Assignment 1: Titanic Disaster Survivability Prediction using ML

Project Location: https://www.kaggle.com/c/titanic

How to Submit:

a. Submit both in Google classroom and Kaggle.

b. For Kaggle submission, create an account, then follow the instruction below:
https://www.kaggle.com/jlawman/complete-beginner-your-first-titanic-submission

c. For classroom submission, use the Assignment 1 link in the classroom.

d. Please include the screenshot of your kaggle submission score in your google classroom submission.

Tasks:

Use the following steps to submit your assignment.

For each step, you need to explain what are the codes below, what they achieve and why you chose certain options/preprocessing steps/model selection/data processing etc. and then include code for that part. If a certain step is not necessary for your case, please also mention why you omitted that step.

STEPS

1. Describe the data used and how you got the data.

2. Initial Exploration of the data (print data, visualize data). Find number of attributes, number of records in the data. Also find if there is any attribute which has missing data.

3. Find correlation of data attributes and target variable (whether survived or not) to explore which attributes have most effects on output (target variable).

4. Compute any new attribute/feature from existing attributes [if necessary]. Explain how these new features contribute to the prediction of target variable.

5. Select your classification model (try at least three model to find best one)

6. Calculate confusion matrix and show classification report using Scikit-learn library

NOTE: In Kaggle you just have to submit an output file, please organize that file according to instruction in Kaggle. In google classroom, you have to submit the .ipynb file and the screenshot of kaggle submission.

a) You can get the .ipynb file from your jupyter notebook easily as shown below:
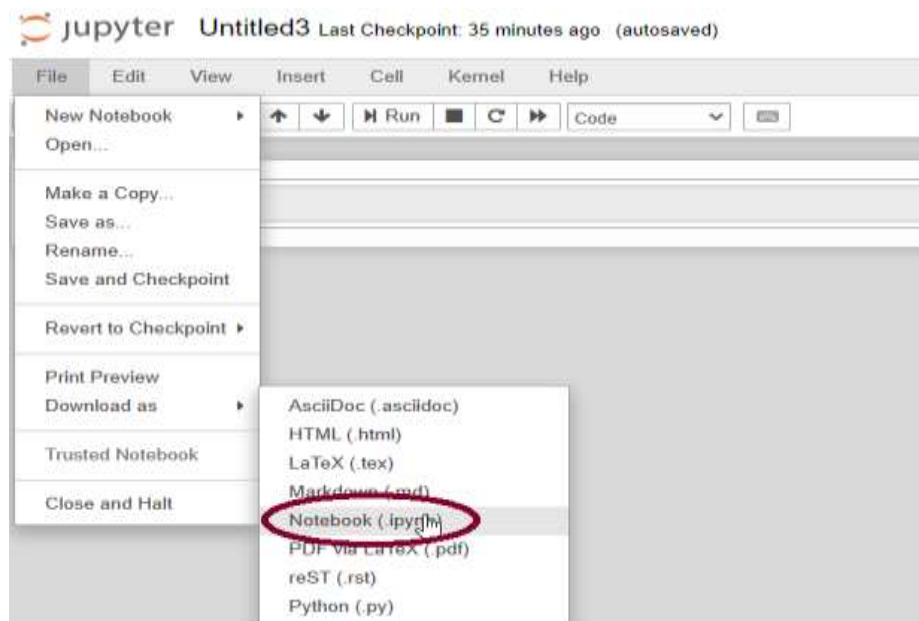
b) You can get the .ipynb file from google collab easily as shown below: