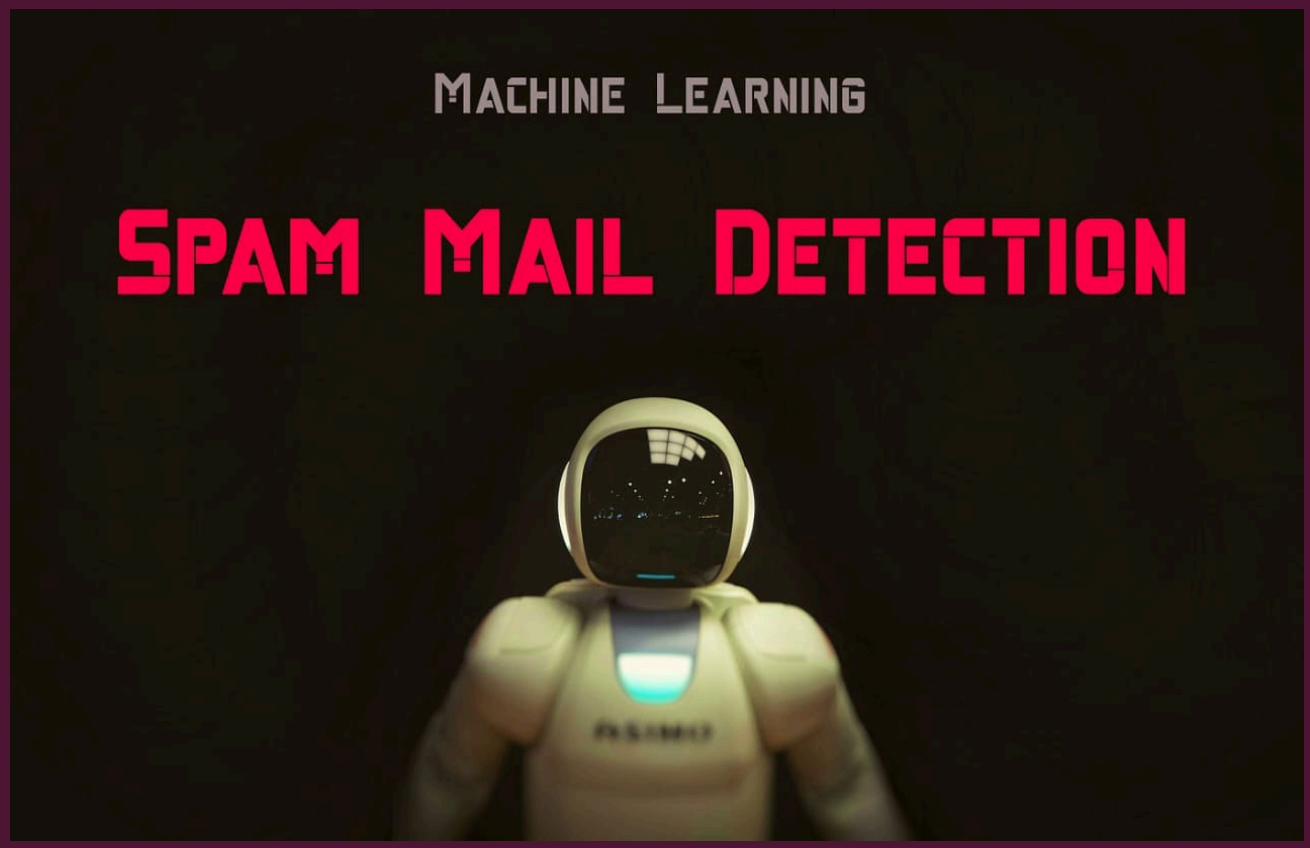
SPAM MAIL DETECTION USING MACHINE LEARNING

MINI PROJECT PRESENTATION

MACHINE LEARNING



Project done by

M Indu sekar
L Vamsi
S Mohammad Gaffar
D Sunil kumar
N Balaji

INTRODUCTION

- Spam emails are unsolicited and unwanted messages sent in bulk.
- They often contain advertisements, phishing links, or malware.
- Due to the huge volume of emails, manual filtering is not practical.
- Machine Learning provides an automated and efficient solution.

WHAT IS SPAM AND HAM?

- Spam Mail:
- - Unwanted or irrelevant emails
- - Often used for fraud or phishing

- Ham Mail:
- - Legitimate and useful emails
- - Sent by trusted sources

PROBLEM STATEMENT

- The goal is to classify incoming emails as Spam or Ham.
- Traditional rule-based filters fail to detect new spam patterns.
- An intelligent system is required that can learn from data.

OBJECTIVES

- - Automatically detect spam emails
- - Reduce human effort in email filtering
- - Improve accuracy and reliability
- - Enhance email security

DATASET DESCRIPTION

- The dataset contains a large number of labeled emails.
- Each email consists of text content and a label (Spam or Ham).
- The dataset is divided into training and testing sets.

DATA PREPROCESSING

- Text preprocessing prepares raw email text for analysis:
- - Convert text to lowercase
- - Remove punctuation and numbers
- - Remove stopwords
- - Tokenization
- - Stemming or Lemmatization

FEATURE EXTRACTION

- Feature extraction converts text into numerical form.
- Common techniques include:
- - Bag of Words (BoW)
- - TF-IDF (Term Frequency – Inverse Document Frequency)
- These features are used by ML models for classification.

MACHINE LEARNING ALGORITHMS

- Naive Bayes:
- - Based on probability theory
- - Works well for text classification

- Logistic Regression:
- - Binary classification algorithm

- Support Vector Machine (SVM):
- - Finds optimal decision boundary

MODEL TRAINING AND TESTING

- The dataset is split into training and testing data.
- The model learns patterns from training data.
- Testing data is used to evaluate performance.

EVALUATION METRICS

- Accuracy: Percentage of correctly classified emails
- Precision: Correctly predicted spam emails
- Recall: Ability to detect actual spam emails
- F1-Score: Balance between precision and recall

RESULTS

- The model achieves high accuracy in detecting spam emails.
- False spam detection is reduced.
- System performs efficiently on large datasets.

APPLICATIONS

- - Email services like Gmail and Outlook
- - Corporate email security systems
- - Banking and financial institutions
- - Spam filtering in messaging platforms

ADVANTAGES

- - Automated spam detection
- - Saves time and resources
- - Scalable to large email systems
- - Improves user experience

LIMITATIONS

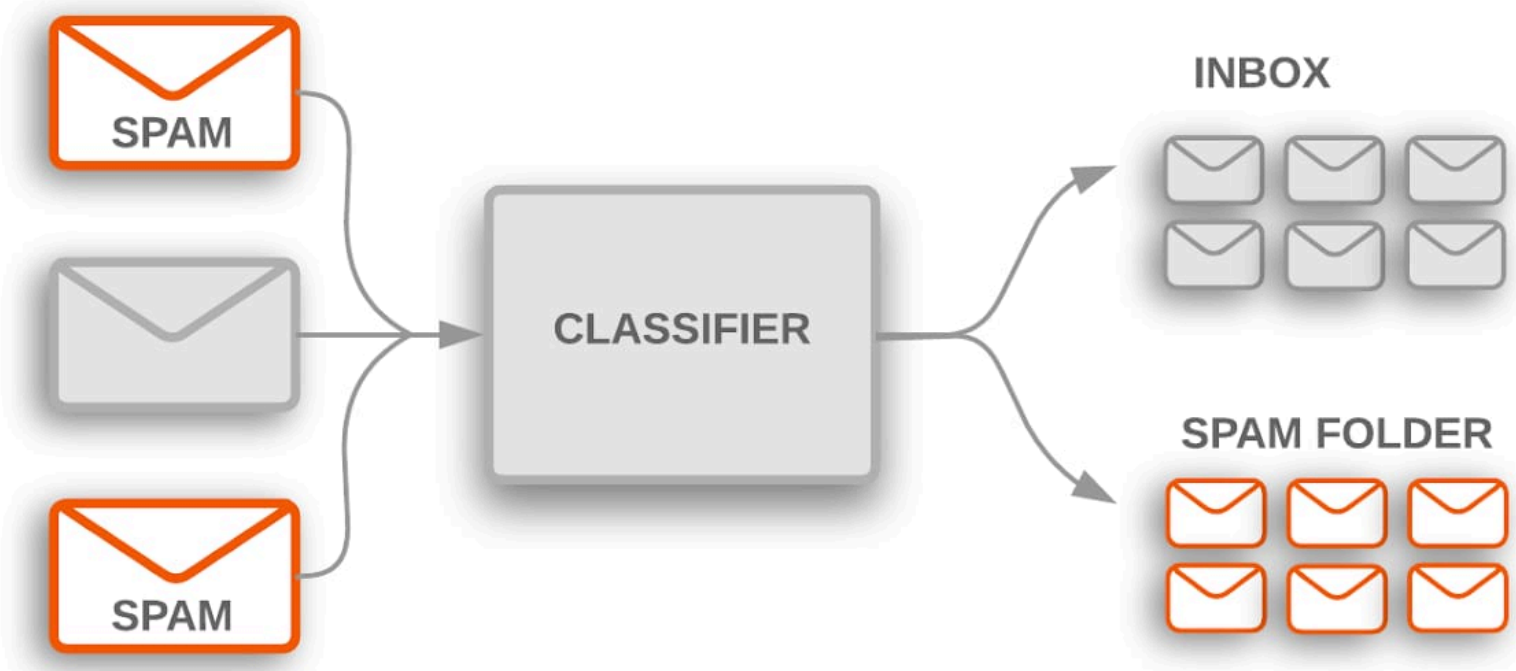
- - Requires large labeled datasets
- - New spam techniques may reduce accuracy
- - Language-dependent models

FUTURE SCOPE

- - Use of Deep Learning models
- - Real-time spam detection
- - Multilingual spam filtering
- - Improved phishing detection

CONCLUSION

- Spam Mail Detection using Machine Learning is an effective approach.
- It automates email filtering and improves security.
- Machine Learning models continuously improve with more data.



THANK YOU

- Thank you for your attention.
- Questions?