

Text Classification

Your task is to train and evaluate a text classifier using the following dataset:

<http://archive.ics.uci.edu/ml/datasets/YouTube+Spam+Collection>.

The task is to classify comments as spam, or not spam (ham). This problem must be solved using Python, but you are free to use any public libraries you wish for data processing, model training, and evaluation. Please adhere to the following constraints:

1. Code must be in a single, executable Python script.
2. Dependencies must be provided in a requirements.txt file
3. You may not use any other outside data or pre-trained models (aside from word vectors or internal parsing/tagging models used by a library such as spaCy)
4. Your code must consider at least two candidate models and automate the process of picking the best model.
5. Decide how to split the data into training, validation, and test sets, and print appropriate evaluation metrics at the end of your script's execution.