

Assignment 07

Natural Language Processing Bag of Words

The spam.csv file contains SMS messages from different users.

The file contains one message per line. Each line contains two columns: v1 contains the label

(ham or spam) and v2 contains the raw text.

Use Bag-of-words model and random forest algorithm to design a classifier that can distinguish spam and ham messages.

Please note the following:

Dataset is added along with

Do the assignment in Jupyter notebook/colab

Add the .ipynb file to GitHub as public and share the link in Paatshala

