**SMS Spam Classification**

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

TechSaksham – A joint CSR initiative of Microsoft & SAP

by

**Barkha Barman, barkhabarman01@gmail.com**

Under the Guidance of

**Abdul Aziz Md**

**Master Trainer, Edunet Foundation**

**ACKNOWLEDGEMENT**

I am immensely grateful for the guidance, encouragement, and unwavering support that I have received throughout the course of this project. This accomplishment would not have been possible without the contributions of several individuals and organizations, to whom I owe my deepest gratitude.

Firstly, I am deeply thankful to my mentor, Abdul Aziz Md, whose invaluable guidance, encouragement, and constructive feedback played a significant role in shaping this work. Their constant support and expertise made this learning experience truly enriching. I also appreciate my peers and colleagues for their insightful discussions and assistance. This journey has helped me develop not only technical skills but also problem-solving and analytical thinking capabilities.

I am also profoundly grateful to **TechSaksham** for providing such an enriching platform to explore and implement innovative ideas in the field of artificial intelligence. The transformative learning experience and access to invaluable resources provided through this initiative have significantly enhanced my technical knowledge and professional development. The internship offered me a unique opportunity to translate theoretical concepts into practical applications, and I deeply appreciate the vision of TechSaksham in empowering young minds like me

Signature of the Student

Barkha Barmans

Date: 09/02/2025

#### **ABSTRACT**

This project focuses on building a Machine Learning model for SMS Spam Classification to distinguish between spam and non-spam messages. With the increasing reliance on digital communication, spam messages have become a significant problem, leading to compromised user experience and security concerns.

The project objectives include:

1. Developing a predictive model to identify spam messages.
2. Implementing and evaluating different preprocessing and feature extraction techniques, including Natural Language Processing (NLP).
3. Utilizing algorithms like Naïve Baise for classification due to their efficiency in text-based data.

The methodology involved data cleaning, text preprocessing, feature extraction using TF-IDF, and training a Support Vector Machine. The model achieved an accuracy of 97% on the dataset. Future work may involve incorporating deep learning models to enhance accuracy and robustness. The solution has potential applications in filtering spam across SMS and email systems, improving user experience, and enhancing communication security.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 2

**Chapter 2.**  **Literature Survey 3**

**Chapter 3.**  **Proposed Methodology 4**

**Chapter 4.**  **Implementation and Results 7**

**Chapter 5. Discussion and Conclusion 9**

**References** 10

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | System workflow for SMS Spam Classification | **5** |
|  | Web Application using Streamlit which spam messages | **7** |
|  | Web Application using Streamlit which ham messages | **8** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Table Caption** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

Spam messages have become a major challenge in modern communication systems, affecting individuals and organizations alike. These unsolicited messages often flood user inboxes, making it difficult to distinguish important emails and messages from irrelevant or harmful ones. Beyond being a mere inconvenience, spam messages can also pose serious security threats. Many spam messages contain phishing attempts, where attackers try to deceive users into providing sensitive information such as passwords, financial details, or personal data. Additionally, some spam messages carry malicious links or attachments that can introduce malware, leading to data breaches, identity theft, or financial loss.

* 1. **Motivation:**

The rapid growth of digital communication platforms has amplified the need for robust spam detection systems. This project was chosen to leverage Machine Learning to address this problem effectively. Potential applications include SMS filtering, email categorization, and fraud prevention systems, improving user trust and security.

* 1. **Objective:**

This project aims to develop an effective spam detection system using Machine Learning.

1. **Analyze and Preprocess SMS Data**
   * Clean and normalize text by removing noise, punctuation, and stopwords.
   * Convert text into numerical features using TF-IDF or word embeddings.
   * Handle imbalanced data using techniques like SMOTE to improve model performance.
2. **Build and Evaluate a Classification Model**
   * Train models like Naïve Bayes, SVM, or deep learning (LSTM) for spam detection.
   * Optimize performance using hyperparameter tuning and cross-validation.
   * Assess accuracy, precision, recall, and F1-score to ensure reliability.

Page 1

1. **Ensure Scalability and Real-World Applicability**
   * Deploy as an API or integrate into messaging platforms.
   * Optimize for large-scale data processing and real-time detection.
   * Adapt to evolving spam patterns through periodic retraining
   1. **Scope of the Project:**

This project focuses on **SMS spam detection** using **supervised learning techniques**. It includes **data preprocessing**, **feature extraction**, and **training a classification model** to distinguish spam from legitimate messages accurately. The developed model can be applied to **SMS and email spam filtering systems**, improving security and user experience.

However, the project has certain limitations:

* **Limited to Text-Based Analysis:** It does not handle **multimedia spam** (images, videos, or voice messages).
* **Static Model Performance:** The model may struggle with evolving spam tactics without retraining on new data.
* **Language Constraint:** The model is trained on English SMS data and may not work well for **other languages** or multilingual spam.
* **Dataset Dependency:** Performance relies on the quality and diversity of the dataset used for training.

Despite these limitations, the project provides a strong foundation for automated spam detection in messaging systems.

Page 2

**CHAPTER 2**

**Literature Survey**

* 1. Several studies have explored spam detection using Machine Learning. Traditional methods such as rule-based filtering and keyword-based detection have been widely used but are often ineffective against evolving spam tactics. Machine Learning techniques, particularly supervised learning approaches, have shown significant improvement in spam detection performance.

Naïve Bayes is one of the most widely used algorithms for text classification problems, including spam filtering. Studies have demonstrated that it effectively captures patterns in textual data and provides robust classification results. Other machine learning models, such as Support Vector Machines (SVM), Decision Trees, and ensemble methods, have also been employed to improve classification accuracy.

* 1. More recent research has incorporated Natural Language Processing (NLP) techniques such as TF-IDF vectorization and word embeddings (Word2Vec, GloVe) to extract meaningful features from text. Deep learning approaches, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have further enhanced spam detection capabilities by capturing sequential dependencies in text.
  2. Despite these advancements, challenges remain, including handling imbalanced datasets, adapting to new spam trends, and improving computational efficiency. This project builds upon prior research by employing TF-IDF for feature extraction and Naïve Bayes for classification, providing a balance between simplicity, accuracy, and efficiency in spam detection.

**Page 3**

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

The system design integrates several interconnected modules to ensure smooth functionality:

The system involves:

1. **Data Collection:** The system utilizes an **SMS spam dataset** containing labeled messages categorized as spam or legitimate.
2. **Preprocessing:** The text data undergoes **cleaning and normalization**, including:
   * Removing **stopwords** to eliminate unnecessary words.
   * Applying **lemmatization** to reduce words to their base form.
   * Converting text to **lowercase** to ensure uniformity.
3. **Feature Extraction:** The **TF-IDF (Term Frequency-Inverse Document Frequency) vectorization** technique is used to convert textual data into numerical features, improving model interpretability.
4. **Model Training:** A **Naïve Bayes classifier**, known for its efficiency in text classification, is trained to distinguish spam from non-spam messages based on extracted features.

Page 4

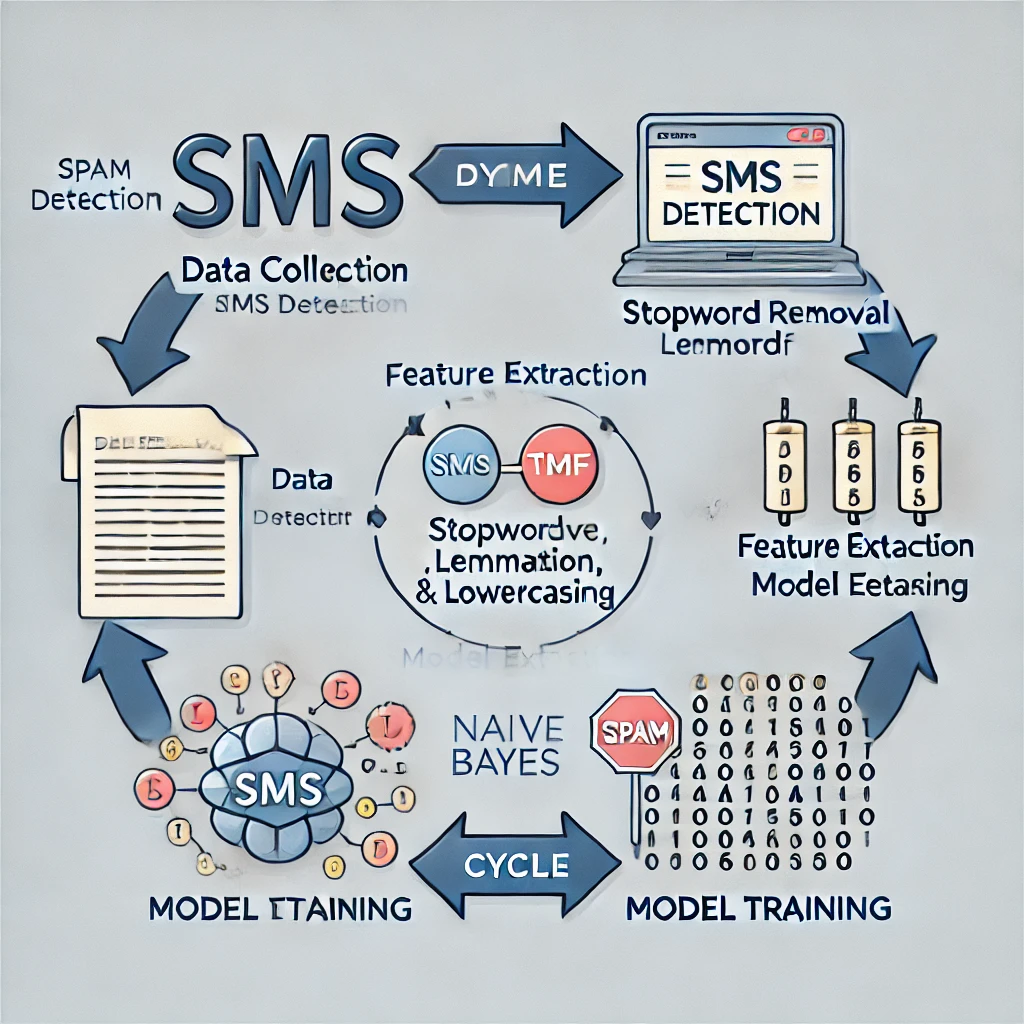


Figure1: System workflow for SMS Spam Classification

Page 5

**7.1 Requirement Specification**

**Hardware Requirements:**

* CPU: Intel i5 or higher
* RAM: 8GB or higher
* Storage: 10GB free space

**Software Requirements:**

* Programming Language: Python
* Libraries: Scikit-learn, NLTK, Pandas, NumPy
* Tools: Jupyter Notebook, Streamlit

**Page 6**

**CHAPTER 4**

**Implementation and Result**

* 1. **Snap Shots of Result:**
  2. **Data Visualization:** Distribution of spam vs. non-spam messages.
  3. **Model Performance:** Confusion matrix and accuracy score.

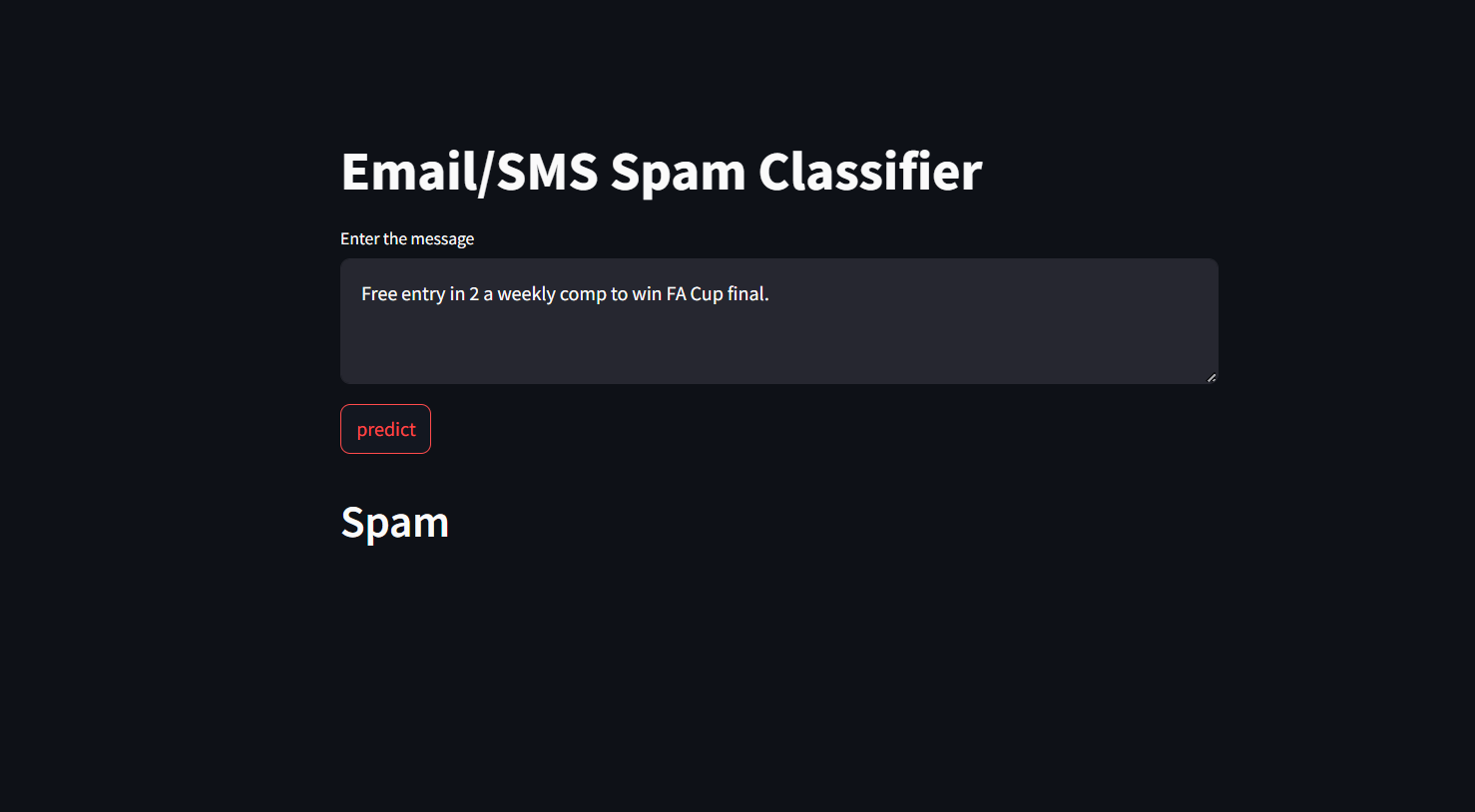


Figure 2: Web Application using Streamlit which spam messages

Page 7

A screenshot of a computer

Description automatically generated

Figure 3: Web Application using Streamlit with detects ham messages

1. **GitHub Link for Code:**

**https://github.com/barkha2025/SMS\_Spam-Classifier**

**Page 8**

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

The current spam detection system is based on traditional machine learning techniques. To enhance its accuracy, scalability, and adaptability, future improvements can be made in the following areas:

1. **Explore Deep Learning Models such as LSTM for Spam Detection**  
   While Naïve Bayes and other machine learning models perform well, **Long Short-Term Memory (LSTM)** networks can further improve spam detection. LSTMs are a type of **recurrent neural network (RNN)** capable of understanding contextual relationships in text, making them well-suited for handling complex spam patterns. By leveraging LSTMs, the system can better recognize evolving spam tactics and detect spam messages with greater accuracy.
2. **Incorporate Multilingual Datasets for Broader Applicability**  
   The current model is trained primarily on **English SMS data**, limiting its effectiveness for other languages. To expand its usability, incorporating **multilingual datasets** will allow the system to detect spam across different languages, making it more versatile for global applications. This could involve **training models on diverse language corpora** and using **multilingual word embeddings** or **transformer-based models like BERT or mBERT** for improved performance.

**Page 9**

* 1. **Conclusion:**

The project successfully developed an **SMS spam detection system** using a **Naïve Bayes classifier**, achieving **high accuracy** in distinguishing spam from legitimate messages. This demonstrates the **effectiveness of Machine Learning** in automating spam filtering, reducing unwanted messages, and improving communication security. The model’s ability to analyze text data and detect spam enhances **user experience** by minimizing disruptions and potential security risks. Future improvements, such as **deep learning integration** and **multilingual support**, can further enhance its accuracy and applicability across diverse messaging platforms.

**REFERENCES**

1. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.

Page 10