|  |
| --- |
| **Task: SMS Spam Filtering**  **Dataset:**  The dataset is a collection of instances of SMS labeled messages as Ham (legitimate) or Spam for spam filtering.The collection is composed of text file, where each line has the correct class followed by the raw message. We offer some examples below:  ham What you doing?how are you?  ham Ok lar... Joking wif u oni...  spam FreeMsg: Txt: CALL to No: 86888 & claim your reward of 3 hours talk time to use from your phone now! ubscribe6GBP/ mnth inc 3hrs 16 stop?txtStop  spam Sunshine Quiz! Win a super Sony DVD recorder if you canname the capital of Australia? Text MQUIZ to 82277. B  spam URGENT! Your Mobile No 07808726822 was awarded a L2,000 Bonus Caller Prize on 02/09/03! This is our 2nd attempt to contact YOU! Call 0871-872-9758 BOX95QU  Note: the messages are not chronologically sorted.  The zip files contains 2 text files:   * train.txt ( 5000 instances) * test.txt ( 525 instances)   Kindly use train.txt for building your classifier model that will classify the incoming text message as a ham or spam. Use test.txt to test the accuracy of your approach classifying the messages in test.txt. Send us your commented code with proper directions for the evaluation along with some explanation of your approach. |
|  |