



## Practice Project: Building a spam classifier

### Introduction

**Spam detection** is one of the major applications of Machine Learning in the interwebs today. Pretty much all of the major email service providers have spam detection systems built in and automatically classify such mail as 'Junk Mail'.

**In this mission we will be using the Naive Bayes algorithm to create a model that can classify dataset SMS messages as spam or not spam, based on the training we give to the model.** It is important to have some level of intuition as to what a spammy text message might look like.

### What are spammy messages?

Usually they have words like 'free', 'win', 'winner', 'cash', 'prize', or similar words in them, as these texts are designed to catch your eye and tempt you to open them. Also, spam messages tend to have words written in all capitals and also tend to use a lot of exclamation marks. To the recipient, it is usually pretty straightforward to identify a spam text and our objective here is to train a model to do that for us!

**Being able to identify spam messages is a binary classification problem as messages are classified as either 'Spam' or 'Not Spam' and nothing else. Also, this is a supervised learning problem, as we know what are trying to predict. We will be feeding a labelled dataset into the model, that it can learn from, to make future predictions.**

[NEXT](#)