Homework 4: NLP Classification
Due date: Oct 8th, Tuesday, 11:59 PM

1. **Preprocessing the data:**
   a. Import the libraries and datasets
   b. Use 're' library to extract the characters [a-zA-Z]
   c. Use 'nltk' library to extract to stem the sentence and remove the stop words.
   d. Leverage countvectorizer to vectorize the data.
   e. Split the train and test set.

2. **Build the models:**
   a. Build Logistic Regression Classifier, Naive Bayes, Decision Tree Classifier, Random Forest Classifier.
   b. Use Grid Search CV to find the best parameters of each model, and train the models with the best parameters.
   c. Predict with the test set

3. **Evaluation:**
   a. Calculate the classification_report and confusion matrix of each model.
   b. Write a brief report evaluating those models.

You can put all your code into one script. Please submit both IPYNB and PDF files.
Read the tsv file with: pd.read_csv(r'Restaurant_Reviews.tsv', sep='\t')