



**Submitted to :**

**Submitted by:** Abhinav Shukla

**University Roll No.:-**2016561

**Guided by :**

Garima Sharma(Mentor)

**Session :** 2022

**CSE-lV Semester**

**Student ID :**20021535

**AIM**: Spam Classification using Machine learning

**INTRODUCTION**: Spam is a kind of unwanted data which are either sent by a bot or sometimes manually by a human.

Promotional messages or advertisements or the messages demanding our personal information can be considered as Spam messages.

The upsurge of such kind of unwanted messages or emails has created a need for development of Spam Classifiers to filter spams.

**MOTIVATION**: Most spam is irritating and consume lot of time but some spam can be very dangerous. Usually it is seen that scammers are trying to get your bank detail or some other personal details so that they can use is for some kind of crime.

Some spam email or messages might include phishing or malware that can harm your device.

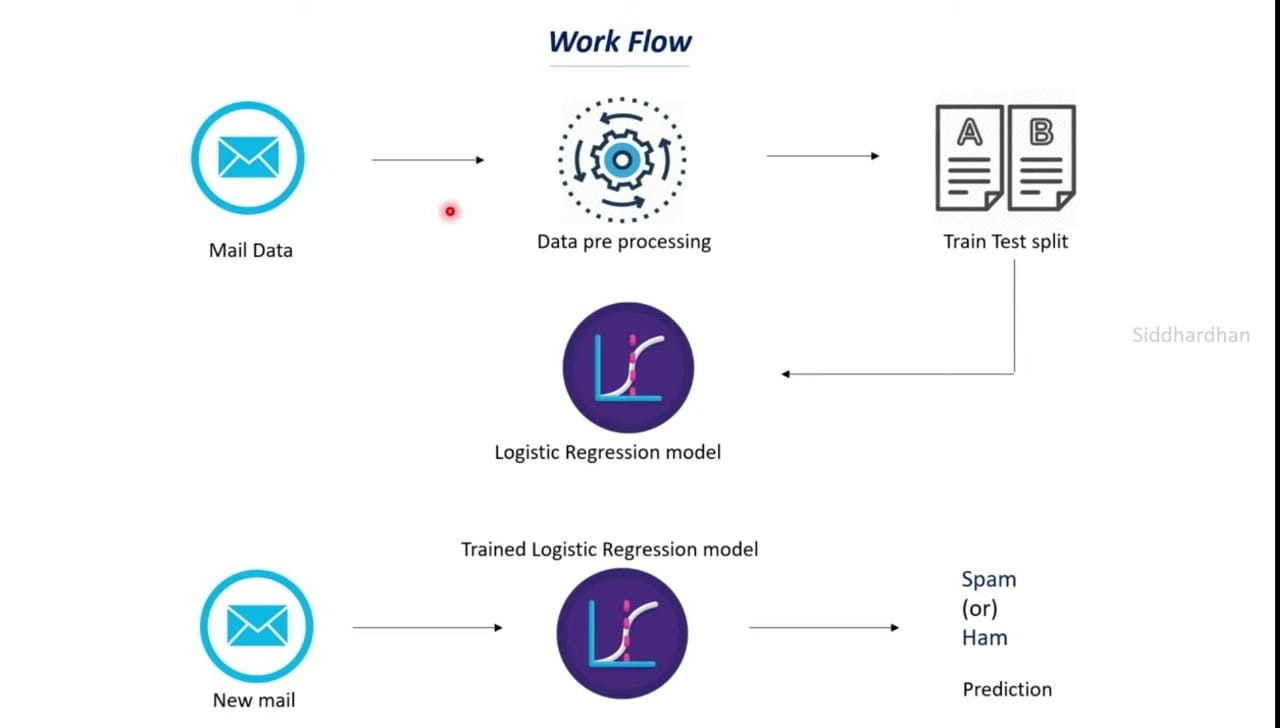
Seeing such examples, it is better to avoid spam emails or messages as much as possible. To avoid spam we can create a Spam Classifier which automatically works as a filter and tells us whether a message is a spam or not.

**TOOLS USED:**

1. Python 3.10.5
2. PyCharm(IDE)
3. Pandas 1.4.3
4. Scikit-learn 0.22

**METHODOLOGY:**





1. Import the required packages
2. Load the Dataset with Spam and Ham messages
3. Data cleaning and preprocessing
4. Test train Split
5. TfidfVectorizer used for feature extraction
6. Training done using logistic regression model

**References**

1. UCI Machine learning repository for Spam collection data set . <https://archive.ics.uci.edu/ml/datasets/sms+spam+collection>
2. Youtuber Code with harry for theoretical learning.

[https://www.youtube.com/watch?v=\_u- PaJCpwiU&list=PLu0W\_9lII9ai6fAMHp-acBmJONT7Y4BSG](https://www.youtube.com/watch?v=_u-%20%20PaJCpwiU&list=PLu0W_9lII9ai6fAMHp-acBmJONT7Y4BSG)

1. <https://github.com/krishnaik06/SpamClassifier>
2. <https://colab.research.google.com/drive/1VsDfAFLVclm7v2IOxpG32F6KdJN2RMCF?usp=sharing>