Assignment 1 - Gender Detection

Build a classification model to classify the gender of a Hollywood movie character based on their dialogues in the movie.

Here’s the link to the dataset and its detailed description:

<https://www.kaggle.com/Cornell-University/movie-dialog-corpus>

You need to solve this project in a jupyter notebook and submit the IPYNB file or the link to your Google Colaboratory notebook (in case you are using Google Colab). **You need to explain in detail:**

* the data-exploration steps including data cleaning, data normalisation/standardisation, feature encoding, and data visualisation
* all the classification model(s) used by you and the Python package(s) used to build the same
* every line of your code

For eg., *“In the above code cell, in line number 2, using the* **pandas.read\_csv()***function,*

*we created a new Pandas DataFrame”.*

You are free to use any classification model to solve this project. Also, while writing your line-by-line code explanations, assume that you are explaining your working to a CBSE board 10th-grade student who has been introduced only to the fundamentals of Python and descriptive statistics.

Your solution will be assessed on the following parameters:

* How well you have introduced different topics/concepts relating to this particular use-case
* How well you have explained your code and your working
* The approach used to solve the project
* The model evaluation results obtained such as recall, precision, and f1-score
* How well you have interpreted the results/outputs

**Statutory warning:** Do not copy anybody’s code from the internet. Write your code. If found guilty of plagiarism, your candidature will be rejected immediately.

**Deadline:** Submit your solution ***within two days*** from the date of receipt of this assignment.