SpamGuard

Technical Architecture & System Design Report

An intelligent Machine Learning–based system for Email and SMS spam detection using Natural Language Processing techniques.

1. Introduction

SpamGuard is a robust and scalable spam filtering system designed to classify incoming messages as Spam or Ham (legitimate). The system leverages classical machine learning algorithms combined with efficient NLP preprocessing pipelines to achieve high precision and reliability.

2. System Architecture Overview

The architecture of SpamGuard follows a layered machine learning pipeline. Each layer is modular, ensuring maintainability, scalability, and ease of future enhancement.

Layer	Description
Data Ingestion	Loads and validates raw SMS/Email data
Preprocessing	Cleans, normalizes, tokenizes, and stems text
Feature Engineering	Transforms text into TF-IDF vectors
Model Layer	Multinomial Naive Bayes classifier
Evaluation	Measures accuracy, precision, recall, and F1-score
Prediction	Classifies new messages as Spam or Ham

3. Dataset Description

SpamGuard utilizes the SMS Spam Collection dataset, a widely used benchmark dataset for spam classification tasks in Natural Language Processing research.

Attribute	Details
Dataset Name	SMS Spam Collection (spam.csv)
Total Samples	5,157 messages (after cleaning)
Features	Message text, Category label
Labels	Spam (1), Ham (0)
Data Type	Textual and Categorical

4. Data Preprocessing Pipeline

The preprocessing stage ensures that raw textual data is transformed into a structured and noise-free format suitable for machine learning algorithms.

Key preprocessing steps include label encoding, duplicate removal, text normalization, tokenization, stop-word elimination, and stemming using the Porter Stemmer.

5. Feature Engineering

TF-IDF (Term Frequency–Inverse Document Frequency) is employed to convert text messages into numerical vectors. This technique effectively represents word importance while handling sparse, high-dimensional data.

6. Model Design

SpamGuard primarily uses the Multinomial Naive Bayes classifier due to its proven effectiveness in text classification problems. Logistic Regression and Support Vector Machines were evaluated for comparison.

7. Training & Performance Evaluation

The dataset is divided into training (80%) and testing (20%) sets. The model demonstrates strong performance, particularly in minimizing false positives.

Metric	**Value**
Accuracy	97.09%
Precision	100%
Recall	76.51%
F1-Score	0.8669

8. Limitations & Future Scope

While SpamGuard achieves high precision, challenges remain in handling slang, abbreviations, and contextual language. Future enhancements include n-gram modeling, deep learning approaches such as BERT, and real-time deployment through APIs.