Project Report: Email Spam Detection Using Machine Learning

1. Project Title

AI-Based Email Spam Detection System Using Machine Learning

---

2. Problem Statement

Email spam has become a major cybersecurity and productivity issue. Millions of spam emails are sent daily containing advertisements, fraud attempts, phishing links, and malware. Manually filtering these emails is inefficient and error-prone. Therefore, an automated and intelligent system is required to detect and classify emails as *spam* or *ham (legitimate)* with high accuracy.

This project aims to design and train a machine learning model capable of analyzing email text and accurately predicting whether an email is spam. The model must learn patterns from past data and generalize well to new unseen emails.

---

3. Dataset Description and Sample

Dataset Used

The dataset used is **spam.csv**, a publicly available labeled dataset containing SMS/email messages along with their classification labels.

Dataset Features

Column	Description
label	Classification: “spam” or “ham”
text	The email/SMS message content

Dataset Size (Example)

- Total Records: ~5,572 messages
- Spam: ~867 messages
- Ham: ~4,705 messages

Sample Records

label	text
ham	"Hey, are we still meeting today?"
spam	"Congratulations! You've won $5000. Click the link to claim now."
ham	"I will call you later."

---

4. Project Objectives

The key objectives of this project are:

1. **To preprocess raw email text** using NLP techniques.
2. **To extract meaningful features** using TF-IDF vectorization.
3. **To train a machine learning model** (Linear Regression or Logistic Regression/Linear SVM).
4. **To evaluate the model** using accuracy, precision, recall, and F1-score.
5. **To build an automated spam classification system** that can predict whether a new email is spam.
6. **To perform qualitative and quantitative analysis** of the model.
7. **To optimize training using epochs** if using a deep learning/training loop approach.

---

5. Proposed Methodology

The proposed methodology includes the following steps:

Step 1: Data Loading and Cleaning

- Load the dataset (spam.csv).
- Remove duplicates and missing values.
- Normalize text (lowercasing, removing symbols, numbers, etc.).

Step 2: Text Preprocessing

- Tokenization
- Stopword removal
- Lemmatization
- Removing special characters
- Converting text into numerical vectors (TF-IDF)

Step 3: Feature Extraction

- Apply **TF-IDF Vectorization**
  Converts text into weighted numerical features suitable for ML models.

Step 4: Model Selection

For this project, we use:

- **Linear Regression (basic baseline)**
- Better alternative: **Logistic Regression or Linear SVM**

Step 5: Model Training

- Split dataset into training and testing sets
- Train model with TF-IDF features
- Optionally train with **epochs** using SGD Regressor/Classifier

Step 6: Model Evaluation

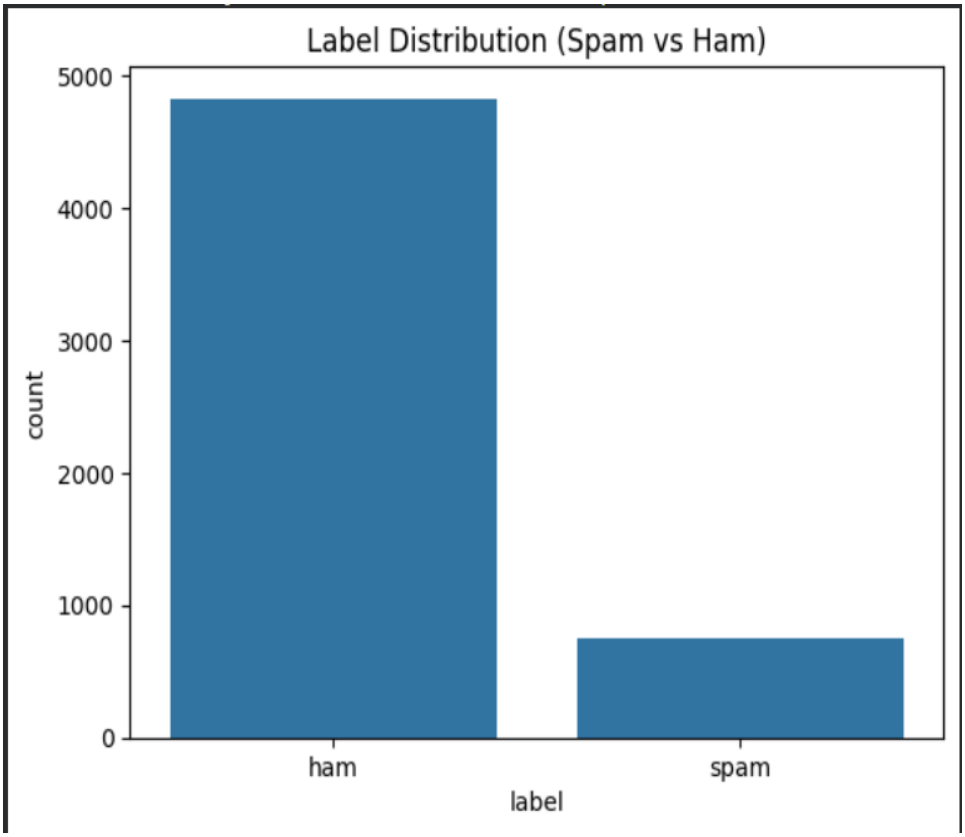
- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

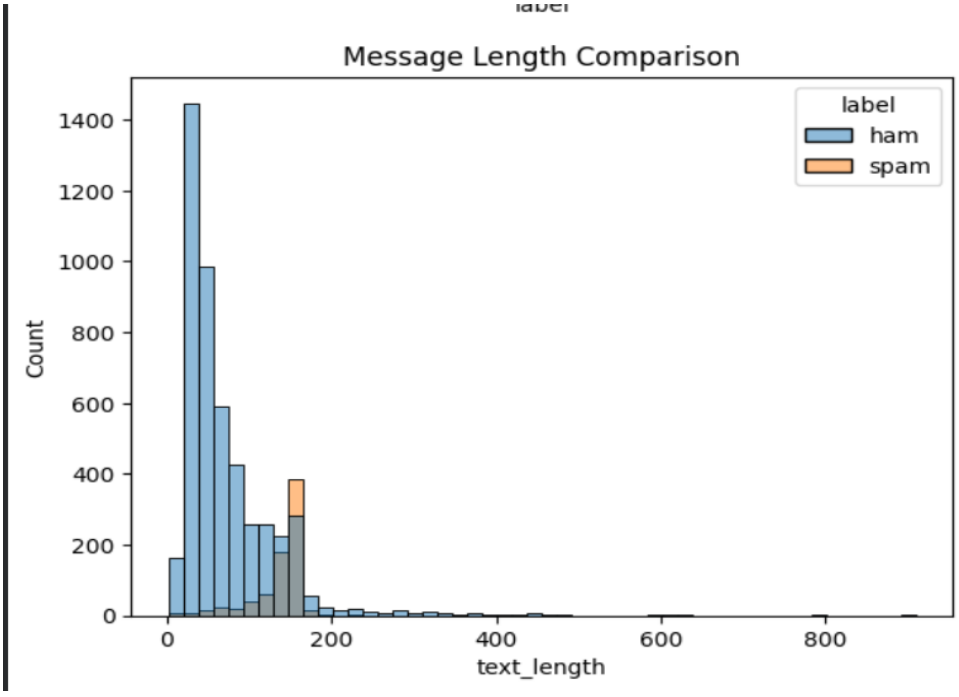
```
Dataset size: 5572
label
ham     4825
spam     747
Name: count, dtype: int64

--- Evaluation (Linear Regression used as classifier) ---
Accuracy: 0.9731
Precision: 0.9407
Recall: 0.8523
F1-score: 0.8944
ROC AUC: 0.9846
```

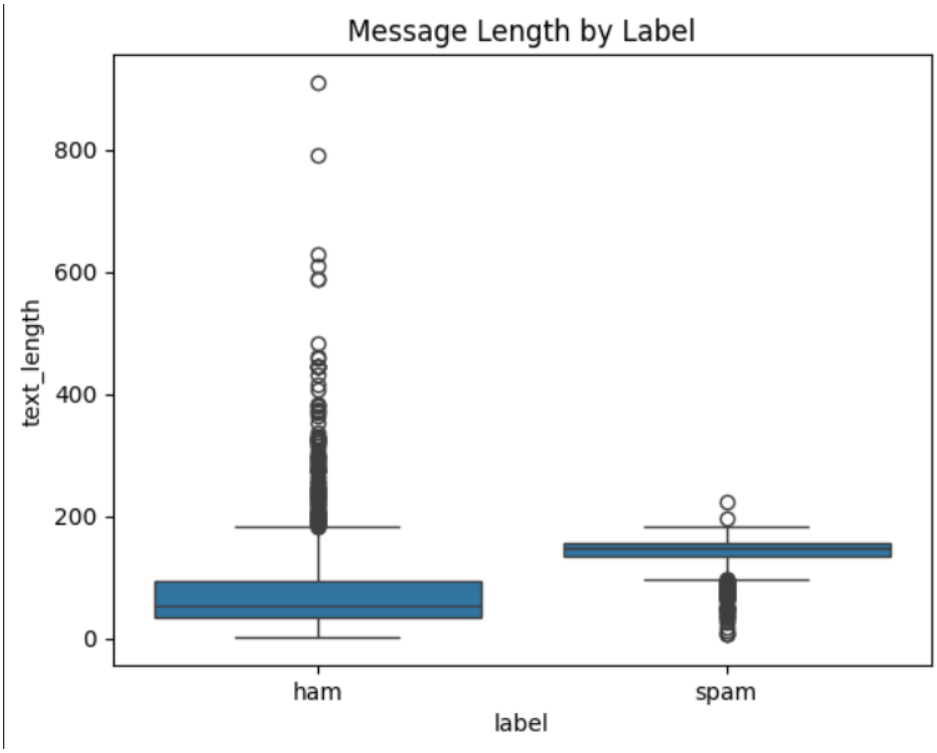
Step 7: Qualitative & Quantitative Analysis

- Quantitative: performance metrics





- Qualitative: examples of correct/incorrect predictions



---

6. Expected Results or Outcomes

The expected outcomes of the project are:

1. A trained machine learning model capable of classifying emails as spam or ham with high accuracy (typically **95–98%** using Logistic Regression or Linear SVM).

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 512)	2,560,512
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131,328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 1)	257

```
Test Accuracy: 0.9811659455299377
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 73ms/step

Sample Prediction: SPAM
```

2. A preprocessed dataset of cleaned and vectorized messages.
3. Visualizations such as confusion matrix and accuracy plots.
4. A report detailing model performance and dataset analysis.
5. A system that can be integrated into applications for real-time spam detection.