Required Packages

Following are the packages required for running the code. To install the packages run 'pip install "package-name" '

- import math
- import en_core_web_sm
- import numpy as np
- import pandas as pd
- import re, string, unicodedata
- from nltk import word_tokenize, sent_tokenize
- from nltk.corpus import stopwords
- from nltk.stem import LancasterStemmer, WordNetLemmatizer
- import inflect
- import textstat
- from gensim.test.utils import common_texts,get_tmpfile
- from gensim.models import Word2Vec
- import gensim
- from nltk.tag import RegexpTagger
- from collections import OrderedDict
- from sklearn import preprocessing
- from sklearn.model_selection import train_test_split
- from sklearn.naive_bayes import GaussianNB
- from sklearn.metrics import f1_score
- from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
- import spacy
- from collections import Counter
- from nltk.stem import PorterStemmer
- import numpy as np
- import os
- from bs4 import BeautifulSoup
- import re, string, unicodedata
- from sklearn.ensemble import RandomForestClassifier
- from sklearn.tree import DecisionTreeClassifier
- from sklearn import svm

NOTE:
Please provide the path as follows:

path_to_csv : Add the master996 file
path_to_books: Path to the folder where the books are placed.

```
path_to_csv =  "D:\\Masters\\Semester II\\5 Advanced Topics in Machine Learning\\Project\\Data\\Gutenberg_English_Fiction_1k
path_to_books = "D:\\Masters\\Semester II\\5 Advanced Topics in Machine Learning\\Project\\Data\\Gutenberg_English_Fiction_1
data = load_data(path_to_csv,path_to_books)
data.to_csv('Loaded_data.csv')
data
```

As pre-processing takes awhile for completion, the preprocessed file (ATiML_Preprocessed_data.csv) can be used. The file can be found in the link below.

https://drive.google.com/file/d/1qkKzi7niGRmmJd99e9I92c0mrdseUxGQ/view?usp=sharing

```
#data = pd.read_csv("C:\\Users\\calid\\Desktop\\ATiML_Preprocessed_data.csv") Calling Preprocessed data file
```

Similarly for the features, a file containing all the extracted features can be uploaded (AtiML_Features.csv)

```
#all_features = pd.read_csv('C:\\Users\\calid\\Downloads\\ATiML_Features.csv') Calling extracted feature file
```