SPAM EMAIL DETECTION

## PROJECT REPORT

***Submitted by***

Sargun Singh (2110993830)

Sohil Dhiman (2110993839)

Nikhal Kumar Rana(2110993861)



**BE-CSE (Artificial Intelligence)**

***Guided by***

## Dr. Kamal Deep Garg Dr. Vandhna Sood

### CHITKARA UNIVERSITY INSITUTE OF ENGINEERING & TECHNOLOGY CHITKARA UNIVERSITY, RAJPURA

**June and 2022**

# CONTENTS

* Acknowledgement
* Abstract
* List of Tables
* List of Figures
* Chapters
  + Introduction
  + Problem Formulation
  + Proposed Solution / Methodology
  + Flowchart/ system flow diagram
  + Software & Hardware requirement
  + Results
  + Conclusion
* Snapshots
* References

# ACKNOWLEDGEMENTS

With immense please we, **Mr Sargun Singh, Mr Sohil Dhiman, Mr Nikhal Kumar Rana**

presenting **“Study Jammers** .” project report as part of the curriculum of ‘BE-CSE (AI)’.

I would like to express my sincere thanks to **Dr. Vandna Sood** (mentor name), for his/her valuable guidance and support in completing my project.

I would also like to express my gratitude towards our dean **Dr. Sushil Narang\_** (dean name)

for giving me this great opportunity to do a project on

**Spam SMS Classifier**

(topic name).

Without their support and suggestions, this project would not have been completed.

Signature………...

Name:

Roll No:

# ABSTRACT

The increasing use of mobile phones and messaging services has led to an increase in the number of spam SMS messages. To tackle this problem, a spam SMS classifier can be developed using machine learning algorithms. This project aims to build a spam SMS classifier that can accurately identify spam messages and prevent them from reaching the recipient's inbox. The methodology involves collecting a dataset of SMS messages, preprocessing the data, extracting features, building a model using the Naive Bayes algorithm, and evaluating the performance of the model using various metrics. The classifier achieved an accuracy of 98%, Precision of 99%. The results demonstrate the effectiveness of the spam SMS classifier in identifying spam messages and improving the user experience with messaging services. The spam SMS classifier can be further improved by incorporating other machine learning algorithms and techniques.

LIST OF FIGURES

1. **Figure 1** – Flowchart (Working of our user interface is represented using)
2. **Figure 2** – Home Page (Shows the working of Home Page)
3. **Figure 3** – Contact Page (Shows the working of Contact Page)

# CHAPTER – 1

**INTRODUCTION**

With the rise in the use of mobile phones and messaging services, the number of spam SMS messages has increased dramatically in recent years. These messages can be harmful and contain fraudulent or unwanted content. Such messages can be annoying, distracting, and time-consuming for the recipients, especially when they are bombarded with spam messages continuously. To combat this problem, a spam SMS classifier can be developed using machine learning algorithms. The classifier can accurately identify spam messages and prevent them from reaching the recipient's inbox.

The concept of spam SMS classification involves the use of machine learning algorithms to develop a model that can predict whether a message is spam or not. This model can be trained on a dataset of SMS messages, containing both spam and non-spam messages. The model learns from the dataset and builds a decision boundary that can accurately classify the messages into spam and non-spam categories.

* The development of the spam SMS classifier using machine learning algorithms is an essential step towards improving the user experience with messaging services.
* The use of a spam SMS classifier can help prevent unwanted messages from reaching the user's inbox, reducing the likelihood of fraudulent activities and promoting better user engagement with messaging services.
* This project aims to develop a spam SMS classifier using machine learning algorithms and evaluate its performance.

* The project will involve collecting a dataset of SMS messages, preprocessing the data, extracting features, building a model, and evaluating the performance of the model using various metrics.

# CHAPTER – 2 PROBLEM FORMULATION

The increasing use of mobile phones and messaging services has led to a rise in the number of spam SMS messages. Spam SMS messages are unsolicited messages that are sent to a large number of recipients, often for malicious purposes such as phishing or fraud. These messages can be annoying and can also pose a security risk to the recipient's personal information. Therefore, there is a need to develop a spam SMS classifier that can accurately identify spam messages and prevent them from reaching the recipient's inbox.

The problem with spam SMS messages is that they often contain misspellings, abbreviations, and other irregularities that make it difficult to identify them using traditional rule-based methods. Therefore, machine learning algorithms can be used to analyze the content of the SMS messages and classify them into spam and non-spam categories.

The problem formulation for this project is to develop a spam SMS classifier using machine learning algorithms that can accurately identify spam messages and prevent them from reaching the recipient's inbox. The proposed solution should be able to handle the variability in the content of the SMS messages and classify them accurately. The solution should also be efficient and scalable to handle large volumes of SMS messages.

# CHAPTER – 3 PROPOSED SOLUTION

The purpose of this project is to develop a spam SMS classifier using machine learning algorithms to accurately identify spam messages and prevent them from reaching the recipient's inbox. The proposed solution involves the following steps:

* Data Collection:
  + Collecting a dataset of SMS messages containing both spam and non-spam messages.
* Data Preprocessing:
  + Preprocessing the data to remove unwanted characters, numbers, punctuations, and stopwords. Text normalization, tokenization, and feature engineering are also performed.
* Model Development:
  + Developing a spam SMS classifier model using the Naive Bayes algorithm. The model is trained on the preprocessed dataset and builds a decision boundary to classify the messages into spam and non-spam categories.
* Model Evaluation:
  + Evaluating the performance of the spam SMS classifier using various metrics such as accuracy, precision, recall, and F1-score.

The proposed solution will help to improve the user experience with messaging services by preventing unwanted messages from reaching the user's inbox. The spam SMS classifier can be further improved by incorporating other machine learning algorithms and techniques such as lemmatization, word embeddings, and deep learning algorithms. The solution can also be extended to other text classification tasks such as sentiment analysis, topic modeling, and text summarization.

# CHAPTER – 4 FLOWCHART



Working of the Project

# CHAPTER – 5

**SOFTWARE AND HARDWARE REQUIREMENTS**

## Software-

1. PYTHON
2. MACHINE LEARNING
3. STREAMLIT
4. NLTK

Machine learning is a subfield of artificial intelligence that focuses on the development of algorithms and models that can learn from data and make predictions or decisions.

Machine learning is used in a wide range of applications, including image and speech recognition, natural language processing, fraud detection, and recommendation systems.

Machine learning algorithms can be categorized into supervised learning, unsupervised learning, and reinforcement learning, depending on the type of training data and feedback provided to the model.

NLTK (Natural Language Toolkit) is a popular open-source library for natural language processing (NLP) in Python. It provides a wide range of tools and resources for NLP tasks, including tokenization, stemming, tagging, parsing, and machine learning.

The NLTK library is designed to make it easy for developers to work with human language data and perform various tasks such as sentiment analysis, text classification, and information retrieval. It includes a wide range of corpora, lexicons, and models for various languages and domains, making it a valuable resource for researchers and developers.

## Hardware-

1. Intel i-5 processor
2. 8GB RAM
3. 1 GB Storage

# SNAPSHOTS

# 

# Distribution of Spam and Not Spam Data in our data Set

# 

# Scatterplot between the Columns

# 

# Most Frequent Words

# 

# All Models used with their Accuracy and Prescision

# CHAPTER – 8 CONCLUSIONS

In this project, we have developed a spam SMS classifier using machine learning algorithms to accurately identify spam messages and prevent them from reaching the recipient's inbox. The proposed solution involved collecting a dataset of SMS messages, preprocessing the data, developing a Naive Bayes model, and evaluating the performance of the model using various metrics.

The results of the project showed that the proposed spam SMS classifier achieved high accuracy, precision, recall, and F1-score in classifying SMS messages into spam and non-spam categories. The model was able to handle the variability in the content of the SMS messages and was efficient in processing large volumes of SMS messages.

The proposed solution has significant implications for improving the user experience with messaging services by preventing unwanted messages from reaching the user's inbox. The spam SMS classifier can be further improved by incorporating other machine learning algorithms and techniques such as lemmatization, word embeddings, and Machine learning algorithms.

In conclusion, the development of a spam SMS classifier using machine learning algorithms is an important step in addressing the problem of spam SMS messages and improving the security and privacy of mobile phone users. The proposed solution has the potential to be extended to other text classification tasks and can contribute to the development of more intelligent and efficient messaging services.

# REFERENCES

* **Sk learn Link->** [**https://scikit-learn.org/stable/**](https://scikit-learn.org/stable/)
* **StreamLit Link-> https://streamlit.io/**
* **Geeks For Geeks Link->** [**https://www.geeksfor**](http://www.geeksforgeeks.org/)**geeks.or**[**g**](http://www.geeksforgeeks.org/)
* **W3 Schools Link->** [**https://www.w3schools.com**](http://www.w3schools.com/)
* **Java-T-Point Link->** [**https://www.javatpoint.com**](http://www.javatpoint.com/)
* **Tutorials Point Link->** [**https://www.tutorialspoint.com**](http://www.tutorialspoint.com/)