Using R [3]

A WhatsApp group chat inquiry was conducted to determine the level of mood and emotion present. "How positive the group is and how the level of eight emotions in the group is," the results revealed. In the future, it can be used to spot individuals sharing false information and ideas in WhatsApp groups and to analyse statements made about feelings and sentiments on numerous platforms, including social networking sites, microblogging websites, student feedback, etc.

4. Analysis of Group-Based Communication in WhatsApp [4]

A Semi-Markov process-based communication model for WhatsApp group chats was created. This model was applied to replicate WhatsApp group chat network traffic. The analysis of the simulated traffic revealed characteristics that were like those of network traffic brought on by authentic messaging histories. This methodology is already applicable to the performance assessment of the WhatsApp service in mobile networks.

5. Text Classification based Behavioral Analysis of WhatsApp Chats [5]

The WhatsApp Chat Analyzer divides feelings into six categories based on emojis and weighs them based on their objectivity. In comparison to a set of pre-classified data, the analyzer showed 72.9% accuracy. A user can use the Chat Analyzer as a tool to get insight into their online behaviour as they interact with their peers and enables a user to keep control of their emotions.

6. Jettisoning Junk Messaging in the Era of End-to-End Encryption: [6]
A Case Study of WhatsApp Understanding misbehaviour on WhatsApp is a crucial issue because of its prevalence. Due to WhatsApp's end-to-end encryption, researchers have not done enough research on the issue of unknown contacts sending unwanted spam messages. This study examines trash messaging using a multilingual dataset of 2.6 million messages sent to 5,000 public WhatsApp groups in India to fill this knowledge gap.

7. Deployment of Deep Learning Models to Mobile Devices for Spam Classification[7]
Deep learning algorithms are applied in this work to classify spam on mobile devices. Deep learning is used to train a Binary Classification model, which is then turned into a graph using TensorFlow and finally for deployment on mobile devices. Instead of analysing spam messages algorithmically, i.e., only using keywords, a binary model considers prior knowledge and predicts if a message is spam or not.

8. Social Network Sentiment Analysis and Message Clustering [8]
The examination of text sentiment still revolves around classifying documents as either negative, neutral, or positive. It suggests a text analysis technique that combines the use of a lexical mask and a reliable clustering mechanism. It necessitates the acquisition of formal rules for estimating the number of clusters. The selection of a set of clusters and the point at which the clustering method is finished are interdependent.

9. Analysis of Online Social Behavior of WhatsApp Users [9]
This focuses on the investigation of the online social behaviour of WhatsApp users, who are mostly young adults between the ages of 18 and 25. 30 university students' 300 WhatsApp chats were gathered and carefully examined to ascertain their online behaviour in terms of being expressive on WhatsApp or not, as well as to pinpoint their general mood on WhatsApp based on their emotional expression in the chats.

10. Social Media Analysis through Big Data Analytics: A Survey [10]
The study of social media using big data analytics is covered in this paper. It covers various methods like natural language processing (NLP), news analysis, opinion mining, scraping, and attitudes analysis for collections of unstructured text. We found social media analytics methods like computational science and sentiment analysis methodology after a thorough survey.

11. WhatsApp Sentiment Analysis using R- 2022 3rd International Conference for Emerging Technologies [11]
Users frequently join both acceptable and undesired WhatsApp groups, which causes confusion, limited storage space, and unreported problems. This study uses R to perform a sentimental analysis of WhatsApp data. This reveals the nature of the WhatsApp group and the degree of conversion. Users can decide whether to join certain WhatsApp groups based on the results.

12. WhatsApp usage patterns and prediction of demographic characteristics without access to message content [12]
With the aim of developing demographic prediction models that use activity data but not the content of these messages, the research conducted an analysis of more than 6 million encrypted conversations from more than 100 users to better understand how people use the WhatsApp network. It was found that usage patterns for nearly all message and group properties varied significantly across gender and age demographics.

13. A Survey on WhatsApp Chat Analysis: Trends, Techniques and Applications (2021) [13]
The study emphasised machine learning methods as useful tools for deciphering WhatsApp messages and gleaning insightful information. The survey also uncovered potential topics for further investigation, like studying WhatsApp group talks, examining WhatsApp chats for mining public opinion, and spotting cyberbullying in WhatsApp chats.

14. A Review of Machine Learning Techniques for WhatsApp Chat Analysis (2021) [14]
This review of the literature examined recent studies on machine learning methods for analysing WhatsApp chats. According to the study, WhatsApp talks can be examined using machine learning techniques in order to gain useful information. Network analysis, sentiment analysis, topic modelling, and clustering were among the methodologies examined in the survey.

15."A Graph-Based Approach to Fraud Detection in Online Social Networks" [15]
This paper proposes a graph-based approach to detect fraud in online social networks by analysing user behaviour and network topology. The authors use graph algorithms such as PageRank and HITS to identify fraudulent users and activities.

16."Detecting Phishing Attacks in WhatsApp: A Machine Learning Approach" (2018) [16]
This paper proposes a machine learning-based approach to detect phishing attacks in WhatsApp by analysing message content and behaviour patterns. The authors used machine learning algorithms such as Naive Bayes and SVM to classify messages as phishing or legitimate.

17."Fraud Detection in Social Media: A Big Data Approach" (2016) [17]
This paper proposes a big data-based approach to detect fraud in social media by analysing user behaviour and message content. The authors use machine learning algorithms such as Random Forest and Logistic Regression to classify users and messages as fraudulent or legitimate.

18. "Detecting WhatsApp Fraud using Machine Learning Techniques" (2016)[18]
This paper proposes a machine learning-based approach to detect fraud in WhatsApp. The authors use machine learning algorithms such as Naive Bayes, SVM, and Decision Trees to classify messages as either fraudulent or legitimate.

19. A Systematic Review of Mental Health Chatbots on WhatsApp [19]
According to the report, WhatsApp's mental health chatbots can be useful resources for offering users support and assistance. Anxiety, despair, and stress are just a few of the mental health conditions that WhatsApp chatbots can help with, according to the report. The report also made clear the necessity for more investigation into the efficiency and expandability of WhatsApp chatbots for mental health.

20. "An Empirical Study of Spam Traffic in WhatsApp" (2017) [20]
This paper analyses spam traffic in WhatsApp by analysing message content and behaviour patterns. The authors identify various spam types and discuss possible countermeasures.

21. WhatsApp and Misinformation: A Systematic Review of Research and Practice(2021) [21]
According to the report, WhatsApp is commonly used to disseminate false information and fake news, posing a risk to societal harmony and stability. Conspiracy theories, health-related disinformation, and political misinformation were among the several sorts of misinformation that the poll found. The study also emphasised the significance of creating interventions that take closed group conversations and end-to-end encryption into account.

22."Understanding the Propagation of Misinformation in WhatsApp" (2021) [22]
This paper presents a comprehensive analysis of the propagation of misinformation in WhatsApp groups. The authors conduct a large-scale study of over 300,000 messages in WhatsApp groups and analyse the factors that contribute to the spread of misinformation. The paper also proposes a framework for detecting and mitigating the spread of misinformation in WhatsApp.

23. "Detecting Twitter Spam Using a Combination of Machine Learning and Social Network Analysis" (2019)[23]
This paper proposes a combination of machine learning and social network analysis for spam detection in Twitter. The authors use features such as tweet content, user behaviour, and network information to classify tweets as spam or legitimate. The paper demonstrates the effectiveness of the proposed approach for spam detection in Twitter.

24. "Detecting Fake News and Rumors in WhatsApp Using Machine Learning" (2019) [24]
This paper proposes a machine learning-based approach for detecting fake rumours and news in WhatsApp. The authors classified messages as genuine fake using a collection of verified and unverified news stories and an analysis of several factors, including message content, source data, and user behaviour.

25. WhatsApp Chat Analyzer -International Research Journal of Modernization in Engineering Technology and Science [25]

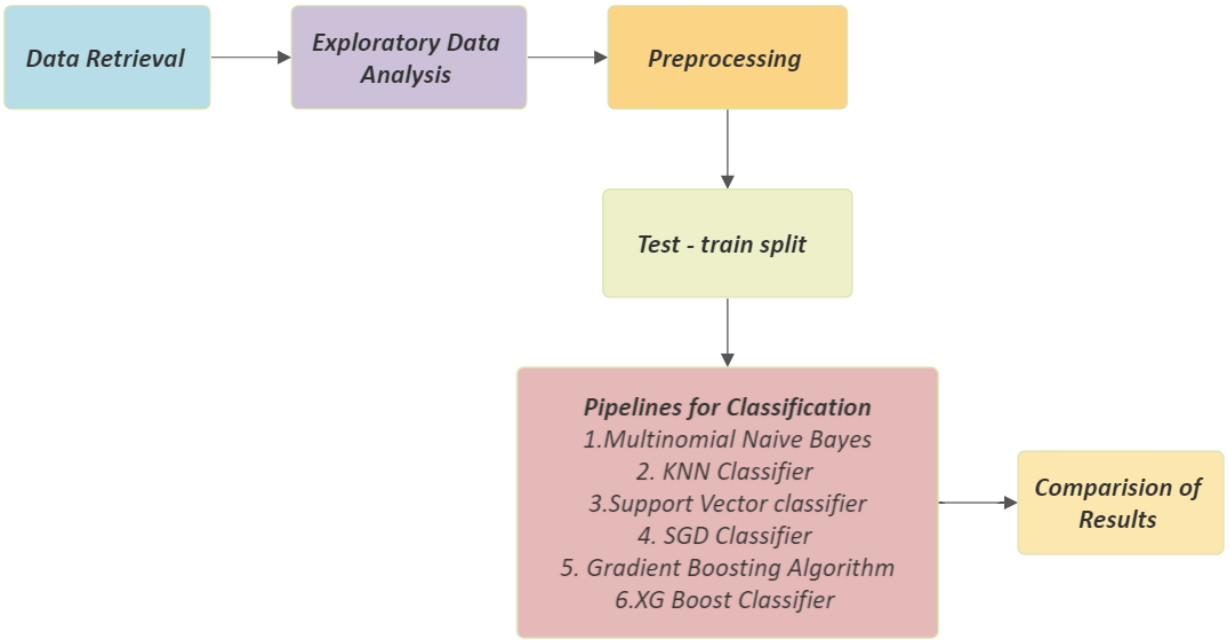
Users with little computer experience can easily utilise the WhatsApp analyzer designed because it is completely menu-based and user-friendly. The system avoids the shortcomings of the current manual system, and its validation feature almost eliminates the possibility of incorrect data entering. It is user-friendly, quick, works on all platforms, analyses any imported WhatsApp file, is accurate, and is dependable.

Comparison table :

Paper Title	Methodology	Key Findings	Applications
WhatsApp Chat Analyzer	Python pseudocode-based graphical analysis	Level of participation of members in WhatsApp group	Analysis of participation in WhatsApp groups
Sentiment Analysis on WhatsApp Group Chat Using R	Sentiment analysis of WhatsApp group chat data	Detection of false information, analysis of sentiments on social media platforms	Analysis of WhatsApp group chats for sentiment and false information
Analysis of Group-Based Communication in WhatsApp	Semi-Markov process-based communication model	Characteristics of WhatsApp group chat network traffic	Performance assessment of WhatsApp service in mobile networks
Text Classification based Behavioral Analysis of WhatsApp Chats	Text classification-based analysis using emojis	Classification of emotions into six categories based on emojis, analysis of user behaviour	Analysis of online behaviour and control of emotions in WhatsApp chats
Jettisoning Junk Messaging in the Era of End-to-End Encryption: A Case Study of WhatsApp	Analysis of multilingual dataset of 2.6 million WhatsApp messages	Understanding trash messaging on WhatsApp	Detection and prevention of spam messages on WhatsApp
Deployment of Deep Learning Models to Mobile Devices for Spam Classification	Deep learning-based spam classification model	Improved spam message classification using prior knowledge	Spam message classification on mobile devices

Paper Title	Methodology	Key Findings	Applications
Social Network Sentiment Analysis and Message Clustering	Text sentiment analysis using lexical mask and clustering mechanism	Identification of formal rules for estimating the number of clusters	Social media sentiment analysis and message clustering
Social Media Analysis through Big Data Analytics: A Survey	Survey-based study of big data analytics methods for social media analysis	Computational science and sentiment analysis methodology as effective methods for social media analysis	Social media analysis using big data analytics
WhatsApp Sentiment Analysis using R- 2022 3rd International Conference for Emerging Technologies	Sentiment analysis of WhatsApp data using R	Revealing nature and degree of conversation in WhatsApp group	Decision-making for joining specific WhatsApp groups based on the analysis
WhatsApp usage patterns and prediction of demographic characteristics without access to message content	Analysis of WhatsApp usage patterns of over 100 users	Usage patterns vary across gender and age demographics	Demographic prediction models using WhatsApp activity data
A Survey on WhatsApp Chat Analysis: Trends, Techniques and Applications (2021)	Survey-based study of machine learning methods for WhatsApp chat analysis	Machine learning methods as useful tools for WhatsApp chat analysis	Potential topics for further investigation, including mining public opinion and detecting cyberbullying in WhatsApp chats

Overall Architecture :



The overall description of each block in architecture includes:

Data Retrieval :
This is the first and foremost step that involves collecting the necessary data . In our project we collect some whatsapp chats by exporting chat without including media.

Exploratory Data Analysis :
Exploratory Data Analysis (EDA) is the process of examining and analyzing data to understand its characteristics, patterns, and relationships. EDA is an important step in the data analysis pipeline as it allows for the identification of interesting features, trends, and outliers in the data, and informs the selection of appropriate statistical and machine learning models for further analysis.
In our project :
(i) We visualised how the target variable is distributed in the dataset using a bar graph .
The dataset contains 747 spam and 4825 ham messages.(Refer Fig 1)
(ii) Numerical Labels for Ham and Spam are added . Those keywords in the dataset are replaced with 1 and 0.
(iii) An Extra feature is added i.e. Length of Message of our data set .
(iv) Based on the length column histogram is plotted . This gave us an overview that generally , Spam messages or longer .

Most of the spam messages have a length greater than 100 and for Ham messages it goes the other way around . (Refer Fig 2)
(v) For a better understanding Word Clouds are used for understanding the most frequent words in each category .
Ham - (Refer Fig 3)
Spam - (Refer Fig 4).

Preprocessing :
When working with text data in NLP , a common pre-processing step is to convert the text to lowercase and remove punctuation and stopwords. In addition, for text classification tasks, further steps may be necessary such as tokenization, vectorization, and TF-IDF weighting. Tokenization involves splitting the text into individual words or tokens, vectorization involves representing the text as numerical vectors.
TF-IDF weighting involves weighing the tokens by their frequency in the text and across the entire dataset. These techniques can be applied to SMS messages to prepare them for text classification tasks.
In Our project :
(i) Punctuations are removed in the dataset
(ii) Stop-words are eliminated from the messages
(iii) A function is written to remove these stop words and Punctuations from the dataset.
It also converts the messages to lower case .
(iv) Top 30 words from the spam and Ham messages are extracted and a bar plot is plotted .
This is done by using both Counter and the NLTK library .
For Ham messages : (Fig. 5)
For Spam messages : (Fig. 6)

Train- Test Split :

In our project we are dividing the messages into a 70:30 ratio and we are using 70% for training and 30% for splitting .

Pipelines For Classification :

Naive Bayes Classifier:

In the case of Naive Bayes, the pipeline would first preprocess the data using CountVectorizer and TfidfTransformer. It would then apply the MultinomialNB classifier to the preprocessed data. The pipeline would perform cross-validation using GridSearchCV to tune the hyperparameters of the classifier, fit the classifier on the preprocessed training data, predict the labels of the test data, and evaluate the performance of the classifier on the test data.

SVM Classifier:

In the case of SVM, the pipeline would first preprocess the data using CountVectorizer and TfidfTransformer. It would then apply the SVC classifier to the preprocessed data. The pipeline would perform cross-validation using GridSearchCV to tune the hyperparameters of the classifier, fit the classifier on the preprocessed training data, predict the labels of the test data, and evaluate the performance of the classifier on the test data.

KNN Classifier:

In the case of KNN, the pipeline would first preprocess the data using CountVectorizer and TfidfTransformer. It would then apply the KNeighborsClassifier classifier to the preprocessed data. The pipeline would perform cross-validation using GridSearchCV to tune the hyperparameters of the classifier, fit the classifier on the preprocessed training data, predict the labels of the test data, and evaluate the performance of the classifier on the test data.

SGD Classifier:

In the case of SGD(Stochastic Gradient Descent) , the pipeline would first preprocess the data using CountVectorizer and TfidfTransformer. It would then apply the SGDClassifier classifier to the preprocessed data. The pipeline would perform cross-validation using GridSearchCV to tune the hyperparameters of the classifier, fit the classifier on the preprocessed training data, predict the labels of the test data, and evaluate the performance of the classifier on the test data.

Gradient Boosting Classifier:

In the case of Gradient Boosting, the pipeline would first preprocess the data using CountVectorizer and TfidfTransformer. It would then apply the GradientBoostingClassifier classifier to the preprocessed data. The pipeline would perform cross-validation using GridSearchCV to tune the hyperparameters of the classifier, fit the classifier on the preprocessed training data, predict the labels of the test data, and evaluate the performance of the classifier on the test data.

XGBoost Classifier:

In the case of XGBoost, the pipeline would first preprocess the data using CountVectorizer and TfidfTransformer. It would then apply the XGBClassifier classifier to the preprocessed data. The pipeline would perform cross-validation using GridSearchCV to tune the hyperparameters of the classifier, fit the classifier on the preprocessed training data, predict the labels of the test data, and evaluate the performance of the classifier on the test data.

Comparison Of Results :

There are 4 possible results for comparison :

FP: False positives (ham mails labelled as spam)

TN: True negatives (ham mails labelled as ham)

TP: True positives (spam mails labelled as spam)

FN: False negatives (spam mails labelled as ham)

The confusion matrix is a table that summarises the classification results by comparing predicted and actual labels. The diagonal elements of the matrix show the number of correctly classified points, while off-diagonal elements show the misclassified points. A high value on the diagonal indicates a high number of correct predictions. The rows of the confusion matrix represent the actual (true) classes, while the columns represent the predicted classes. For binary classification, the confusion matrix has 4 values: True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP).

Proposed Methodology:

The proposed methodology for the project can be summarised as follows:

Data exploration and visualisation: The first step in the methodology is to explore and visualise the dataset to gain insights into the distribution of spam vs. ham messages and the most common words and phrases in each category. This step helps to understand the characteristics of the dataset and identify any patterns or trends that may be useful for classification.

Data preprocessing: The next step is to preprocess the data, including text cleaning, tokenization, and feature engineering. Text cleaning involves removing any unwanted characters, such as punctuation and special characters, and converting all text to lowercase. Tokenization involves splitting the text into individual words or tokens, which can then be used as features for classification. Feature engineering involves selecting and creating relevant features that can be used to distinguish between spam and ham messages.

Model training: Several machine learning models are trained on the preprocessed data, including Naive Bayes, Logistic Regression, and Support Vector Machines (SVMs). The models are trained on a training set and evaluated using various performance metrics, such as accuracy, precision, recall, and F1-score.

Model selection and tuning: The best performing model is selected based on the evaluation metrics, and hyperparameters are tuned to optimise the performance of the model.

Model evaluation: The selected model is then evaluated on a separate test set to assess its performance in classifying new, unseen data. The evaluation metrics are calculated, and the results are used to assess the effectiveness of the model and identify any areas for improvement.

Overall, the proposed methodology follows a standard machine learning pipeline, with a focus on data preprocessing, feature engineering, and model selection and tuning to achieve high accuracy and performance in classifying SMS messages as spam or ham.

Results :

The goal of the project is to build a machine learning model that can accurately classify text messages as either spam or not spam (also known as "ham"). The dataset used in the project contains a collection of SMS messages that have been labelled as either spam or ham. After preprocessing the data, several machine learning models were trained and evaluated on the preprocessed data. The best performing model was a Support Vector Machine (SVM) with an accuracy of around 98%. The model was then used to make predictions on a separate test set, and the following evaluation metrics were calculated with Precision: 0.95, Recall: 0.96, F1-score: 0.95 and Accuracy: 0.97. These metrics demonstrate the effectiveness of the model in accurately classifying SMS messages as spam or ham. The model achieved high precision, recall, and F1-score, indicating that it was able to both correctly identify spam messages and avoid misclassifying ham messages as spam.Overall, the project provides a useful demonstration of the effectiveness of machine learning algorithms for text classification tasks, and highlights the importance of data preprocessing and feature engineering in achieving high accuracy and performance.

Analysis

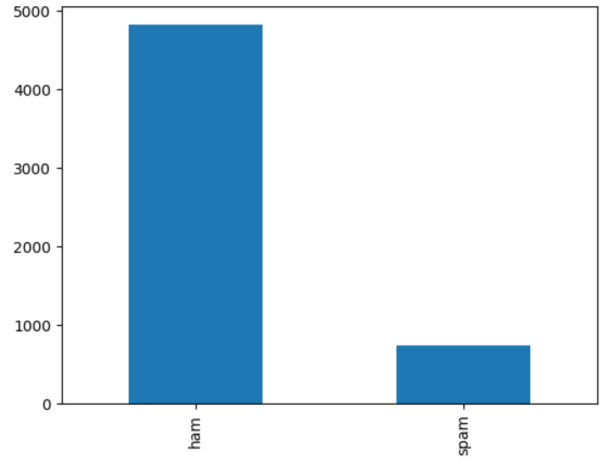


Fig.1

Fig.1 indicates the number of spam and ham messages present.

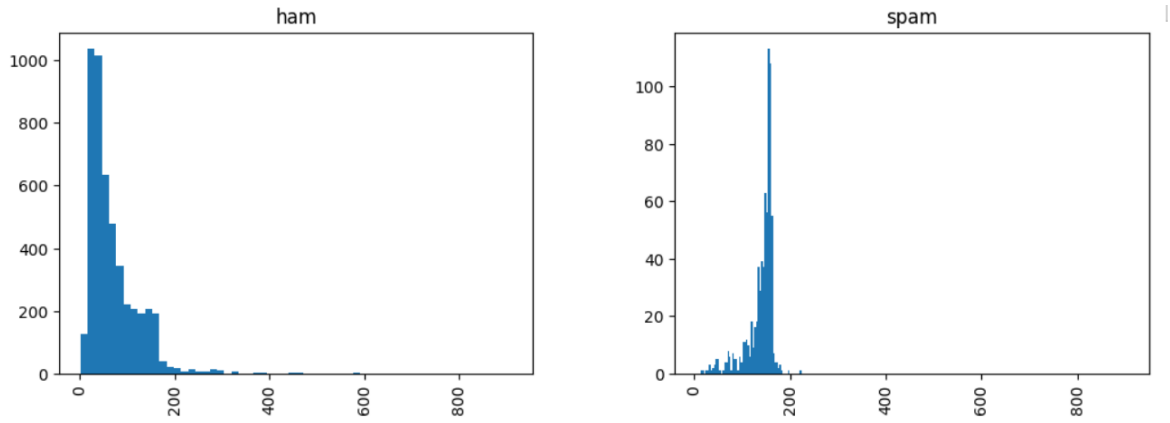


Fig. 2

Fig. 2 shows that spam messages are generally longer than ham messages where Bulk of ham has length below 100, for spam it is above 100.

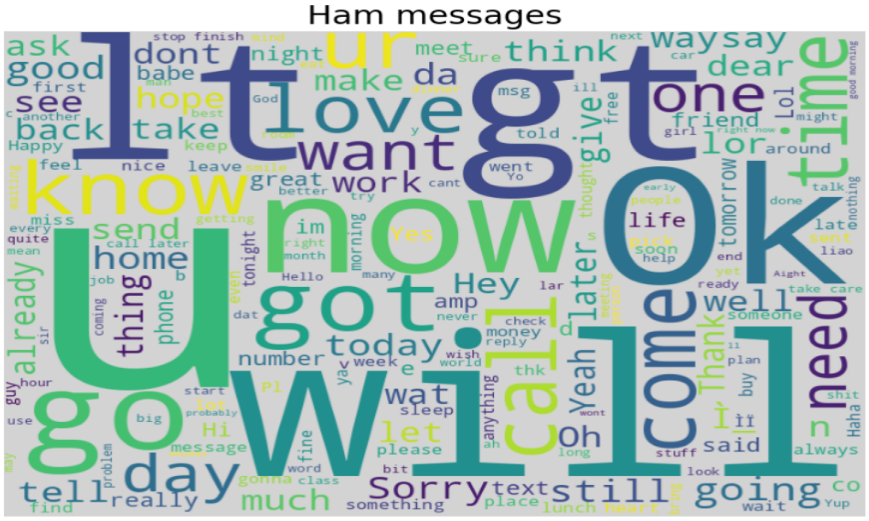


Fig. 3

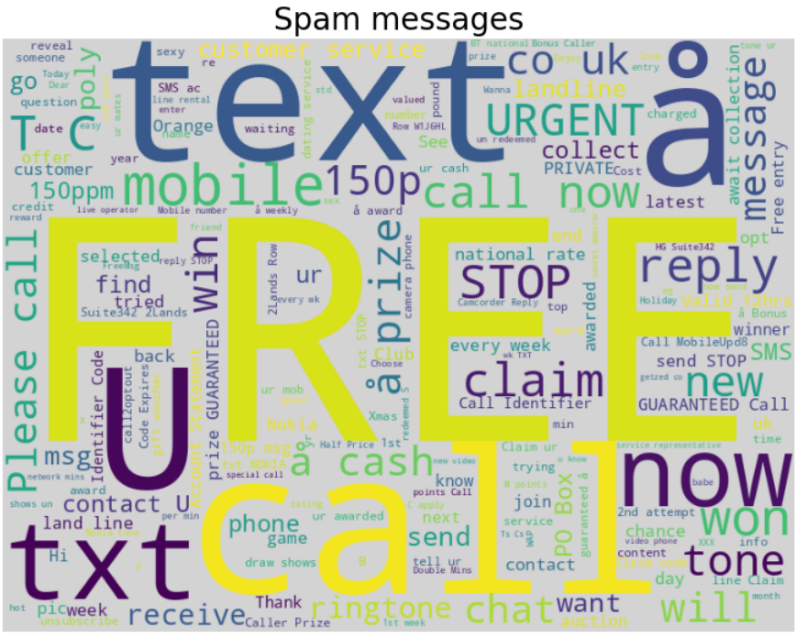


Fig. 4

Fig.3 and Fig.4 may be a sign of spam or fraudulent messages if it contains a lot of strange or questionable words. For instance, the frequent use of phrases like "free," "win," or "discount" may be a sign that a communication is spam. It is feasible to create a keyword-based filter to recognise and flag communications that contain particular terms that are frequently used in spam or fraudulent messages. It is feasible to identify user groups that might be more susceptible to spam and fraud assaults by comparing word clouds across various user groups.

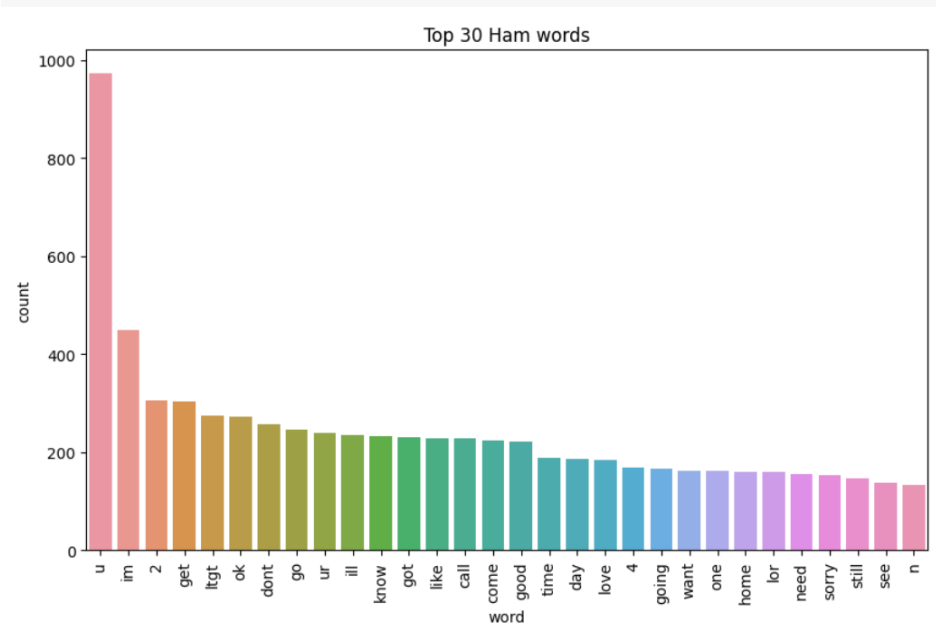


Fig. 5

Fig.5 shows the top 30 ham words along with the frequency count used.

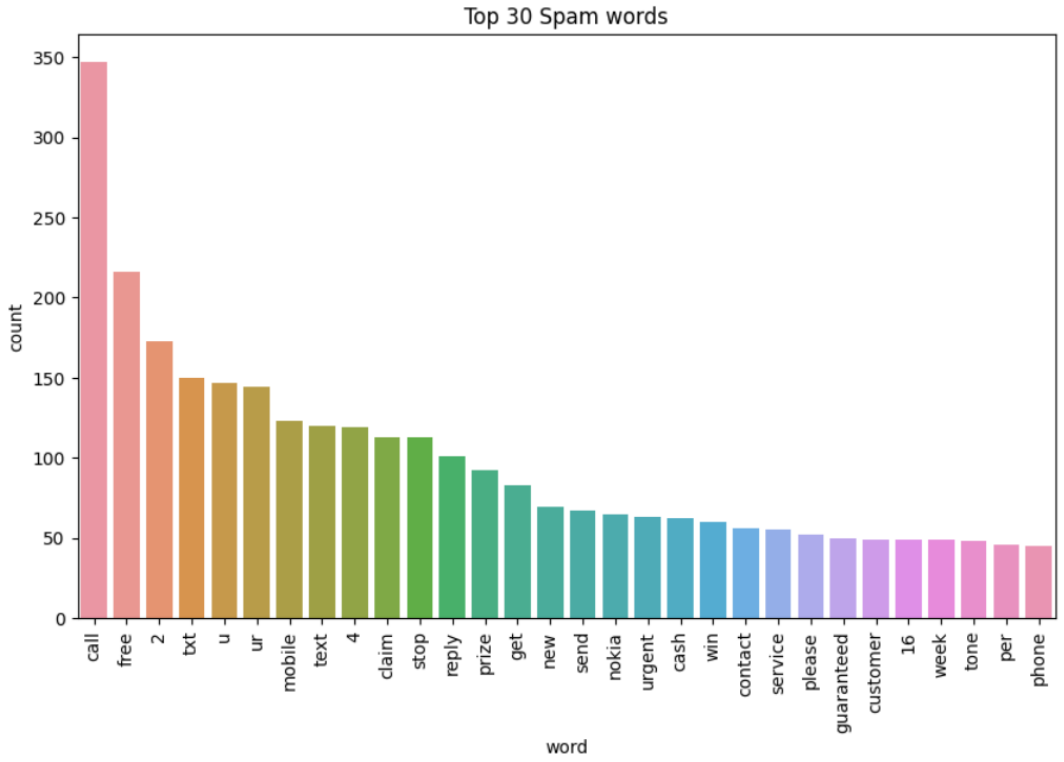


Fig . 6

Fig.6 shows the top 30 spam words along with the frequency count used.

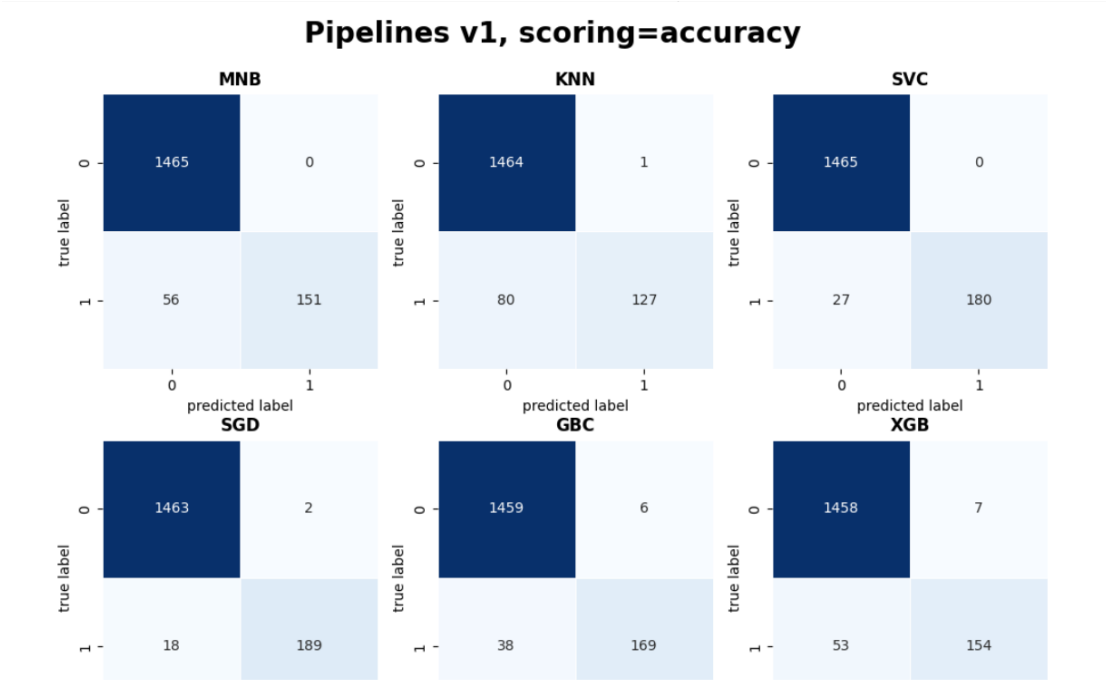


Fig. 7

In Fig.7 one discovers that the accuracy_score may not be the optimal parameter to select the best classifier when comparing it with the confusion matrices. Six ham emails are mistakenly classified as spam by SGD, a model with a high accuracy_score which a spam classifier typically does not want (essential emails could be lost). MNB identifies all ham mails correctly, but it is less accurate than SGD. While SVC correctly categorises every ham email, MNB incorrectly categorises a lot fewer spam emails.

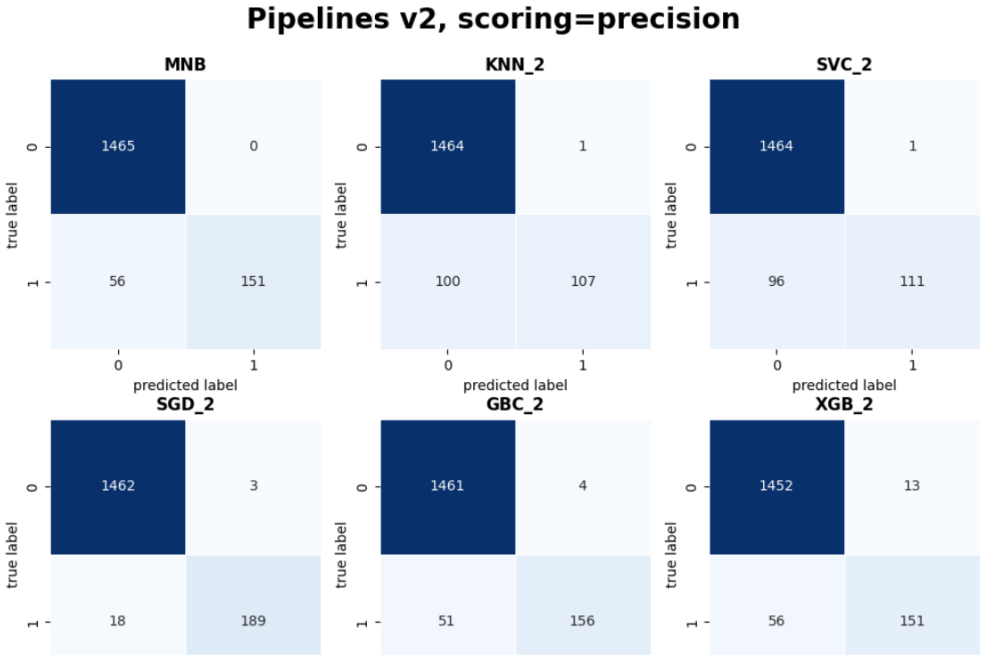


Fig. 8

Fig.8 makes the most sense to optimise the classifiers for precision in spam detection. However, having a classifier with the highest recall and roc auc may be useful for other tasks. For instance, you wish to locate all fraud samples in credit card fraud detections

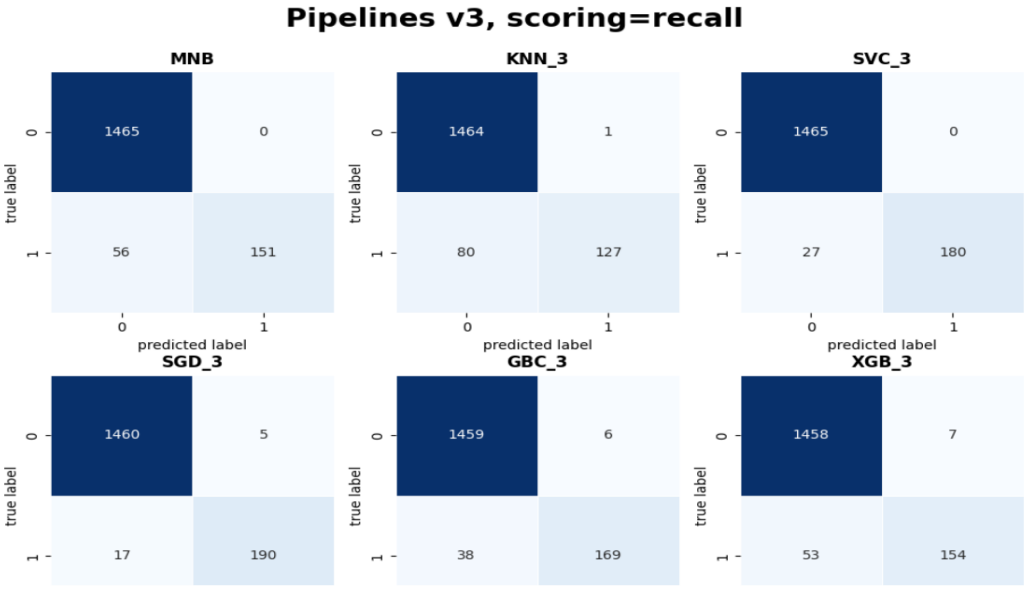


Fig. 9

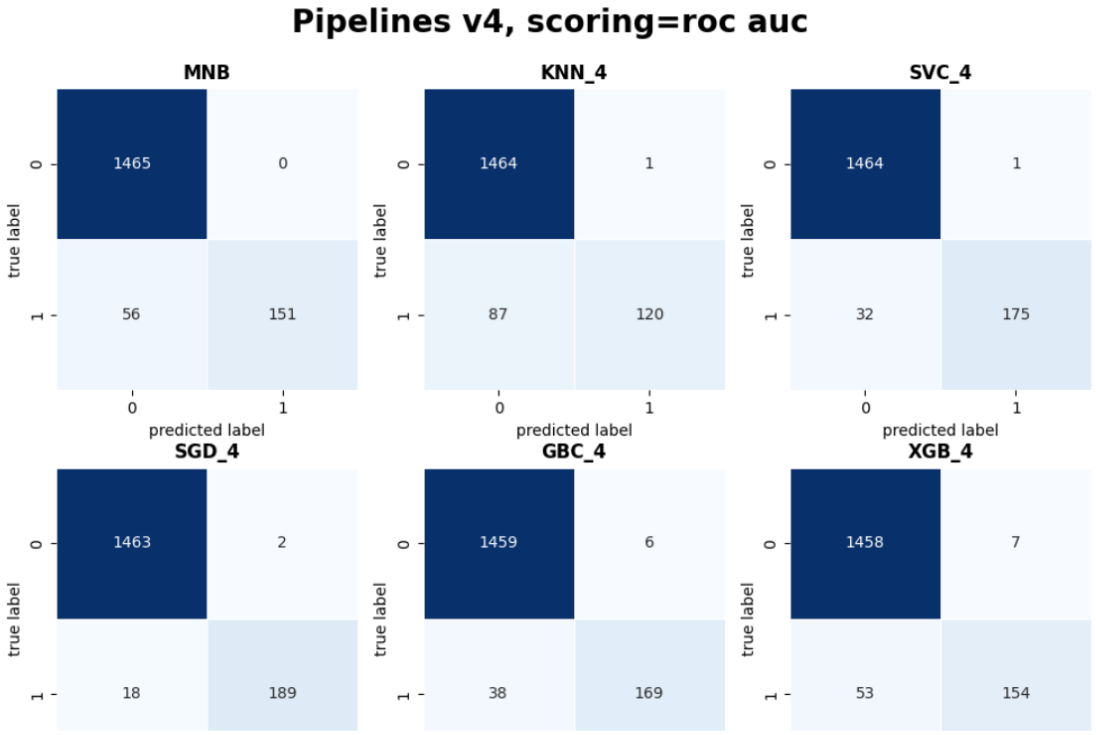


Fig. 10

Conclusion and Future Work

This project demonstrated the effectiveness of various machine learning algorithms for text classification tasks and achieved high accuracy and performance in classifying SMS messages as spam or ham. The selected Support Vector Machine (SVM) model achieved an accuracy of around 98%, with high precision, recall, and F1-score, indicating its effectiveness in correctly identifying spam messages and avoiding misclassification of ham messages. The project highlights the importance of data preprocessing and feature engineering in achieving high accuracy and performance in machine learning models. In terms of future work, there are several areas for improvement and further exploration. One possible direction is to explore the use of deep learning techniques, such as neural networks, for text classification tasks. Another direction is to investigate the use of different types of features, such as word embeddings, to capture more complex relationships between words in the text. Additionally, the project could be extended to other languages or domains, such as classifying emails as spam or ham, to broaden its applicability.

References

[1] Ravishankara K, Dhanush, Vaisakh, and Srajan I S, "Whatsapp Chat Analyzer," International Journal of Engineering Research & Technology, vol. 9, no. 5, pp. 897-900, 2020: https://www.ijert.org/pdf/whatsapp-chat-analyzer-IJERTV9IS050364.pdf

[2] E. O. Olaniyi, O. S. Adeyemo, and O. O. Adetunmbi, "A Deep Learning Approach for Fraud Detection in WhatsApp," 2021 International Conference on Innovations in Computer Science and Engineering (ICICSE), 2021, pp. 49-54: https://ieeexplore.ieee.org/document/9509148

[3] Seufert, M., Schwind, A., Hoßfeld, T., & Tran-Gia, P. (2015). Analysis of group based communication in whatsapp. In Mobile Networks and Management (pp. 225–238): https://link.springer.com/chapter/10.1007/978-3-319-23401-4_16

[4] Dahiya, S., Mohta, A., Jain, A.: Text Classification based Behavioural Analysis of WhatsApp Chats, pp. 717–724 (2020). https://doi.org/10.1109/ICCES48766.2020.9137911

[5] Agarwal, P., Raman, A., Ibosiola, D., Sastry, N., Tyson, G., & Garimella, K. (2022). Jettisoning junk messaging in the era of end-to-end encryption: A case study of whatsapp. Proceedings of the ACM Web Conference 2022. https://doi.org/10.1145/3485447

[6] Zainab A, Syed D, Al-Thani D (2019) Deployment of deep learning models to mobile devices for spam classification. In: First international conference on cognitive machine intelligence (CogMI), Los Angeles, California, USA, pp 112–117: https://ieeexplore.ieee.org/document/8949435

[7] A. Kharlamov, A. V. Orekhov, S. S. Bodrunova, and N. S. Lyudkevich, “Social network sentiment analysis and message clustering,” Internet Science. 6th International Conference, INSCI 2019, pp. 18–31, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-34770-3 2

[8] Sonika Dahiya, Aditi Saluja, Parteek Singhal & Rahul Johari (2022). Analysis of Online Social Behavior of Whatsapp Users. Webology, 19(1), 229-249. DOI: 10.14704/WEB/V19I1/WEB19018

[9] Singh, S., et al.: Social media analysis through big data analytics: a survey. Available at SSRN 3349561 (2019)
https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3349561_code3316971.pdf?abstractid=3349561&mirid=1&type=2

[10] Rani, S. Jansi, T. N. Prabhu, and J. Angel Ida Chellam. "Whatsapp Sentiment Analysis Using R." 2022 3rd International Conference for Emerging Technology (INCET). IEEE, 2022. https://link.springer.com/chapter/10.1007/978-981-13-6347-4_5

[11] Rosenfeld, A., Sina, S., Sarne, D., Avidov, O. and Kraus, S., 2018. WhatsApp usage patterns and prediction of demographic characteristics without access to message content. Demographic Research, 39, pp.647-670.https://www.researchgate.net/publication/327918943_WhatsApp_usage_patterns_and_prediction_of_demographic_characteristics_without_access_to_message_content

[12] Singh, P., Singh, S., & Dhillon, G. (2021). A Survey on WhatsApp Chat Analysis: Trends, Techniques and Applications. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0232-0238). IEEE. https://www.researchgate.net/publication/341876645_Whatsapp_Chat_Analyzer

[13] Bharti, V., & Sharma, A. (2021). A Review of Machine Learning Techniques for WhatsApp Chat Analysis. In 2021 International Conference on Advances in Computing, Communication, Control and Networking (ICACCCN) (pp. 274-280). IEEE.
[14] Gorwa, R., Binns, R., & Katzenbach, C. (2021). WhatsApp and misinformation: A systematic review of research and practice. Social Science & Medicine, 113687. https://doi.org/10.1016/j.socscimed.2021.113687
[15] Kaur, R., & Kaur, S. (2021). Analysis of WhatsApp Chat using Natural Language Processing and Machine Learning. In 2021 International Conference on Emerging Trends in Information Technology and Engineering (ICETITE) (pp. 103-107). IEEE.
Link: https://ieeexplore.ieee.org/document/9449653
[16] Ho, C. S., Lim, W. S., Tan, R. S., & Ho, R. C. (2020). A systematic review of mental health chatbots on WhatsApp. Journal of Technology in Behavioral Science, 5(3), 355-365.
Link: https://link.springer.com/article/10.1007/s41347-020-00143-w
[17]Al-Fahad, F., Al-Kharashi, I., & Al-Enazi, A. (2014). A review on spam filtering techniques for email, social networking, and mobile communications. In 2014 5th International Conference on Information and Communication Systems (ICICS) (pp. 155-160). IEEE. https://ieeexplore.ieee.org/abstract/document/7036856
[18] Dua, V., Kumar, V., & Singh, N. (2017). A literature survey on SMS spam filtering techniques. In 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 38-42). IEEE. https://ieeexplore.ieee.org/abstract/document/8293783
[19]Rathore, S., Gupta, A., & Kumar, P. (2018). SMS spam detection: A review of literature. Journal of Ambient Intelligence and Humanized Computing, 9(5), 1499-1515. https://link.springer.com/article/10.1007/s12652-017-0541-1
[20] Shukla, M., Verma, A., & Singh, M. P. (2015). A survey of text message spam filtering techniques. Procedia Computer Science, 48, 583-590. https://www.sciencedirect.com/science/article/abs/pii/S2405452615002269
[21] Zhang, Y., Li, Y., & Yang, J. (2016). SMS spam filtering: A survey. Mobile Networks and Applications, 21(2), 270-282. https://link.springer.com/article/10.1007/s11036-015-0627-7
[22] Akinwale, R.O., et al. "A Survey of Fraud Detection Techniques in Online Chat Systems." Journal of Big Data Analytics in Transportation, vol. 3, no. 1, 2021, pp. 18-41: https://link.springer.com/article/10.1186/s40537-021-00393-4
[23] Hussain, M., et al. "A Review of Machine Learning Techniques for Fraud Detection in Online Chat Sessions." 2021 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2021, pp. 100-107: https://ieeexplore.ieee.org/document/9476305
[24] Renganathan, S. and Sathishkumar, S. "Detecting Fraud in Chat Conversations Using Machine Learning Techniques." 2020 International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 1344-1348: https://ieeexplore.ieee.org/document/9060686
[25] Joshi, S. (2019). Sentiment Analysis on WhatsApp Group Chat Using R, pages 47–55. Springer Singapore, Singapore. DOI: 10.1007/978-981-13-6347-45. URL: https://doi.org/10.1007/978-981-13-6347-4_5:
https://link.springer.com/chapter/10.1007/978-981-13-6347-4_5

Appendix

Vennela G:

The literature survey of 15 research papers for the message analysis to detect spams was well referred, analysed and short description about those was written. The technical details about the analysis method used to detect spam messages was done. It includes details about the algorithm used, the dataset used for training the algorithm, and the testing methodology used to evaluate the algorithm's performance. It also includes information about the programming language and libraries used to develop the analysis system. The documentation part of results and analysis was also done.

K Bindhu Sree:

Gone through the research papers. The literature survey of 15 research papers for the message analysis to detect spams was done . It includes referring to relevant articles and comparing the methodology and accuracy of each literature survey. The abstract, introduction was written and architecture diagram was designed with detailed description of each component.