Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	13 October 2022
Team ID	PNT2022TMID15966
Project Name	Project – Stastiscal Machine Learning Approaches to Liver Disease Prediction
Maximum Marks	4 Marks

ARCHITECTURE:

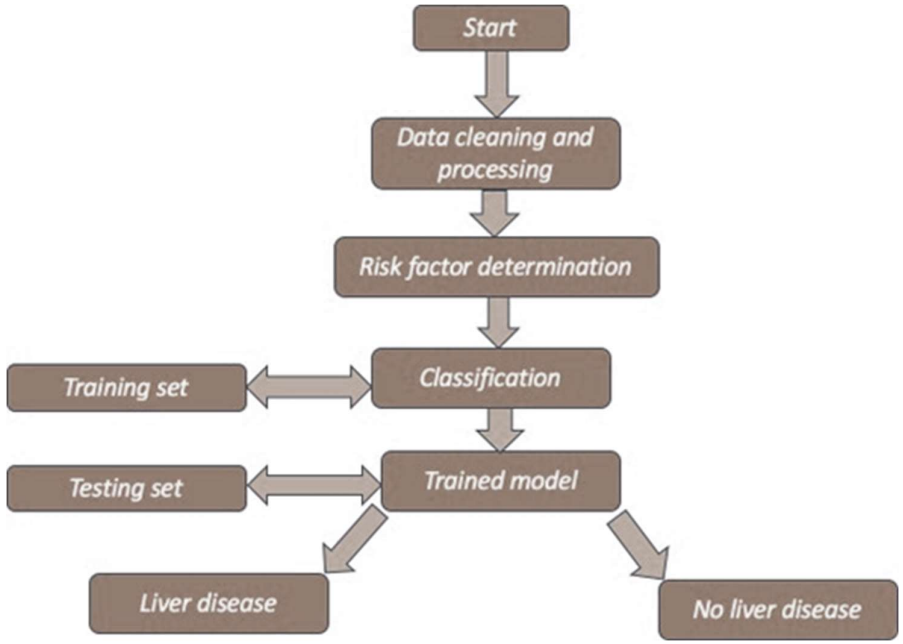

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	Importing data	Data Import lets you upload data from external sources and combine it with data you collect via Analytics	Python, NumPy, pandas.
2.	Data Cleaning	Data cleaning is a process by which inaccurate, poorly formatted, or otherwise messy data isorganized and corrected	Python
3.	Data Preprocessing	Data preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure	Python
4.	Training data	Training data is the subset of original data that is used to train the machine learning model,	Python
5.	Testing data	Test data is data which has been specifically identified for use in tests, typically of a computer program.	Python
6.	Machine learning model	A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data	Python
7.	Improve model performance	Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right.	Python
8.	Checking accuracy	A data accuracy check, sometimes called a data sanity check, is a set of quality validations that take place before using data.	Python

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Collection of data	Data collection is the process of gathering, measuring, and analyzing accurate data from avariety of relevant sources to find answers to research problems, answer questions, evaluate outcomes, and forecast trends and probabilities.	Python, NumPy, pandas
2.	EDA Analysis	Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summaryand graphical representations.	Python
3.	Train & Test split of data	The train-test split is used to estimate the performance of machine learning algorithms that are applicable for prediction-based Algorithms/Applications. This method is a fast and easy procedure to perform such that we can compare our own machine learning model results to machine results.	Python
4.	Model prediction	Predictive modeling is a commonly used statistical technique to predict future behavior.	Python