

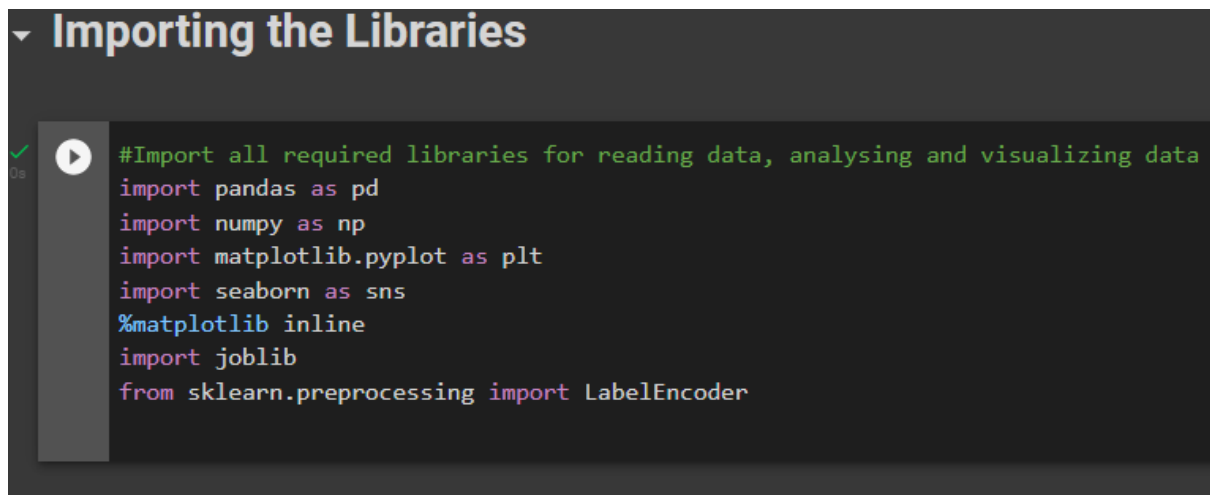
DATA PRE-PROCESSING

IMPORT THE LIBRARIES

Date	19 November 2022
Team ID	PNT2022TMID04221
Project Name	Project – Statistical Machine Learning Approaches to Liver Disease Prediction

The first step is usually importing the libraries that will be needed in the program.

The required libraries to be imported to Python script are:

A screenshot of a Jupyter Notebook interface. The title bar at the top reads 'Importing the Libraries'. Below the title bar, there is a code cell with a play button icon on the left. The code cell contains the following Python code:

```
#Import all required libraries for reading data, analysing and visualizing data
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import joblib
from sklearn.preprocessing import LabelEncoder
```

- **Numpy:** It is an open-source numerical Python library. It contains a multi-dimensional array and matrix data structures. It can be used to perform mathematical operations on arrays such as trigonometric, statistical, and algebraic routines.
- **Pandas:** It is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.
- **Matplotlib:** Visualisation with python. It is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- **Seaborn:** Seaborn is a library for making statistical graphics in Python. Seaborn helps you explore and understand your data. Its plotting functions operate on dataFrames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.
- **Joblib:** The joblib module implements serialization protocol, which provides an ability to save and later load Python objects using special binary format.