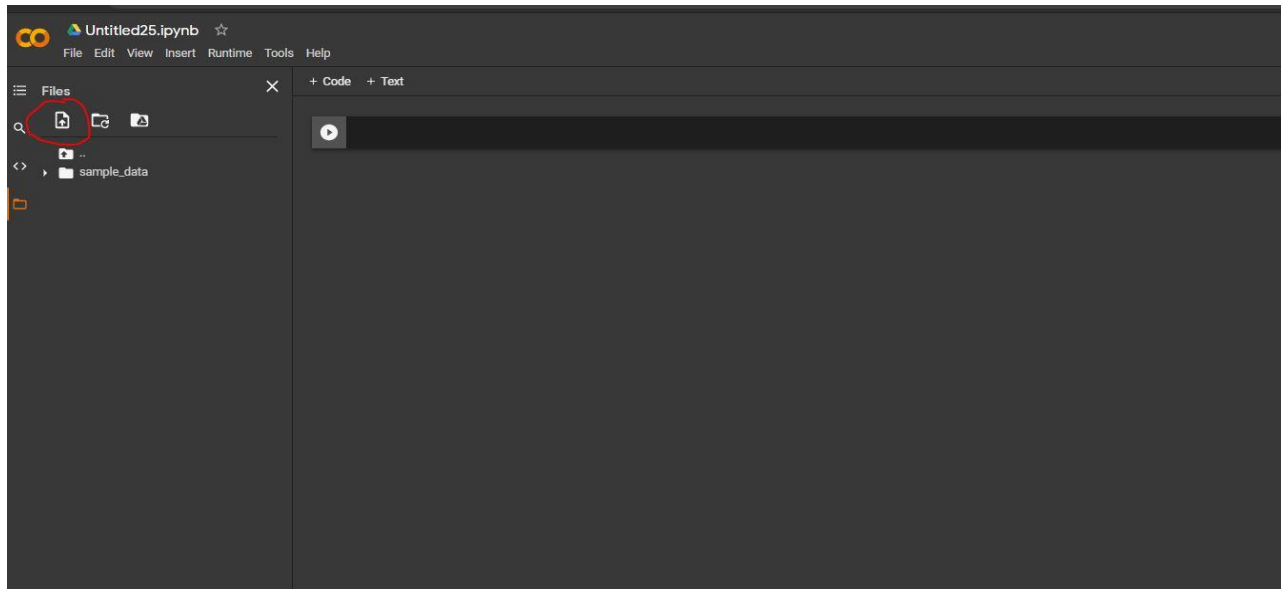
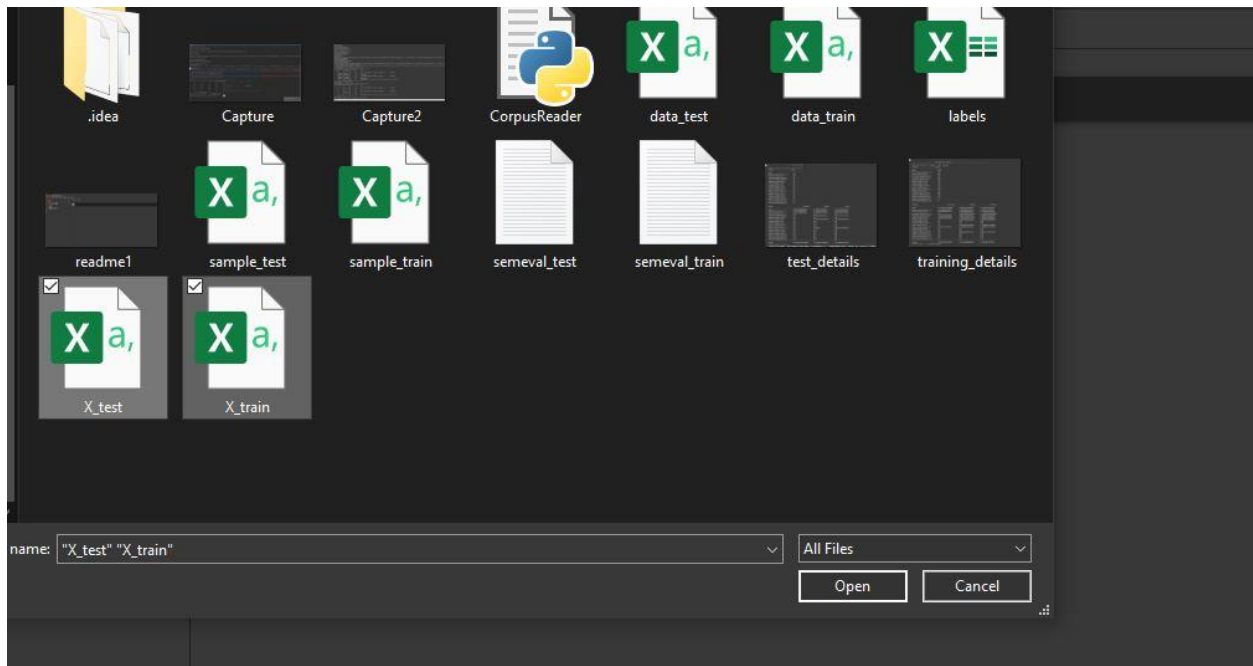


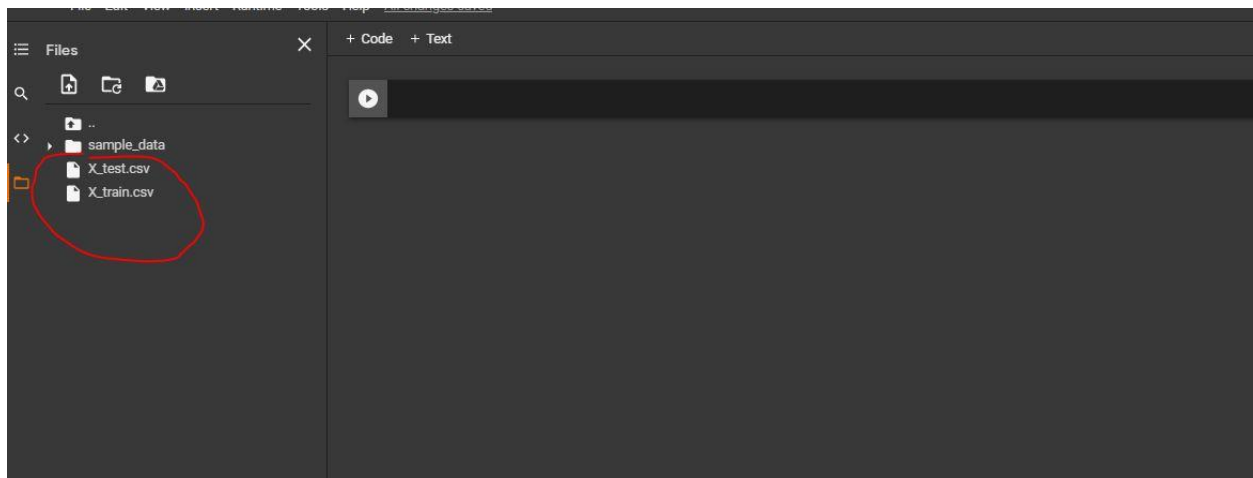
1. Open new notebook in google collab
2. Upload the training and testing dataset attached in the collab
 - a) Click on the upload file icon



- b) Select the test and training csv file



- c) Dataset should be uploaded outside sample_data folder



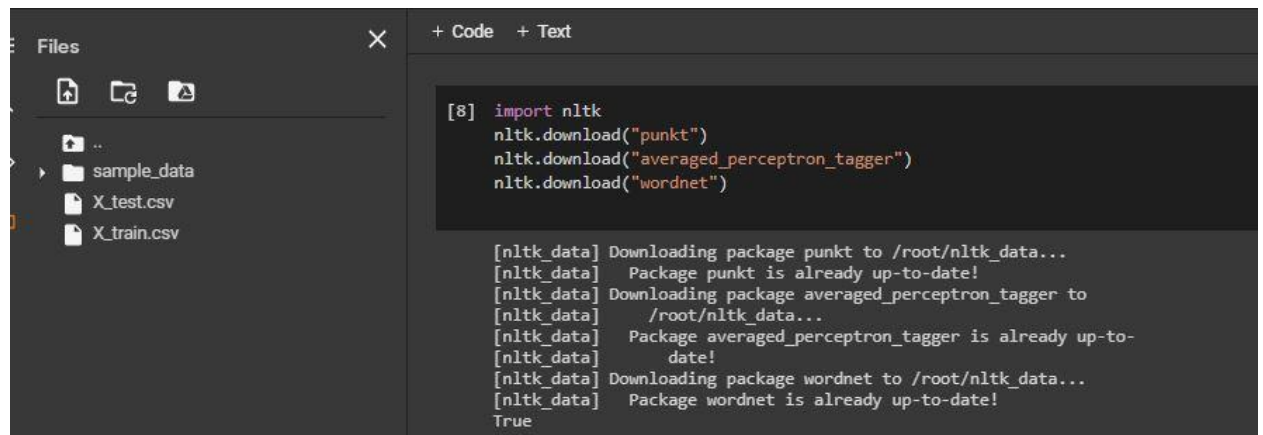
3. Run following commands in collab to download the dependencies

Import nltk

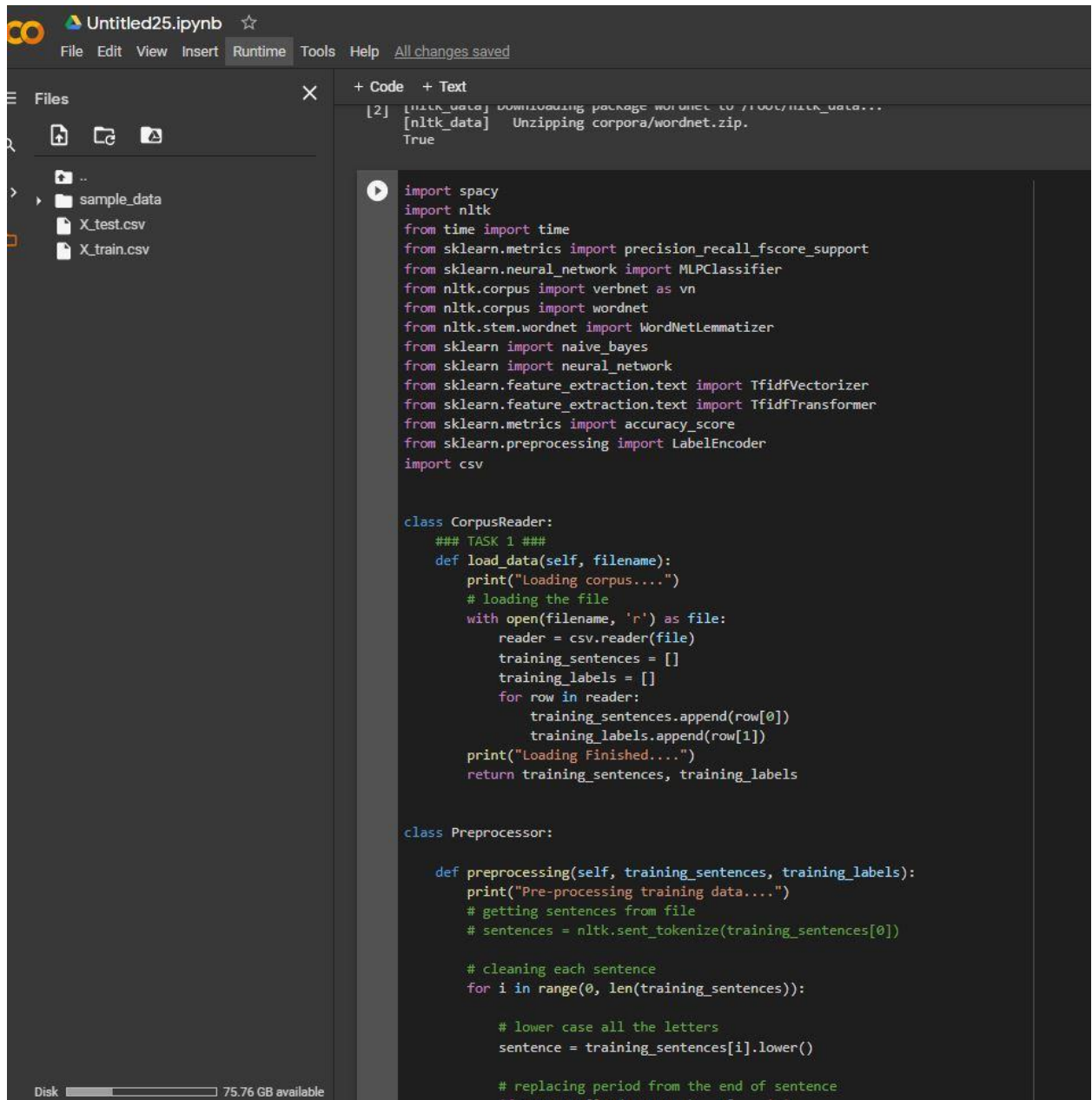
```
nltk.download("punkt")
```

```
nltk.download("averaged_perceptron_tagger")
```

```
nltk.download("wordnet")
```



- Copy the attached code in the collab



The screenshot shows a Jupyter Notebook titled 'Untitled25.ipynb'. The left sidebar displays a file explorer with a folder named 'sample_data' containing two files: 'X_test.csv' and 'X_train.csv'. The main area shows a code cell with the following Python code:

```
[2] [nlTK_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Unzipping corpora/wordnet.zip.
True

import spacy
import nltk
from time import time
from sklearn.metrics import precision_recall_fscore_support
from sklearn.neural_network import MLPClassifier
from nltk.corpus import verbnet as vn
from nltk.corpus import wordnet
from nltk.stem.wordnet import WordNetLemmatizer
from sklearn import naive_bayes
from sklearn import neural_network
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import LabelEncoder
import csv

class CorpusReader:
    """ TASK 1 """
    def load_data(self, filename):
        print("Loading corpus....")
        # loading the file
        with open(filename, 'r') as file:
            reader = csv.reader(file)
            training_sentences = []
            training_labels = []
            for row in reader:
                training_sentences.append(row[0])
                training_labels.append(row[1])
        print("Loading Finished....")
        return training_sentences, training_labels

class Preprocessor:

    def preprocessing(self, training_sentences, training_labels):
        print("Pre-processing training data....")
        # getting sentences from file
        # sentences = nltk.sent_tokenize(training_sentences[0])

        # cleaning each sentence
        for i in range(0, len(training_sentences)):

            # lower case all the letters
            sentence = training_sentences[i].lower()

            # replacing period from the end of sentence
            if sentence[-1] == '.':
```

- Run the collab
 - It takes approximately 74 mins to run both the training and test data