



Spam SMS Detection

The Spam SMS Detection project is a machine learning initiative focused on distinguishing between spam and non-spam (ham) messages. The project utilizes a logistic regression model trained on a set of messages, each labeled as either 'spam' or 'ham'. The model's performance is then evaluated based on its accuracy in categorizing new messages. The project employs a workflow of data preprocessing, training-testing data split, and prediction using the trained model. The model's prediction is then used to classify new messages as either 'spam' or 'ham'.

We can divide the mail into two types:-

1. **Spam Mails:** These are also known as unwanted emails, spam emails are designed to catch the recipient's attention or deceive them into action. Spam emails commonly include irrelevant, promotional, and often fraudulent content, such as offers for unknown products, shortened links, and all capped uppercase messages with the use of exclamation marks.
2. **Ham Mails:-** These are also known as non-spam emails, ham emails typically contain information or messages that the recipient expects or finds useful.

Examples of ham email content include personal messages, work-related information, subscription updates, and notifications.

Example of SPAM Mail:- Free entry in 2 a weekly comp to win FA Cup final tickets 21st May 2005. Test FA to 87121 to receive the entry question.

Example of HAM Mail:- Please go ahead with watts. I just wanted to be sure. Do have a great weekend.

WORKFLOW

- Mail Data
- Data Preprocessing
- Train Test Split
- Logistic Regression Model
- New mail
- Trained LRM
- Spam or Ham Mail Prediction

Data link:- <https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>



Code