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

Anna Tomson,

annatomson27@gmail.com

Under the Guidance of

Rathod Jay

ACKNOWLEDGEMENT

I would like to extend my heartfelt gratitude to all those who have supported and guided me throughout this project. My sincere thanks to Rathod Jay for their invaluable mentorship, constructive feedback, and constant encouragement that helped shape this project into its current form. I am also deeply grateful to the TechSaksham initiative for providing this exceptional learning platform that enabled me to explore advanced AI concepts and develop practical skills. Additionally, I want to thank my peers and colleagues for their collaboration and insights during various stages of this project. Lastly, my appreciation goes to my family and friends for their unwavering support and motivation, which inspired me to overcome challenges and achieve my goals.

ABSTRACT

The SMS Spam Detection System Using NLP addresses the challenge of identifying and filtering spam messages efficiently. Traditional rule-based filters struggle to adapt to evolving spam techniques, leading to inefficiencies in message filtering. This project leverages Natural Language Processing (NLP) and machine learning to develop a robust spam detection system capable of analyzing text patterns and adapting to new spam types. Key objectives include automating the classification process, improving accuracy, and enhancing adaptability. The methodology involves preprocessing text data, feature engineering using TF-IDF, and training models such as Logistic Regression and Naïve Bayes. Results demonstrate significant improvement in spam detection, achieving high precision and recall rates. This project underscores the potential of NLP in improving communication systems and mitigating spam-related challenges.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

The proliferation of spam SMS messages has become a significant challenge, overwhelming users and diminishing the efficiency of communication systems. Users often receive a large volume of messages daily, making it increasingly difficult to manually sort important messages from spam. Traditional rule-based spam filters often fail to keep up with the constantly evolving tactics used by spammers, leading to inefficiencies and high false-positive rates. Moreover, these conventional systems lack the ability to adapt dynamically to new types of spam, further exacerbating the problem.

To address these challenges, there is a need for an intelligent, automated system capable of analyzing and classifying SMS messages effectively. By leveraging machine learning and Natural Language Processing (NLP) techniques, such a system can detect patterns, adapt to evolving spam strategies, and significantly enhance spam detection accuracy. The goal of this project is to design and implement a robust SMS Spam Detection System that helps users focus on the messages that truly matter while ensuring minimal manual intervention and maximum adaptability.

1.2Motivation:

The project aims to automate the spam detection process, minimizing manual effort and improving message relevance. With spam causing distractions and potential security threats, a machine learning-based system offers a scalable and adaptive approach. Spam messages often contain misleading or harmful content, which can result in financial losses or compromise personal information. Traditional spam filters fail to address the dynamic nature of spam, making it essential to explore innovative approaches. By using advanced NLP and machine learning techniques, this project aims to provide a scalable and adaptive solution that addresses these shortcomings, ensuring users receive only the most relevant and trustworthy messages.

1.3Objective:

1) Automate the classification of SMS messages as spam or non-spam.
2) Develop a scalable model that adapts to evolving spam patterns.
3) Improve spam detection accuracy using advanced NLP techniques.

1.4Scope of the Project:

The project focuses on developing a spam detection model using NLP techniques and machine learning. It includes dataset preprocessing, feature engineering, model training, and evaluation. Limitations include dependency on the quality of training data and potential biases in classification.

CHAPTER 2

Literature Survey

Traditional spam detection systems rely on rule-based methods and keyword matching. While easy to implement, these systems struggle with new and evolving spam patterns, often leading to high false-positive rates.

Recent studies highlight the effectiveness of machine learning models like **Naïve Bayes**, **Support Vector Machines (SVMs)**, and **Neural Networks** in improving spam detection accuracy. These models analyze text patterns and adapt to diverse datasets, reducing the limitations of rule-based methods. Advanced techniques such as **Natural Language Processing (NLP)** have further enhanced feature extraction, enabling more accurate classification of spam messages.

Despite these advancements, challenges such as computational complexity and handling imbalanced datasets remain. This project addresses these gaps by combining NLP techniques with machine learning algorithms to create a more adaptive and efficient spam detection system.

CHAPTER 3

Proposed Methodology

3.1 System Design

The proposed solution comprises the following steps:

1. **Data Collection**: Using publicly available SMS spam datasets like the UCI SMS Spam Collection Dataset to ensure diverse and comprehensive training data.
2. **Preprocessing**: Preparing the data by cleaning, tokenizing, lemmatizing, and removing stop words to ensure high-quality input for the model.
3. **Feature Engineering**: Utilizing the TF-IDF (Term Frequency-Inverse Document Frequency) technique to convert text data into numerical vectors for machine learning.
4. **Model Training**: Experimenting with machine learning algorithms like Logistic Regression and Naïve Bayes to build the classification model.
5. **Evaluation**: Measuring model performance using metrics such as precision, recall, F1-score, and accuracy to ensure robustness.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

- **Memory**: At least 8 GB RAM for handling large datasets efficiently.
- **Processor**: Intel i5 Processor or higher to enable faster model training and testing.

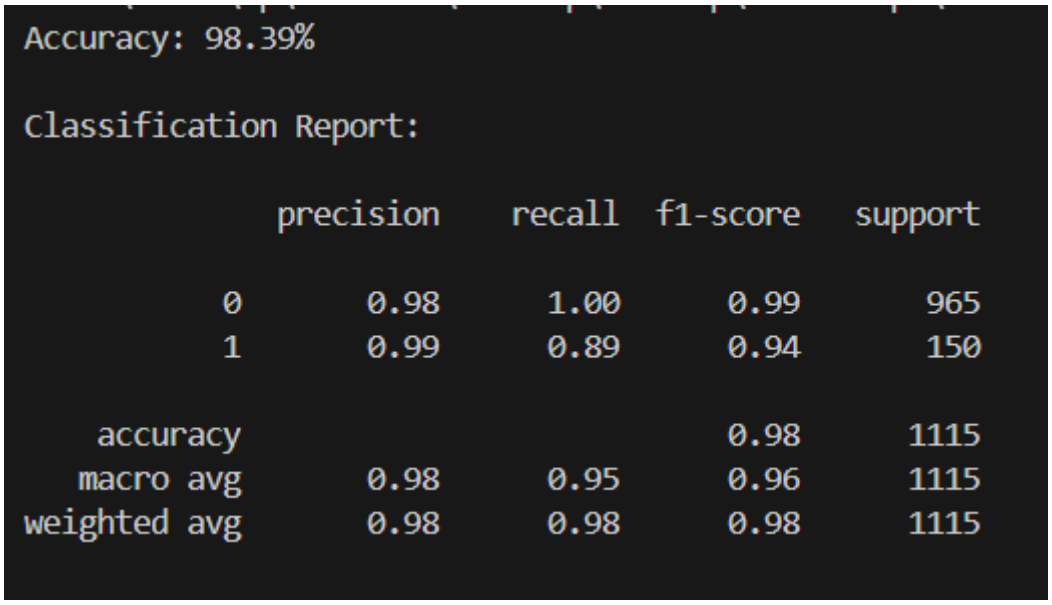
3.2.2 Software Requirements:

- **Programming Language**: Python, known for its extensive libraries and tools for machine learning and NLP.
- **Libraries**: Scikit-learn for machine learning, NLTK for text preprocessing, Pandas for data manipulation, and Matplotlib/Seaborn for data visualization.
- **IDE**: Jupyter Notebook for interactive code development and visualization.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:



Accuracy: 98.39%

Classification Report:

	precision	recall	f1-score	support
0	0.98	1.00	0.99	965
1	0.99	0.89	0.94	150
accuracy			0.98	1115
macro avg	0.98	0.95	0.96	1115
weighted avg	0.98	0.98	0.98	1115

Fig 1: The accuracy of the obtained result.

```
--- SMS Spam Detector ---
Enter a message (or type 'exit' to quit): Congratulations! You've won a free trip to Paris!
Prediction: SPAM
```

Fig 2: Given an input, the Spam Detector predicts it as SPAM message.

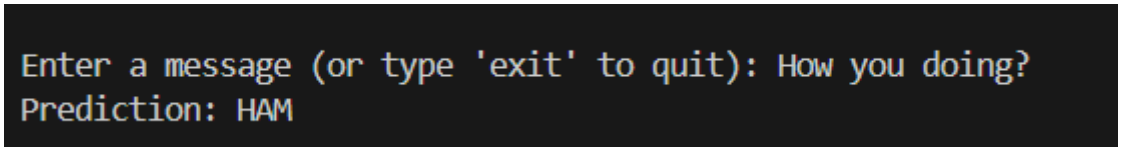


Fig 3: Given an input, the Spam Detector predicting the output as Not Spam (HAM).

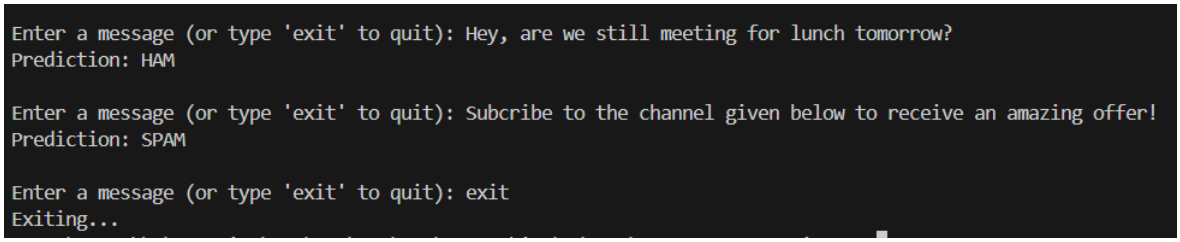


Fig 4: More inputs and outputs and finally exiting.

4.2GitHub Link for Code:

https://github.com/AnnaTomson/SMS_Spam_Detection

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

1. Expand the dataset with real-time SMS data to improve model generalization.
2. Explore advanced NLP models, such as BERT or GPT, for enhanced performance.
3. Integrate the spam detection system into mobile or web-based applications for practical deployment.

5.2 Conclusion:

The SMS Spam Detection System effectively automates spam classification, achieving significant improvements in accuracy and adaptability. By leveraging NLP and machine learning techniques, this project highlights the potential to mitigate spam-related challenges, paving the way for more secure and efficient communication systems.

REFERENCES

1. UCI ML Repository - SMS Spam Collection Dataset.
2. Scikit-learn Documentation.
3. "Text Mining with TF-IDF" - Research Paper.