

```
# import packages
import sys

def load_data(data_file):
    # read in file

    # clean data

    # load to database

    # define features and label arrays

    return X, y

def build_model():
    # text processing and model pipeline

    # define parameters for GridSearchCV

    # create gridsearch object and return as final model pipeline

    return model_pipeline

def train(X, y, model):
    # train test split

    # fit model

    # output model test results

    return model

def export_model(model):
    # Export model as a pickle file

def run_pipeline(data_file):
    X, y = load_data(data_file) # run ETL pipeline
    model = build_model() # build model pipeline
    model = train(X, y, model) # train model pipeline
    export_model(model) # save model

if __name__ == '__main__':
    data_file = sys.argv[1] # get filename of dataset
    run_pipeline(data_file) # run data pipeline
```