SMS Spam Detection System Using NLP

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

Ayush Singh, 2ayushsingh@gmail.com

Under the Guidance of

Rathod Jay Sir

ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere gratitude to everyone who helped me, directly or indirectly, during this thesis work.

First, I would like to express my special thanks of gratitude to **Rathod Jay Sir** and **P Raja Sir** for mentoring, guiding, and motivating me throughout my project, SMS Spam Detection using NLP.

I would also like to express my thanks to **Pavan Sumohana Sir** and **Edunet Foundation** for their support to attend Internship.

I would also like to acknowledge my family and friends for their support and contributing their ideas and perspective, which enriched my project.

ABSTRACT

Today **spam** messages are very common problem, most of the population is affected by spam messages and it is very necessary to identify these type of messages as spam messages may leads to malware distribution, scam, financial or personal data loss. The **SMS Spam Detection System** project aims to identify these **spam** messages and help users protect their data.

The main aim of this project is to build a simple GUI based detection system which can easily identify spam messages leaving the legitimate messages as it is.

The building of this project approach involves collecting a dataset of SMS messages and preprocessing (removing error, duplicate words, empty values, stop-words, etc.) the dataset. Then, techniques like **TF-IDF** (Term Frequency - Inverse Document Frequency) is used to convert the text into a format that machine learning models can understand. After testing the models, including Naïve Bayes, Support Vector Machine (SVM), and Random Forest for finding the most accurate one.

In conclusion, this project demonstrates the use of NLP and machine learning for spam messages identification ensuring the security of the user from cyber criminals. In future larger datasets, deep learning and automation can be used for better results.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

In today's digital era, most of the public services are linked to people's mobile phone number on which they get messages. Cyber criminals also take advantage of this feature to trick people for their gain by **Smishing** (sending **fraudulent** or **spam** messages).

So, it is necessary to distinguish between the **spam** and **non-spam (ham)** messages but sometimes it is not so easy to distinguish as criminals are also evolving and using more sophisticated ways to trick people.

Thus, we need a tool to make this task easy and reduce our efforts. **SMS Spam Detection using NLP** is such a tool for our help. It is designed by using **Natural Language Processing** (NLP) for distinguishing between spam and ham messages.

Why is this problem significant?

I. This problem directly impacts individual of each sector whether the person belongs to technical or non-technical sector spam messages are a threat for everyone.
II. The consequences of not addressing this issue includes increase in the number of victims of frauds by spam SMS and financial loss.

1.2 Motivation:

This project was chosen to address the growing concern about the spam messages. According to a report **(Smishing Scams: The Rising Threat in India)** by Quick Heal which is known for making antivirus software, "The global prevalence of **smishing text** attacks has been on rise, with research indicating a 328% increase in smishing attemps worldwide in 2020 alone. In India, the situation is equally alarming, with the country ranking among the top fine nations most affected by **smishing security** threats." Additionally, the project aligns with my interest in both **Artificial Intelligence** and **Cyber Security** which motivates me of understanding this serious problem and work on this project.

The **SMS Spam Detection System** has numerous potential applications. It can help individuals filter out unwanted messages, protecting them from phishing attempts and scams, businesses can implement these systems to ensure that employees receive only

relevant communications, mobile applications like Truecaller, utilize spam detection algorithms to automatically filter out spam SMS. There are plenty of such applications like this.

The project's impact is expected to be significant. Using this detection system, the number of victims of spams messages can be decreased, resulting in very less financial losses, and less number of SMS related cyber-crimes.

1.3 Objective:

- The primary objective of this project is to develop a system for the detection of **spam** messages to prevent any type financial and personal data loss.
- To develop a simple and attractive Graphical User Interface (GUI) for better user experience and easy classification.
- To implement different machine learning algorithms for optimal result.

1.4 Scope of the Project:

Scope:

- This project aims to develop a web-based GUI where users can detect **spam** messages.
- The programming is done using **Python** language and the GUI and **Streamlit** (Python) has been used to make web GUI.
- It uses **NLP** and **Machine Learning** algorithms.

Limitations:

- This detection system can understand only English language.
- We have to manually paste the messages to the input box of the detection system to differentiate.
- The classification is **not** 100% accurate. It may sometime classify a **spam** message as **non-spam**.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

The word spam generally means some unwanted text sent or received through sms, social media sites such as Facebook, Twitter, e-mail, etc. It is generated by spammers to divert the attention of the users for the purpose of marketing, financial data theft, personal data theft or spreading malware, etc. The spam messages are sent in bulk to various users, with the intention of tricking them into clicking on fake advertisements and spreading malware on their devices.

There are several studies have been conducted all over the world on **SMS Spam detection System.** These studies reveals a variety of approaches and methodologies employed by researchers over the years. Here are some of them:

1. **Machine Learning and Deep Learning Techniques:**
   - A study by Gupta et al. (2018) compared various machine learning classifiers, including Decision Trees, SVM, and CNN, for SMS spam detection. The CNN model achieved the highest accuracy of 99.19% across two datasets, demonstrating the effectiveness of deep learning in this domain.
   - In another research project, a Long Short-Term Memory (LSTM) model was utilized, achieving an accuracy of 98.18%. This model showed a strong ability to detect spam messages with a misclassification rate of only 0.74%, highlighting the potential of deep learning techniques in improving spam detection accuracy.
2. **Dataset Challenges and Methodologies:**
   - Almedia et al. (2011) pointed out the limitation of publicly available datasets and emphasized the need of larger datasets for effective spam detection. They collected a substantial dataset and found that SVM performed best with an accuracy of 97%.
   - Reaves et al. (2016) focused on enhancing detection precision by using clustering methods, which raised precision form 23.8% to 94.1% and recall from 61.3% to 88.8%. Their work highlighted the importance of dataset size and feature selection in improving spam detection outcomes.
3. **Systematic Literature Reviews:**
   - A systematic literature review conducted by Sisodia et al. (2020) analyzed various SMS spam detection techniques published between 2006 and 2016. The review highlighted the challenges unique to SMS spam detection, such as

regional content and abbreviated language, which complicate classification efforts.

- The review also noted that while there is substantial research on email spam detection, SMS spam detection remains underexplored, indicating a significant opportunity for future research in this area.

4. **Hybrid Approaches and Feature Selection:**
   - Research by Roy et al. (2020) explored both machine learning methods (like Naive Byes and Random Forest) and deep learning methods (such as CNN and LSTM). They reported and impressive accuracy of 99.4% with CNN, but acknowledged that their study was limited to English text messages.
   - The work of Alzahrani and Ghorbani (2016) introduced a mobile botnet detection module using clustering algorithms, achieving accuracies around 96% across different datasets, showcasing the versatility of clustering methods in spam detection.
5. **Limitations in Existing Solutions:**
   - Many studies have indicated that existing solutions that often struggle with evolving spam tactics and insufficient dataset diversity, which can lead to high false-negative rates in real-world applications.
   - Additionally, while some models achieve high accuracy rates, they may note generalize well across different languages of types of SMS content, limiting their applicability in diverse contexts.

These studies collectively highlight the program made in SMS spam detection while also identifying gaps that future research address, such as improving dataset diversity, adapting to new spamming techniques, and exploring multilingual capabilities in spam detection systems.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

Identifying spam messages from inbox has been done by various methodologies. Below are some of the approaches.

i. Sethi, P., Bhandari, V., &Kohli, B. (2017). “SMS spam detection and comparison of various machine learning algorithms.” 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN).

ii. DelviaArifin, D., Shaufiah, &Bijaksana, M. A. (2016).“Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FP-growth and Naive Bayes Classifier.”2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob).

iii. Agarwal, S., Kaur, S., &Garhwal, S. (2015). “SMS spam detection for Indian messages.” 2015 1st International Conference on Next Generation Computing Technologies.

iv. Gupta, M., Bakliwal, A., Agarwal, S., &Mehndiratta, P. (2018). "A Comparative Study of Spam SMS Detection Using Machine Learning Classifiers." 2018 Eleventh International Conference on Contemporary Computing (IC3).

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

1. **High False Positive and False Negative Rates:**
   Some models incorrectly classify legitimate messages as spam (false positives) or fail to detect actual spam (false negatives).
   To reduce this gap or limitation feature extraction technique like **TF-IDF (**Term Frequency Inverse Document Frequency) and testing different machine learning algorithms, we aim to minimize classification errors and improve accuracy.

2. **Limited Generalization Across Different Datasets:**
   Many existing models perform well on specific datasets but fail when applied to different SMS datasets due to variations in language, message structure, and spam tactics.
   We will train and test our model on a diverse dataset to improve generalization and ensure robustness across different types of spam messages.

3. **Inability to Adapt to Evolving Spam Patterns:**
   Spammers continuously change their messaging patterns, making it difficult for static models to keep up.
   We will implement a regular retraining mechanism so the model can learn from new spam trends and remain effective over time.

CHAPTER 3

Proposed Methodology

3.1 System Design

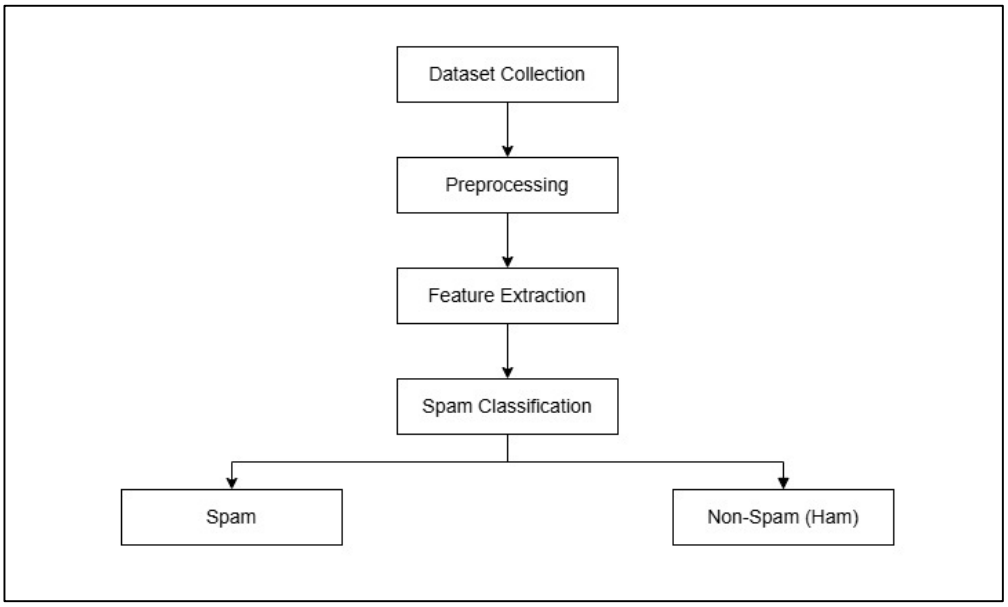

Fig 1: System Design

i. **Dataset Collection:**
This is the first phase in spam identification, dataset comprises both spam and non-spam (ham) messages from different sources.
Some of the datasets are: spam.csv (Kaggle), SMS spam collection (UCI).

ii. **Preprocessing:**
Pre-processing is a significant technique for cleaning the raw data in a dataset. It is necessary to remove any type of unwanted data from the dataset before training the model. The pre-processing involves removal of redundancy, errors, empty values, stop words, punctuations and special characters.

iii. **Feature Extraction:**
Because many machine learning algorithms rely on numerical data rather than text, it is required to convert the text input into numerical vectors. The goal is to extract meaningful information from a text. Some of the feature extraction strategies are: Bag of words (**BoW**), N-grams, Term frequency-inverse document frequency (**TF-IDF**), Word2Vec, etc.

iv. **Spam Text Classification Techniques:**
Text Classifiers can organize and categorize practically any sort of material, including documents and internet text. Text classification is an important stage in natural language processing, with applications ranging from sentiment analysis to subject labelling and spam detection. Some of the techniques of classifying the text are rule based system, machine learning and hybrid approach.

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

- **Processor:** Intel Core i3 or higher (or equivalent AMD Ryzen 3 or higher)
- **RAM:** 8 GB (16 GB recommended)
- **Storage:** 256 GB SSD and 512 GB HDD (minimum)
- **Peripherals:** Standard keyboard and mouse
- **GPU (optional):** NVIDIA GTX 1650 or higher

3.2.2 Software Requirements:

i. **Programming Languages:** Python
ii. **Backend:** Streamlit (Python)
iii. **Libraries Used:** Scikit-learn, nltk, wordcloud
iv. **Operating System:** Windows/ Linux/ Mac
v. **Version Control:** Git
vi. **Development Environment:** Jupyter Notebook/ Google Colab for development of the model and Visual Studio Code/ PyCharm for debugging, and version control.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

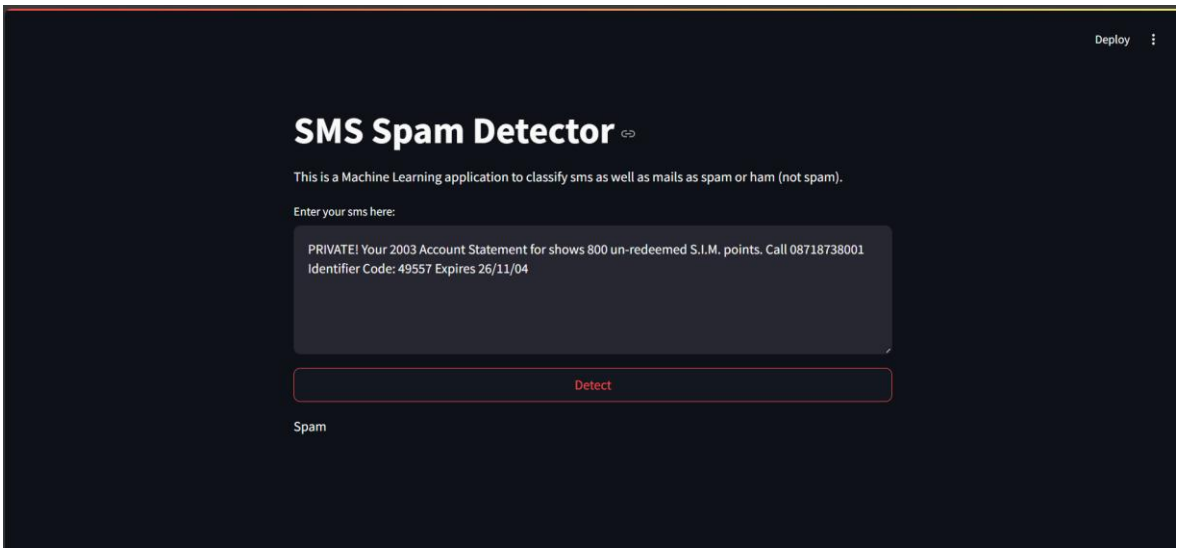

Fig.2: Snapshot-1

In Fig. 4.1, it is shown that a SMS has been classified as **spam** by the **SMS Spam Detection Model**. This image is showing that the detection model is working properly.

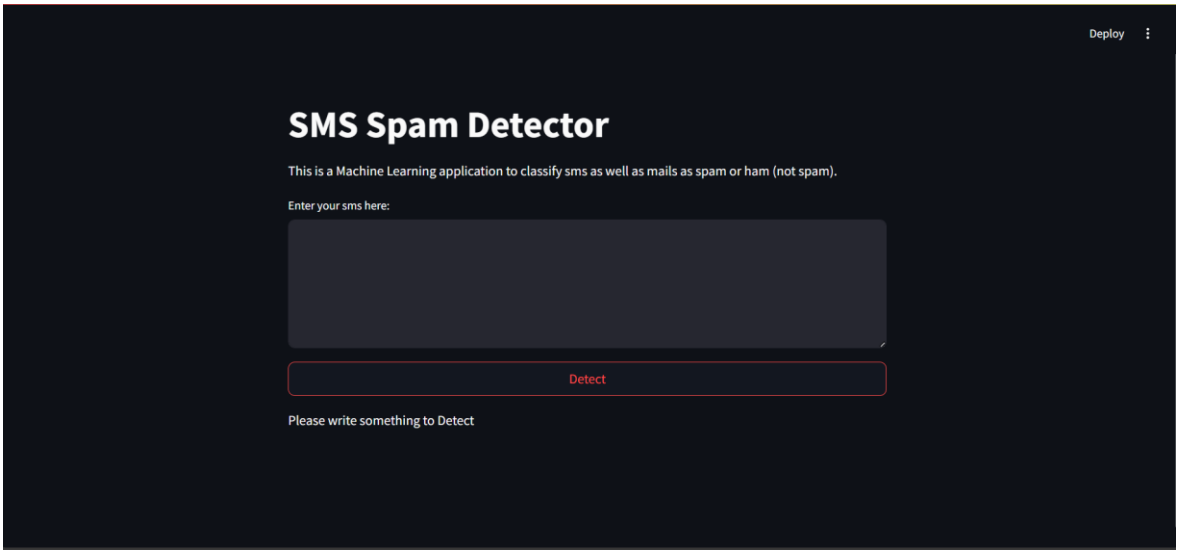

Fig.3: Snapshot-2

In Fig.4.2, it is shown that when the input box is left empty then a message appears for the user to write something in the input box to classify by the **SMS Spam Detection Model**. This image is showing that the detection model is working properly.

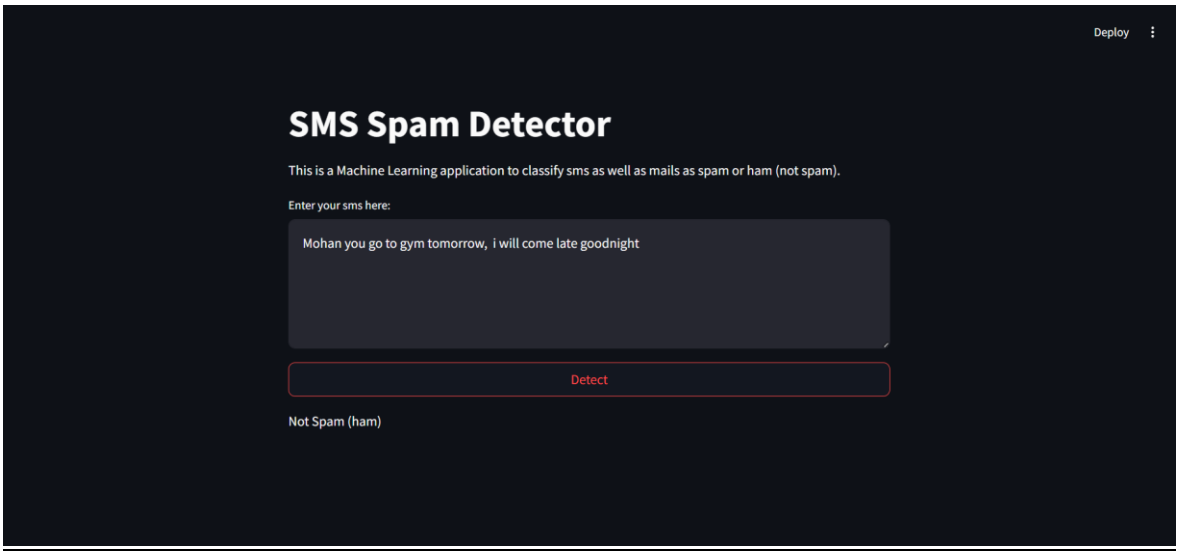

Fig.4: Snapshot-3

In Fig.4.3, it is shown that a SMS has been classified as **non-spam (ham)** by the **SMS Spam Detection Model**. This image is showing that the detection model is working properly.

Fig.5: Snapshot-4

In Fig.4.3, it is shown that a **spam** SMS has been classified as **non-spam (ham)** by the **SMS Spam Detection Model**. This image is showing that the detection model is **not** working properly.

4.2 GitHub Link for Code:

https://github.com/ayushSingh0112/AICTE-Edunet-Internship-SMS_Spam-Detection-System.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

As the model sometimes fails to identify the spam message, it can be improved by using a larger dataset.

The model can be implemented using **Deep Learning** Techniques for better outcomes.

This project can be integrated with messaging applications so that we don't have to manually check for spam messages. It will automatically classify the incoming messages as spam or ham and alert the users for spam messages.

5.2 Conclusion:

The **SMS Spam Detection System** was developed to easily identify the **spam** messages by providing a very simple Graphical User Interface. It is a very easy to use web-application. It can be used by individuals to identify **spam** messages from a bulk of unclassified messages keeping them safe from any malware attack or losing personal or financial data to an ill intended criminal.

While the project meets its objectives, future enhancement could be integrated to smartphones and automating this tools so that it can directly fetch the incoming messages for real-time identification of spam.

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.

[2]. Luo GuangJun, Shah Nazir, Habib Ullah Khan, Amin Ul Haq, “Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms”, Wiley Online Library, Security and Communication Networks, Volume 2020, Issue 1/8873639, July, 2020.

[3]. Smishing Scams: The Rising Threat of Cybercrime in India, Quick Heal.

[4]. Ravi H Gedam, Sumit Kumar Banchhor, “SMS Spam Detection Using Machine Learning”, Journal of Computational Analysis and Applications, Volume. 33, No. 4, 2024.

[5]. Bollam Pragna, M. RamaBai, “Spam Detection using NLP using techniques", International Journal of Recent Technology and Engineering (JRTE) 7, ISSNN: 2277, Vlumer-8, Issue-2S11, September 2019

[6]. “Smishing Scams: The Rising Threat of Cybercrime in India”, Quick Heal, December 2024.