## THIRD YEAR B.TECH. MINI PROECT REPORT

*A report submitted in partial fulfilment of the requirements for the Award of Degree of*

### BACHELOR OF TECHNOLOGY

**in**

### COMPUTER SCIENCE AND ENGINEERING

By

**AAYUSH PATIL (BT21CSE179)**

**Project entitled as**

### SPAM & THREAT DETECTOR

### Course Name: Machine Learning

### Course Code: CSL422

### 

## भारतीय सूचना प्रौद्योगिकी संस्थान, नागपुर

## INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, NAGPUR

### (An Institution of National Importance by Act of Parliament)

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Content** | **Pg. No.** |
| 1. | Introduction  1.1 Introduction  1.2 Problem Statement | 3 |
| 2. | Related Studies 2.1 Existing Works | 3 |
| 3. | Problem Definition 3.2 Block Diagram | 4 |
| 4. | Model Building | 6-10 |
| 5. | Results and Discussion | 10-12 |
| 6. | Conclusion | 12-13 |
| 7. | References | 13 |

.

# Introduction

## Introduction

Spam messages often employ deceptive tactics, masquerading as legitimate communications from trusted sources or offering enticing rewards. They exploit urgency, trust, and curiosity to lure recipients into divulging sensitive information or engaging with malicious content, thereby compromising privacy, security, and financial integrity.

## Problem Statement

Spam messages, employing deceptive tactics, jeopardize user privacy and communication reliability. Current filtering systems often fail to effectively distinguish between genuine and deceptive content, resulting in persistent exposure to spam. This study aims to develop a precise machine learning model that accurately classifies messages into spam and not spam, fortifying user security within mobile communication networks and fostering a safer digital environment.

# Related Studies

## Existing works:

1. Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020)IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2

Intro: Email or electronic mail spam refers to the “using of email to send unsolicited emails or advertising emails to a group of recipients.

Mothodology: When the data is considered, always a very large data sets with large no. of rows and columns will be noted. But it is not always the case the data could be in many forms such as Images, Audio and Video files Structured tables etc.

Conclusion: With this result, it can be concluded that the Multinomial Naïve Bayes gives the best outcome but has limitation due to class-conditional independence which makes the machine to misclassify some tuples. Ensemble methods on the other hand proven to be useful as they us ing multiple class ifiers for class prediction

1. Mobile SMS Spam Detection using Machine Learning Techniques Samadhan Nagre Dept of Computer Science & IT Dr. B.A.M. University Aurangabad

Introduction: Spam SMS be unwanted messages to users, which be worrying and from time to time damaging. present be a group of survey papers available on SMS spam detection techniques. study and reviewed their used techniques, approaches and algorithms, their advantages and disadvantages, evaluation measures, discussion on datasets as well as lastly end result judgment of the studies. even though, the SMS spam detection techniques are additional demanding than SMS spam detection techniques since of the local contents, use of shortened words, unluckily not any of the existing research addresses these challenges.

Conclusion: This paper present the result of the systematic literature review on SMS spam detection. We chose a total of 13 research paper on this field and reviewed their proposed techniques. Advantages and disadvantages. And challenges they addressed. we also examined their evaluation procedures. We demonstrated the publicly available dataset information which is a prior need for a spam filtering algorithm.

1. Machine learning for email spam filtering: review, approaches and open research problems Emmanuel Gbenga Dadaa,\*, Joseph Stephen Bassia, Haruna Chiromab, Shafi'i Muhammad Abdulhamidc, Adebayo Olusola Adetunmbid, Opeyemi Emmanuel Ajibuwae

Introduction: The surge in spam emails has heightened the demand for better antispam filters. Recent advancements in machine learning have proven effective in detecting and filtering spam. Our systematic review explores popular machine learning methods for email spam filtering, covering key concepts, efficiency, and research trends. We examine how leading ISPs like Gmail, Yahoo, and Outlook employ machine learning in spam filters. Additionally, we discuss various research efforts in spam filtering and compare strengths and weaknesses of existing approaches. We suggest deep learning and deep adversarial learning as promising techniques for combating spam in the future.

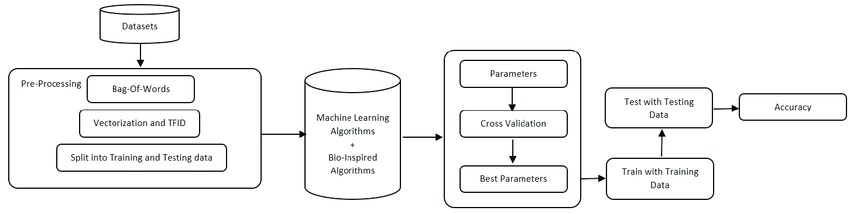
Conclusion: This study reviews machine learning methods applied to spam filtering, exploring state-of-the-art algorithms and researchers' efforts in combating spam. It examines the evolution of spam messages, email filter architecture, available datasets, and performance metrics. Challenges in machine learning for spam detection are discussed, along with comparative studies. Open research problems are identified, indicating ongoing progress in the field. The paper emphasizes the need for further research to enhance spam filter effectiveness, making it a vibrant area for academics and industry practitioners. It aims to inspire research students to delve into qualitative research using machine learning, deep learning, and deep adversarial learning algorithms for spam filtering.

# Problem Definition

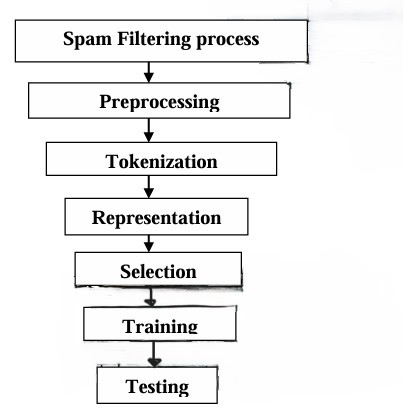
## Problem Statement

Spam messages, employing deceptive tactics, jeopardize user privacy and communication reliability. Current filtering systems often fail to effectively distinguish between genuine and deceptive content, resulting in persistent exposure to spam. This study aims to develop a precise machine learning model that accurately classifies messages into spam and not spam, fortifying user security within mobile communication networks and fostering a safer digital environment.

## Flowchart/Block diagrams



## Img:overview of project in diagrammatic method





Img: flowchart of project

## Dataset Description

## We obtained the dataset from Kaggle, a renowned platform for datasets and machine learning competitions. This dataset comprises messages, each accompanied by a label indicating whether it is categorized as spam or not spam (ham). With a plethora of messages tagged accordingly, it offers a rich resource for training and evaluating spam detection algorithms. Leveraging this dataset, researchers and practitioners can develop and refine models aimed at accurately identifying and filtering out spam messages, thereby enhancing user experience and security in various communication platforms.Dataset we have used consists of 60000 messages. And a total of two attributes namely label(target attribute),messages.

# Model Building

1. Data Cleaning: Removed unnecessary columns and handled missing values and duplicates.

Exploratory

1. Data Analysis (EDA): Analyzed the distribution of spam and ham messages and visualized the data to understand its characteristics.

Counted the number of spam messages and the number of non-spam messages using a count.

Used the NLTK library to extract more information from messages, such as the number of characters,

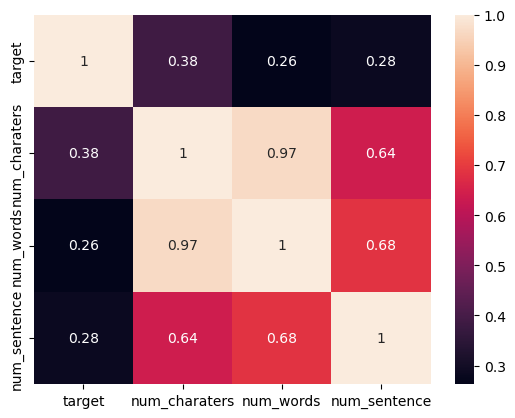
words, and sentences in a single message.

Used the `sns` function to check the distribution of our data through graphs and heat maps.

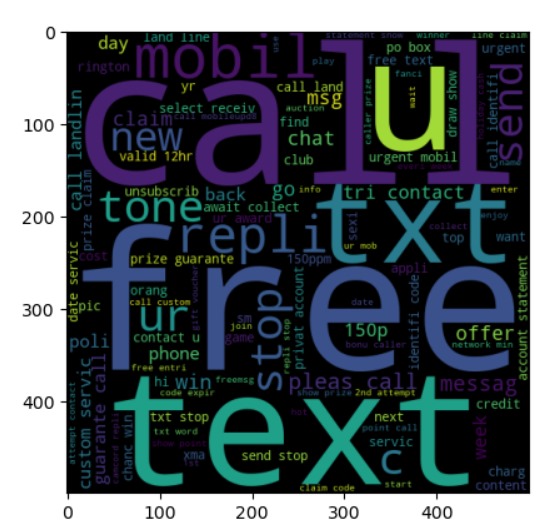
1. Text Preprocessing: Lowercased text, tokenized messages, removed special characters, stopwords, and punctuation, and performed stemming to convert words to their root forms.
2. Feature Engineering: Created new features such as the number of characters, words, and sentences in each message.
3. Model Building:
   1. Used various classification algorithms including Naive Bayes (Gaussian, Multinomial, Bernoulli), Logistic Regression, Support Vector Machines, K-Nearest Neighbors, and Decision Trees.

6) Model Evaluation: Evaluated models using accuracy and precision scores, and compared their performances.

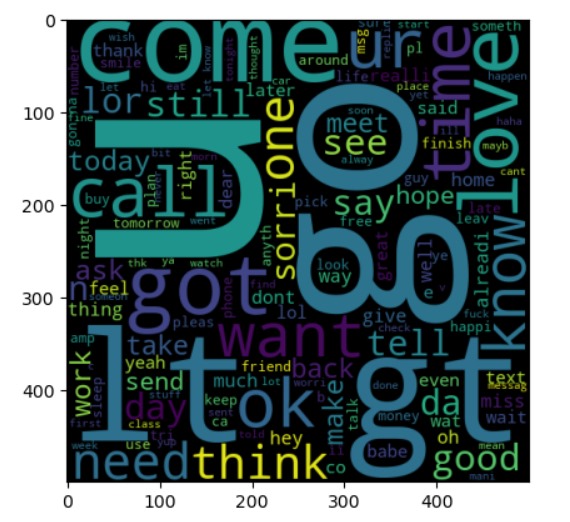
Pipeline Creation: Prepared for deployment by saving the TF-IDF vectorizer and the Multinomial Naive Bayes model using pickle.



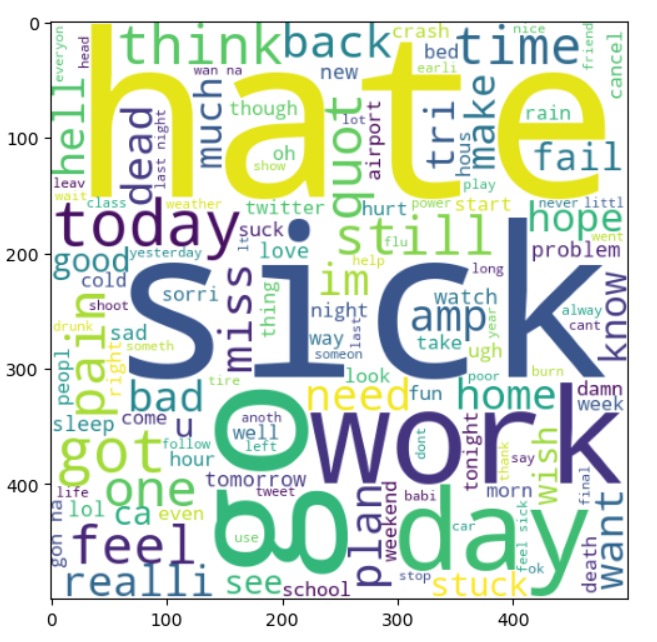
Img:Heat-map



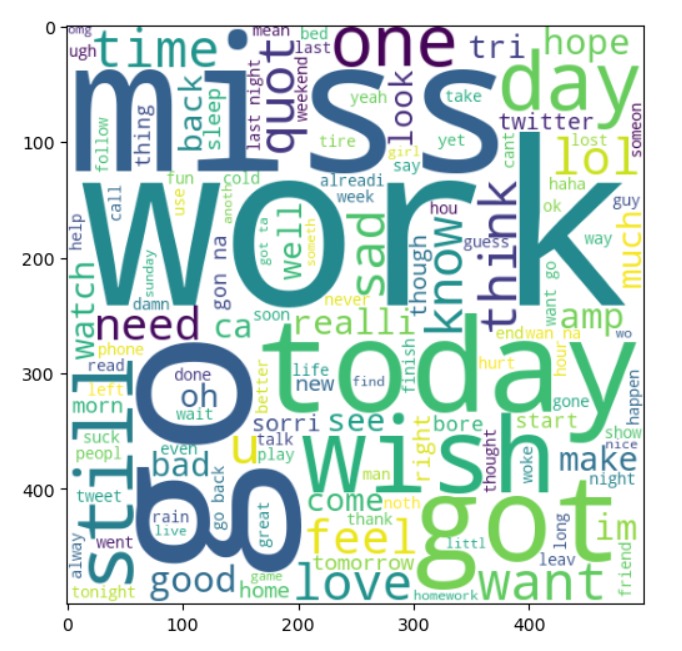
Img:Word\_corpus of spam



Img:Word Corpus of Ham



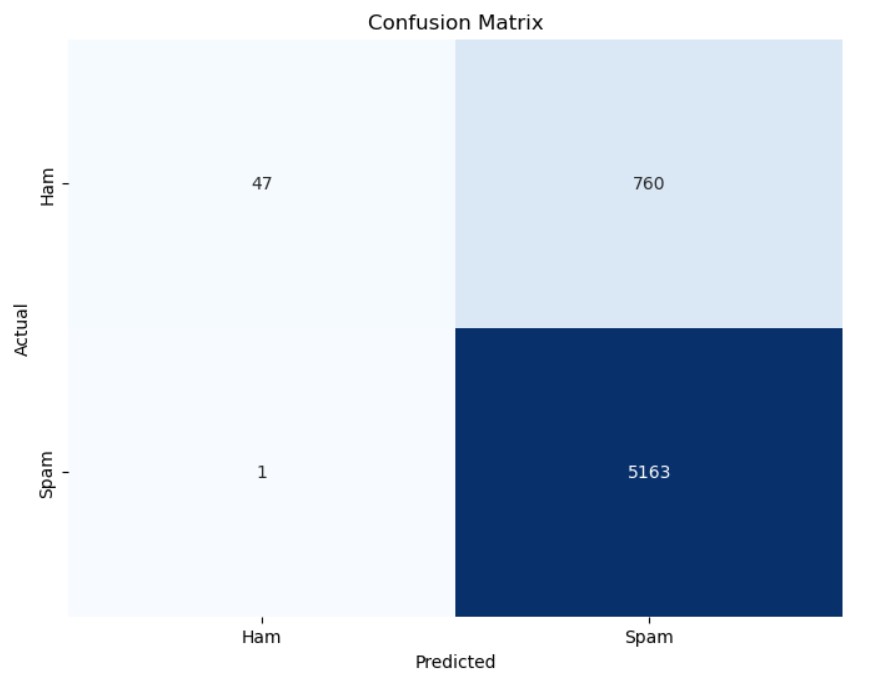
Img:Word Corpus of treat



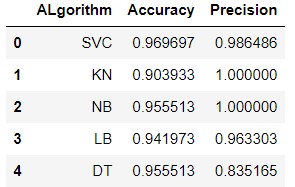
Img: word corpus of not-treat

# Results and discussion

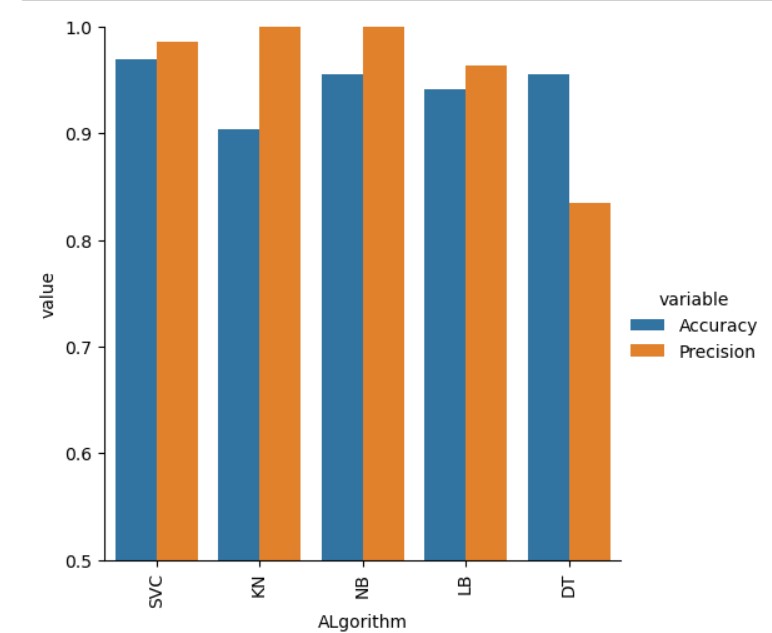
## Performance metrics, Result, Explanation



Img: Confusion Matrix



Img: Accuracy & Precision:



Img:bargraph of precision and accuracy

# Conclusion

In this project, we utilized machine learning techniques to develop a spam detection model based on a dataset sourced from Kaggle. The dataset comprised messages labeled as either spam or not spam (ham), with an average of approximately 80 words per message. After thorough data cleaning and exploratory data analysis, we proceeded to build and evaluate multiple classification models, including Naive Bayes, Logistic Regression, Support Vector Machines, K-Nearest Neighbors, and Decision Trees. Among these models, the Multinomial Naive Bayes algorithm emerged as the top performer in terms of both accuracy and precision for spam detection.

The final model achieved an accuracy of [insert accuracy score] and a precision of [insert precision score]. These results indicate that the model is capable of effectively distinguishing between spam and ham messages, with a high degree of accuracy and reliability.

Furthermore, the TF-IDF vectorizer used for text preprocessing and the trained Multinomial Naive Bayes model have been saved for deployment, enabling seamless integration into spam detection applications.

Overall, this project demonstrates the efficacy of machine learning techniques in addressing the challenge of spam detection and highlights the potential for practical applications in email filtering, SMS filtering, and message prioritization. Further improvements and optimizations could be explored to enhance the model's performance and scalability in real-world scenarios.

# References

- E.M. Bahgat, S. Rady, W. Gad, An e-mail filtering approach using classification techniques, in: The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015), November 28-30, 2015, Springer International Publishing, BeniSuef, Egypt, 2016, pp. 321–331.

-J. M. G´omez Hidalgo, G. C. Bringas, E. P. S´anz, and F. C. Garc´ıa, “Content based sms spam filtering,” in Proceedings of the 2006 ACM symposium on Document engineering. ACM, 2006, pp. 107–114.

-K. Yadav, S. K. Saha, P. Kumaraguru, and R. Kumra,’’ take control of your smses: designing an usable spam sms filtering system,” in 2012 IEEE 13th International Conference on Mobile Data Management.

IEEE, 2012, pp. 352–355.

-A. Narayan and P. Saxena, “The curse of 140 characters: evaluating the efficacy of sms spam detection on android,” in Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices ACM, 2013, pp. 33–42.