SPAM SMS - CLASSIFICATION AND TOPIC MODELLING

1. SMS Classification

Reference:

https://machinelearningmastery.com/deep-learning-for-nlp/ (https://machinelearningmastery.com/deep-learning-for-nlp/)

https://machinelearningmastery.com/clean-text-machine-learning-python/?fbclid=IwAR1Zu4lFlGkdW7bdEZz01sLEo2VNj-7yhKN92ZLOqlTfeBQWC8XSiNkCpks (https://machinelearningmastery.com/clean-text-machine-learning-python/?fbclid=IwAR1Zu4lFlGkdW7bdEZz01sLEo2VNj-7yhKN92ZLOqlTfeBQWC8XSiNkCpks)

https://machinelearningmastery.com/develop-n-gram-multichannel-convolutional-neural-network-sentiment-analysis/?fbclid=IwAR2KthsKRrP9eZ_ezZQEse_5bwWJ8pH1r9DvXtv5UtCTRpvanuE6BbKtehU (https://machinelearningmastery.com/develop-n-gram-multichannel-convolutional-neural-network-sentiment-analysis/?fbclid=IwAR2KthsKRrP9eZ_ezZQEse_5bwWJ8pH1r9DvXtv5UtCTRpvanuE6BbKtehU)

About Dataset:

"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 acording being ham (legitimate) or spam".

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.

https://www.kaggle.com/uciml/sms-spam-collection-dataset (https://www.kaggle.com/uciml/sms-spam-collection-dataset)

Library and Data

In [1]:
```
import tensorflow as tf
```