# Project: Spam or Ham message Classification

# This projects focuses to build a model that predicts whether or not a given text message is spam.

# Dataset:

**Context** The SMS Spam Collection is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,574 messages, tagged according being ham (legitimate) or spam.

**Content** The files contain one message per line. Each line is composed by two columns: v1 contains the label (ham or spam) and v2 contains the raw text.

# Technique applied:

# To classify the text messages as spam or ham I have used some basic natural language processing and then used Naïve Bayes classifier.

# Conclusion:

# I have got an accuracy of 97.9% on the test data and AUC is 0.977.