Email Spam Detector

Machine Learning Project

Presented by:

Ashik BC(232BDA31)

Introduction

- An email spam detector is a software or service designed to identify and filter out unwanted emails, commonly referred to as "spam" or "junk mail." The primary purpose of an email spam detector is to improve the quality and security of email communication for users by performing several important functions:

- 1. Reduce Inbox Clutter
- 2. Protect Against Phishing
- 3. Save Time
- 4. Enhance Security
- 5. Reduce Annoyance
- 6. Improve Productivity

Data Source

The dataset has taken from kaggle.com

Preview of the dataset(csv):-

v1	v2
ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
ham	Ok lar... Joking wif u oni...
spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
ham	U dun say so early hor... U c already then say...
ham	Nah I don't think he goes to usf, he lives around here though
spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, Â£1.50 to rcv
ham	Even my brother is not like to speak with me. They treat me like aids patent.
ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
spam	WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
ham	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.

Prerequisites

Basic Algebra:

- Familiarity with basic algebraic operations like addition, subtraction, multiplication, and division.

Statistics Fundamentals:

- Understanding of key statistical concepts like mean, variance, and standard deviation.
- Awareness of probability distributions and probability theory.

Data:

- Familiarity with data. Understanding of variables (independent and dependent) and data points.

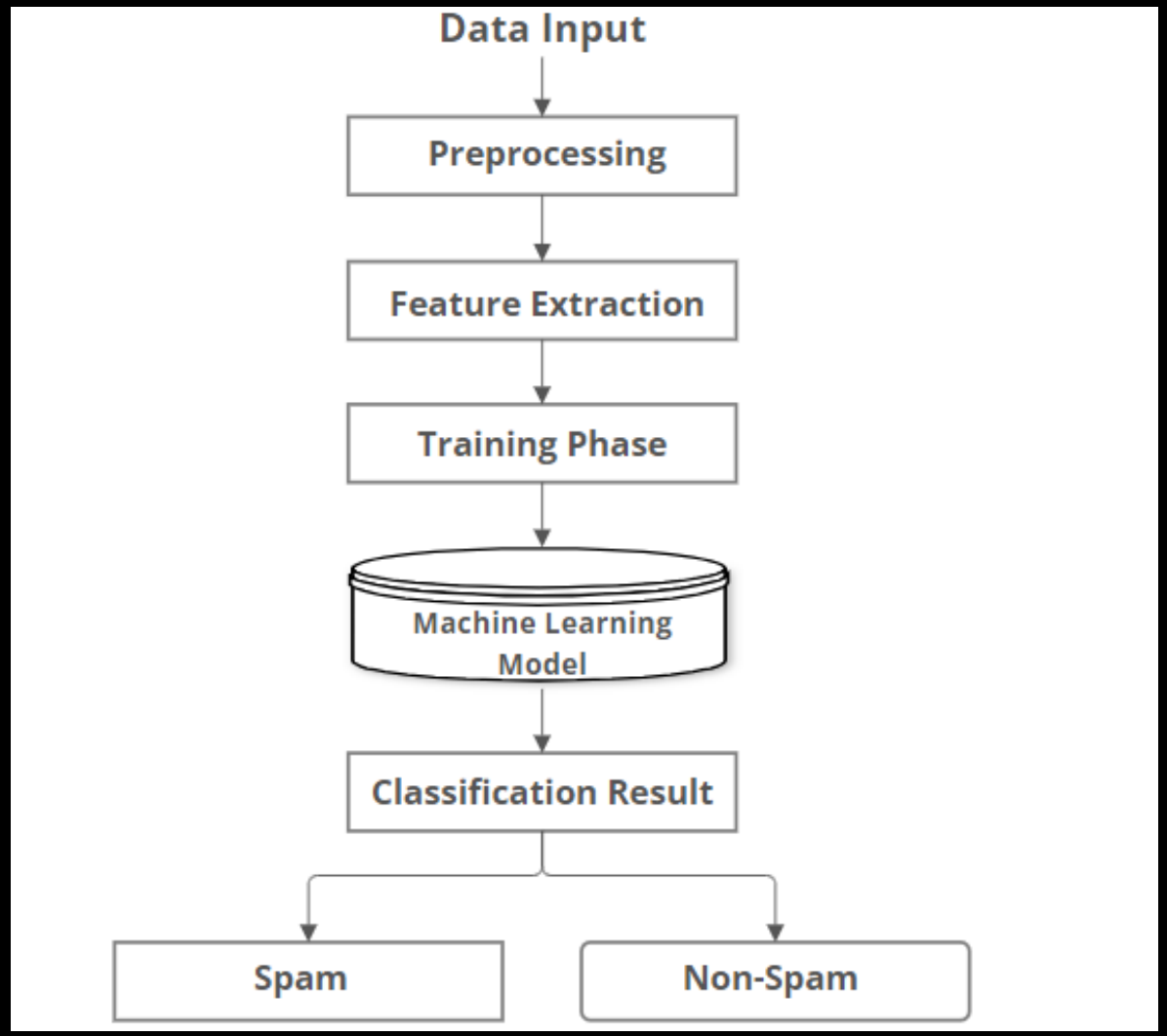
Terminology:

- Basic knowledge of terms such as dependent variable, independent variable, regression, and correlation.

Residuals:

- Understanding of what residuals are and how they relate to the differences between the actual data points and the predictions made by a model.

Architecture of Logistic Regression Model



Algorithms Used

Logistic Regression:-

Logistic regression aims to establish a relationship between one or more independent variables (predictors) and a binary dependent variable (target) by fitting a logistic curve that predicts the probability of the target variable being in one of the two categories. It is widely used for classification tasks where the goal is to predict categorical outcomes based on predictor variables.

SVM(Support Vector Machine):-

A Support Vector Machine (SVM) is a supervised machine learning model that is used for classification and regression tasks. SVMs are particularly popular for classification problems and are known for their ability to handle both linear and non-linear data separation effectively.

Working of the ML model

The model works by learning to classify emails as either spam or non-spam (ham) based on various features of the emails. Here's an overview of how it works:-

1. Data Preparation
2. Feature Extraction
3. Model Training
4. Prediction
5. Model Evaluation
6. Model Optimization

Evaluation metrics scores:

Logistic Regression:
Accuracy: 0.9802690582959641
Precision: 0.9922480620155039
Recall: 0.8590604026845637

Linear SVM:
Accuracy: 0.9820627802690582
Precision: 0.9849624060150376
Recall: 0.8791946308724832

Future Plans Biggest Learning

- The future plan is to build a application that will offer a user-friendly interface that seamlessly integrates with popular email platforms, providing a robust defence against unwanted and potentially harmful email content.
- The goal is to create a sophisticated and highly accurate spam detection system, leveraging advanced machine learning techniques and features like real-time email analysis, automatic updates, user customization, and feedback mechanisms to continually improve its performance.
- This application will empower users to enjoy a clutter-free inbox, reduce security risks, and save valuable time by efficiently managing their email communications.

Thank You