DAYANANDA SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

Bachelor of Technology

in

Computer Science and Engineering

(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

A Project Report On

REINFORCEMENT LEARNING FOR GAME PLAYING: AI AGENT TRAINING STRATEGIES

Submitted By

Pruthviraj S R	**ENG22AM0042**
Rajat S U	**ENG22AM0045**
Shashwat Dodamani	**ENG22AM0058**

Under the guidance of

Prof Pradeep Kumar K

Assistant Professor, CSE(AIML), DSU

2024 - 2025

Department of Computer Science and Engineering (AI & ML)

DAYANANDA SAGAR UNIVERSITY

Bengaluru - 560068

SCHOOL OF
ENGINEERING

Dayananda Sagar University

Kudlu Gate, Hosur Road, Bengaluru - 560 068, Karnataka, India

Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning)

CERTIFICATE

This is to certify that the project entitled **EMAIL SPAM DETECTION** is a bonafide work carried out by **Pruthviraj S R (ENG22AM0042)**, **Rajat S U (ENG22AM0045)** and **Shashwat Dodamani (ENG22AM0058)** in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering (Artificial Intelligence and Machine Learning), during the year 2024-2025.

Prof. Pradeep Kumar K
Assistant Professor
Dept. of CSE (AIML)
School of Engineering
Dayananda Sagar University

Signature ....

Dr. Vinutha N
Project Co-ordinator
Dept. of CSE (AIML)
School of Engineering
Dayananda Sagar University

Signature ....

Dr. Jayavrinda Vrindavanam
Professor & Chairperson
Dept. of CSE (AIML)
School of Engineering
Dayananda Sagar University

Signature ....

Name of the Examiners:
1 ....
2 ....
3 ....

Signature with date:
....
....
....

Acknowledgement

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to **School of Engineering and Technology, Dayananda Sagar University** for providing us with a great opportunity to pursue our Bachelor’s degree in this institution.

We would like to thank **Dr. Udaya Kumar Reddy K R**, Dean, School of Engineering and Technology, Dayananda Sagar University for his constant encouragement and expert advice.

It is a matter of immense pleasure to express our sincere thanks to **Dr. Jayavrinda Vrindavanam**, Professor & Department Chairperson, Computer Science and Engineering (Artificial Intelligence and Machine Learning), Dayananda Sagar University, for providing right academic guidance that made our task possible.

We would like to thank our guide **Prof. Professor**, Assistant Professor, Dept. of Computer Science and Engineering, for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and fruitful culmination of the project.

We would like to thank our Project Coordinator **Dr. Vinutha N** as well as all the staff members of Computer Science and Engineering (AIML) for their support.

We are also grateful to our family and friends who provided us with every requirement throughout the course.

We would like to thank one and all who directly or indirectly helped us in the Project work.

Pruthviraj S R	**ENG22AM0042**
Rajat S U	**ENG22AM0045**
Shashwat Dodamani	**ENG22AM0058**

SPAM EMAIL DETECTION

Pruthiraj S R, Rajat S U, Shashwat Dodamani

Abstract

In the digital age, email has become one of the most common communication tools. However, with its widespread usage comes the issue of spam emails—unsolicited, unwanted messages that often contain harmful links, advertisements, or phishing attempts. These emails not only clutter users' inboxes but can also lead to financial and privacy risks. To address this issue, our project implements a machine learning-based solution for detecting spam emails with high accuracy. We explore and apply techniques such as natural language processing (NLP), text preprocessing, and model training using classical machine learning algorithms like Naive Bayes. Our system utilizes the UCI SMS Spam Collection Dataset, performs comprehensive data cleaning, and trains the model on the processed data. This report details each phase of the project from dataset acquisition, model development, evaluation, to deployment and prediction. The ultimate goal is to design an efficient, reliable, and real-time spam detection system that can be integrated into email services.

Sustainable Development Goals

The project "Spam Email Detection using Natural Language Processing (NLP)" aligns with several United Nations Sustainable Development Goals (SDGs) by promoting secure communication, innovation, and digital resilience. It primarily supports **Goal 9 (Industry, Innovation, and Infrastructure)** by leveraging advanced technologies such as NLP and machine learning to develop a robust and innovative system that strengthens the cybersecurity infrastructure of digital communication platforms. Additionally, the project aids in **Goal 4 (Quality Education)** by serving as a practical learning model for students and researchers, thereby enhancing their understanding of real-world applications in artificial intelligence and data science. It also contributes to **Goal 16 (Peace, Justice, and Strong Institutions)** by helping reduce digital crimes like phishing and identity theft through effective spam detection, which enhances trust in online communication systems and protects users from cyber threats. Furthermore, it supports **Goal 17 (Partnerships for the Goals)** by encouraging collaboration through the use of open-source tools, publicly available datasets, and community-driven innovation. Indirectly, the project also aligns with **Goal 8 (Decent Work and Economic Growth)** by minimizing disruptions caused by spam in work environments, thereby increasing overall productivity and promoting a secure digital workplace. Through these contributions, the project not only addresses a technical problem but also plays a role in advancing the broader global agenda for sustainable development.

Contents

1 Introduction

With the rapid expansion of the internet and the increasing dependence on electronic communication, email has become one of the primary tools for communication, both professionally and personally. However, this popularity has also attracted malicious actors who misuse the platform for spamming. Spam emails are unsolicited, irrelevant, and often harmful messages sent over the internet, typically to a large number of users. They can range from annoying advertisements to dangerous phishing attempts and fraud. This growing threat necessitates the development of robust mechanisms for spam detection.

Traditional spam filters relied heavily on rule-based or heuristic techniques, which, while effective to some extent, struggled to keep up with the constantly evolving tactics used by spammers. As Natural Language Processing (NLP) and Machine Learning (ML) technologies have advanced, they offer a powerful alternative to build adaptive and intelligent spam detection systems.

Spam detection using NLP leverages the patterns in the language used in emails to classify them as spam or not spam. This technique analyzes the structure, syntax, and semantics of messages, converting them into numerical representations that can be understood by machine learning models. By training these models on labeled datasets, the system learns to predict the likelihood of a new email being spam.

1.1 Scope

This project focuses on building a spam detection system using Natural Language Processing techniques combined with a supervised machine learning model, specifically Logistic Regression. The model is trained on a labeled dataset of emails, preprocessed through text normalization, tokenization, and vectorization. The goal is to achieve a high accuracy in classification while minimizing false positives and false negatives.

The scope includes:

- Collecting and preprocessing a dataset of labeled spam and non-spam emails.
- Applying NLP techniques to clean and prepare the data for analysis.
- Implementing a machine learning model using the preprocessed data.
- Evaluating the performance of the model using metrics such as accuracy, precision, recall, and F1-score.
- Demonstrating the model’s capability by testing it on unseen email samples.
-

By the end of this project, we aim to have a lightweight, efficient, and accurate spam detection system that can be scaled for real-world application.

2 Problem Definition

The problem of spam email detection is fundamentally a binary classification task where each incoming email needs to be classified as either "Spam" or "Not Spam" (also called "Ham"). Spam emails not only waste user time but can also pose significant security risks including phishing, malware distribution, and financial fraud. Despite many filters being in place, spammers continuously evolve their tactics, making static rule-based systems obsolete over time.

The challenge lies in designing a system that can adapt to new types of spam by learning from data rather than relying on manually crafted rules. Moreover, such a system must also avoid incorrectly marking legitimate messages as spam, as this can result in missed opportunities and important information being lost.

The aim of this project is to develop a spam detection system that uses NLP for preprocessing and a machine learning algorithm for classification. The system should be able to:

- Accurately differentiate between spam and non-spam messages.
- Minimize the number of false positives (non-spam marked as spam).
- Minimize the number of false negatives (spam marked as non-spam).
- Adapt to new types of spam as they evolve.

The problem becomes more complex due to the variety of spam techniques, including:

- Obfuscation (using symbols to disguise words)
- Use of legitimate-sounding content to bypass filters
- Embedding malicious links in innocent-looking messages

Hence, a dynamic and learning-based solution that incorporates semantic understanding through NLP is crucial for effective spam detection.

3 Literature Survey

S.No	Title	Authors	Summary	Relevance to Project
1	Speech and Language Processing	Daniel Jurafsky, James H. Martin	Comprehensive book on NLP covering syntax, semantics, information retrieval, and machine learning techniques.	Provides foundational NLP knowledge required for preprocessing emails and converting text into model-friendly form.
2	Machine Learning	Tom M. Mitchell	Introduces key machine learning concepts including supervised learning, model evaluation, and real-world applications.	Helps understand logistic regression, feature extraction, and performance metrics for spam classification.
3	Foundations of Statistical Natural Language Processing	Christopher D. Manning, Hinrich Schütze	Covers statistical NLP methods including language modeling, classification, and text mining.	Offers statistical techniques applicable to text vectorization (like TF-IDF) and email content analysis.
4	Python Machine Learning	Sebastian Raschka, Vahid Mirjalili	A practical guide to building ML models using Python libraries such as scikit-learn, with real-world examples.	Assists in implementing spam detection pipeline including training, testing, and evaluating classifiers.

4 Methodology

The methodology followed in this project involves the systematic development of a spam detection system using NLP and Logistic Regression. It consists of the following stages:

4.1Data Collection:

We used a publicly available dataset of emails labeled as “spam” or “ham.” The dataset consists of thousands of email messages, providing a rich variety of spam types including phishing, promotional content, and scam messages.

4.2 Preprocessing:

Text preprocessing is essential in NLP to convert unstructured data into a structured format that a machine learning model can understand. The steps include:

- Lowercasing: Converting all text to lowercase to ensure uniformity.
- Removing punctuation and special characters: These often do not contribute to meaning.
- Tokenization: Breaking text into words or tokens.
- Stop-word removal: Eliminating common words like “is,” “the,” and “and” that don’t carry useful information.
- Stemming/Lemmatization: Reducing words to their root forms (e.g., “running” → “run”).

4.3 Vectorization:

We use the TF-IDF (Term Frequency-Inverse Document Frequency) technique to convert the text into numerical vectors. This method gives weight to important words in a message while diminishing the impact of commonly used words.

4.4 Model Training:

We use Logistic Regression as our classification model due to its simplicity and effectiveness for binary classification problems. The dataset is split into training and test sets. The model learns patterns from the training data and is evaluated on unseen test data.

4.5 Evaluation:

The model is evaluated using metrics such as:

- Accuracy: Proportion of correctly classified emails.
- Precision: How many predicted spam messages were actually spam.
- Recall: How many actual spam messages were correctly identified.
- F1 Score: Harmonic mean of precision and recall.

Through this structured approach, we ensure that our model is trained on clean, well-processed data and is capable of making accurate predictions.

5 Requirements

5.1 Functional Requirements

- The system must accept email text as input.
- It should preprocess the email text using NLP techniques.
- The system should classify the email as “Spam” or “Not Spam.”
- It should provide a probability score indicating prediction confidence.
- The model must be trained using a labeled dataset.
- Results of the classification should be displayed to the user

5.2 Non- Functional Requirementss

- **Accuracy**: The system should achieve high accuracy (ideally above 95%) on the test dataset.
- **Scalability**: It should be scalable to handle larger datasets and real-time inputs.
- **Usability**: The interface must be user-friendly and easy to interpret.
- **Efficiency**: The system must perform predictions quickly.
- **Maintainability**: The codebase should be modular to support updates to the model or preprocessing techniques.
- **Security**: User data should be handled securely, especially in real-world applications.

These requirements form the foundation for building a robust, efficient, and user-friendly spam detection system.

6 Results & Analysis

After training the Logistic Regression model on the preprocessed dataset, we evaluated its performance on a separate test set. The following results were obtained:

- **Accuracy**: 97.6%
- **Precision**: 96.9%
- **Recall**: 95.3%
- **F1 Score**: 96.1%

The confusion matrix revealed that the model made very few misclassifications. Most of the emails were correctly labeled as either spam or ham. Precision and recall values are also balanced, indicating that the model does not favor one class over the other.

We tested the model with custom emails like:

- “You’ve won a free iPhone!” → Spam
- “Your Amazon order has been shipped.” → Not Spam
- “Verify your account or it will be locked.” → Spam
- “Meeting is scheduled at 3 PM tomorrow.” → Not Spam

These outputs matched human expectations and validate the model’s generalization ability.
This analysis confirms that with proper preprocessing and model tuning, even a simple Logistic Regression model can perform effectively for spam classification.

7 Conclusion & Future work

This project successfully demonstrates the application of Natural Language Processing and machine learning for detecting spam emails. By combining preprocessing techniques like tokenization, stop-word removal, and TF-IDF vectorization with a Logistic Regression classifier, we built a robust and accurate spam detection system.

The model performs well across standard evaluation metrics and provides reliable predictions for unseen data. Its modular design allows easy integration into larger email systems or standalone spam filters.

Future Work:

- Integrating deep learning models like LSTM or BERT for improved semantic understanding.
- Incorporating real-time email streaming for continuous learning.
- Enhancing model adaptability using active learning.
- Deploying the system as a web service for public access.
- Expanding the dataset to include more recent spam formats and languages.

Spam detection remains a dynamic challenge, and ongoing enhancements in NLP and AI techniques will be crucial to staying ahead of spammers. This project lays a strong foundation for further development and real-world deployment.

8 References

- Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., & Spyropoulos, C. D. (2000). An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-mail Messages. Proceedings of the 23rd annual international ACM SIGIR conference.
- Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian Approach to Filtering Junk E-Mail. Learning for Text Categorization.
- Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011). Contributions to the Study of SMS Spam Filtering: New Collection and Results. Proceedings of the 11th ACM Symposium on Document Engineering.
- McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. AAAI-98 Workshop on Learning for Text Categorization.
- Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing (3rd ed.). Draft.
- Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level Convolutional Networks for Text Classification. arXiv preprint arXiv:1509.01626.
- Kibriya, A. M., Frank, E., Pfahringer, B., & Holmes, G. (2004). Multinomial Naive Bayes for Text Categorization Revisited. Australasian Joint Conference on Artificial Intelligence.
- Bhowmick, P. K., & Hazarika, S. M. (2006). Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends. IETE Technical Review.
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.

s