***Capstone Project - Tanisha***

1. **Problem Statement**

My problem statement is that many people die of heart attacks every year. When doctors know the chances of heart attack and they can prevent it through treatments. My capstone project predicts the chances of a person getting heart attacks based on the other variable like age or cholesterol level that affect the chances of getting heart attack.

1. **Explain about Dataset**

For my Capstone project I have taken a dataset from Kaggle that contains different factors that affect the chances of getting a heart attack. It contains 14 columns including age, chest pain type etc. I inferred from the data set that I would need a classification model as I need to predict whether or not a person would get a heart attack.

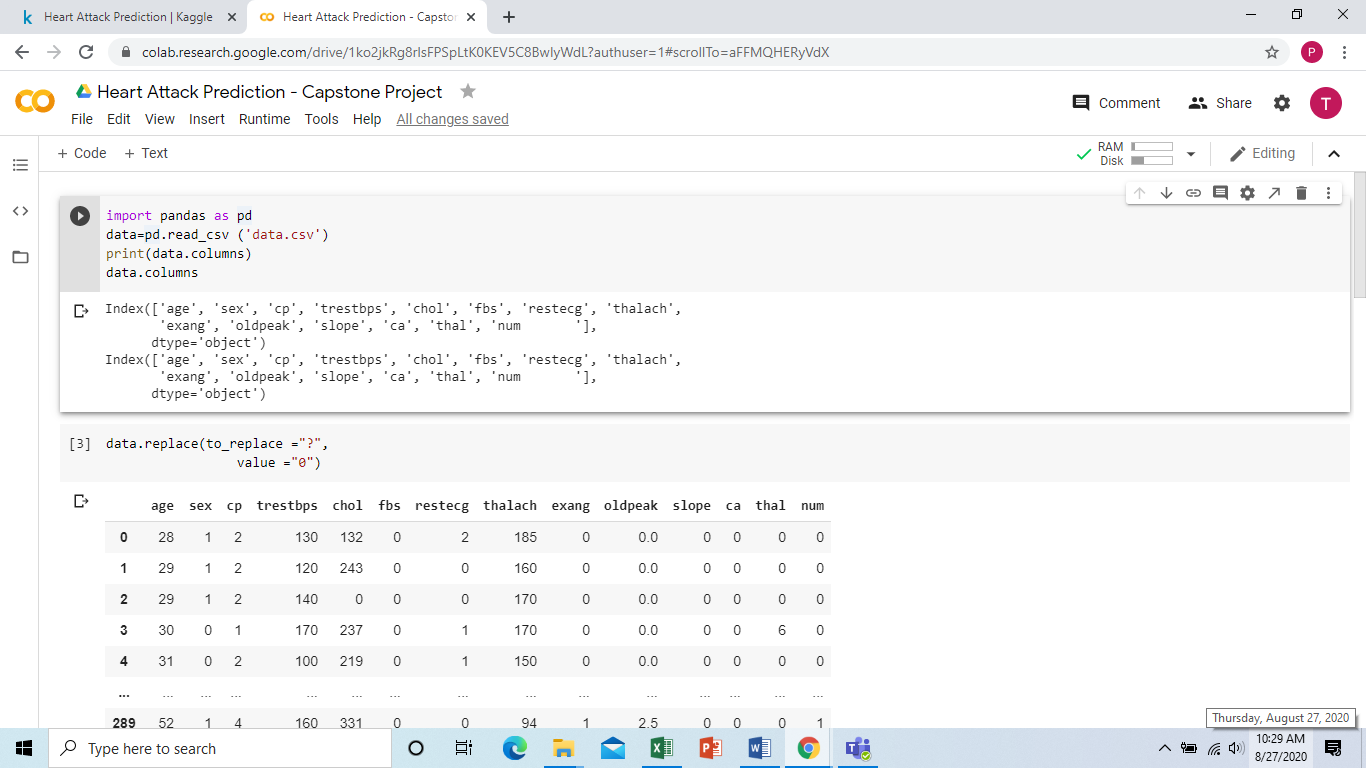
1. **Why you need AI to solve this problem**

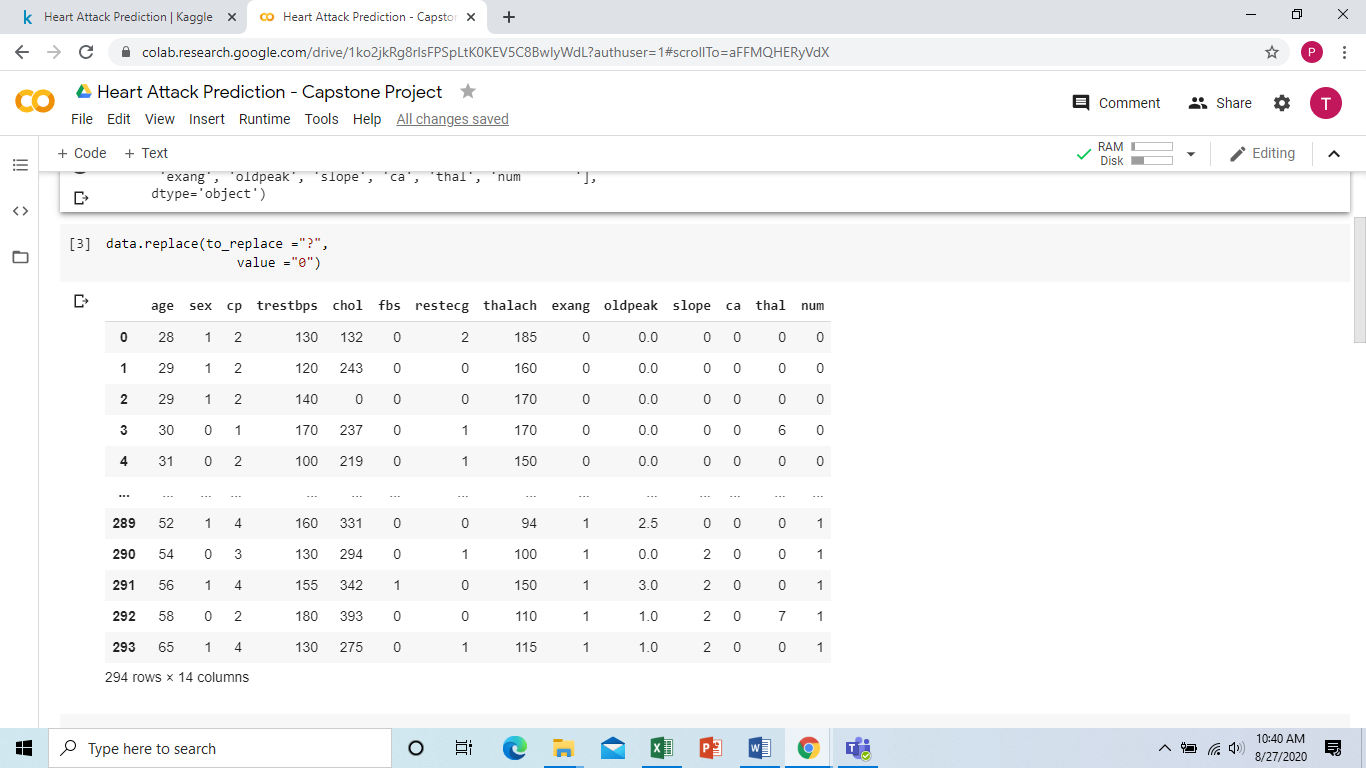
We need AI to solve this problem because AI is able to classify a large amount of data within minimal time … humans would not be able to predict. AI is very precise and can name the exact chances of a person getting a heart attack. Now a days heart attacks have the highest death toll apart from pandemic and epidemics. If AI is able to predict the heart attack beforehand the doctors would help the patients prevent the heart attack or at least reduce the intensity of it via treatments. This way it would also help save lives of many people.

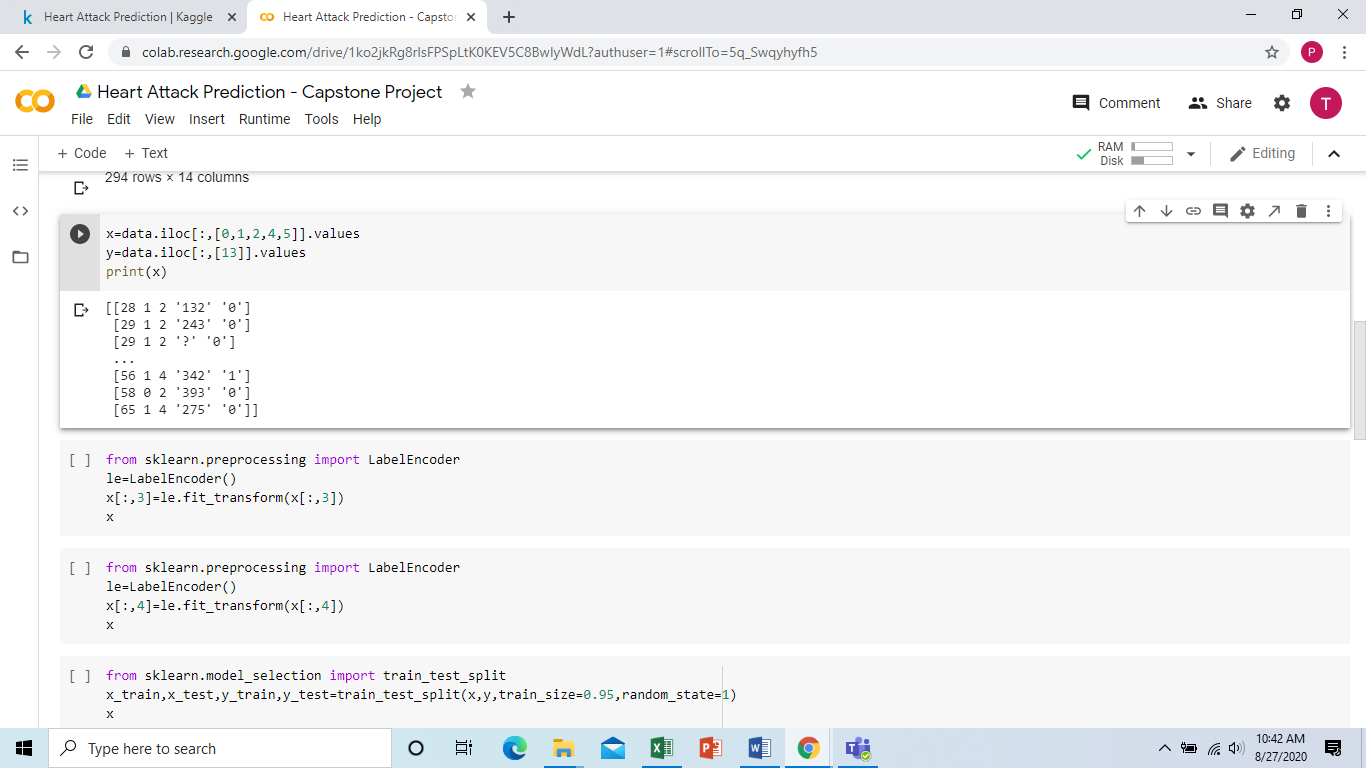
1. **Explain algorithm**

For this project I am using the KNN classification algorithm. KNN or K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. KNN algorithm is one of the simplest classification algorithm. Even with such simplicity, it can give highly competitive results. I used a classification algorithm as for this dataset we had to predict whether a person is prone to getting a heart attack or not.

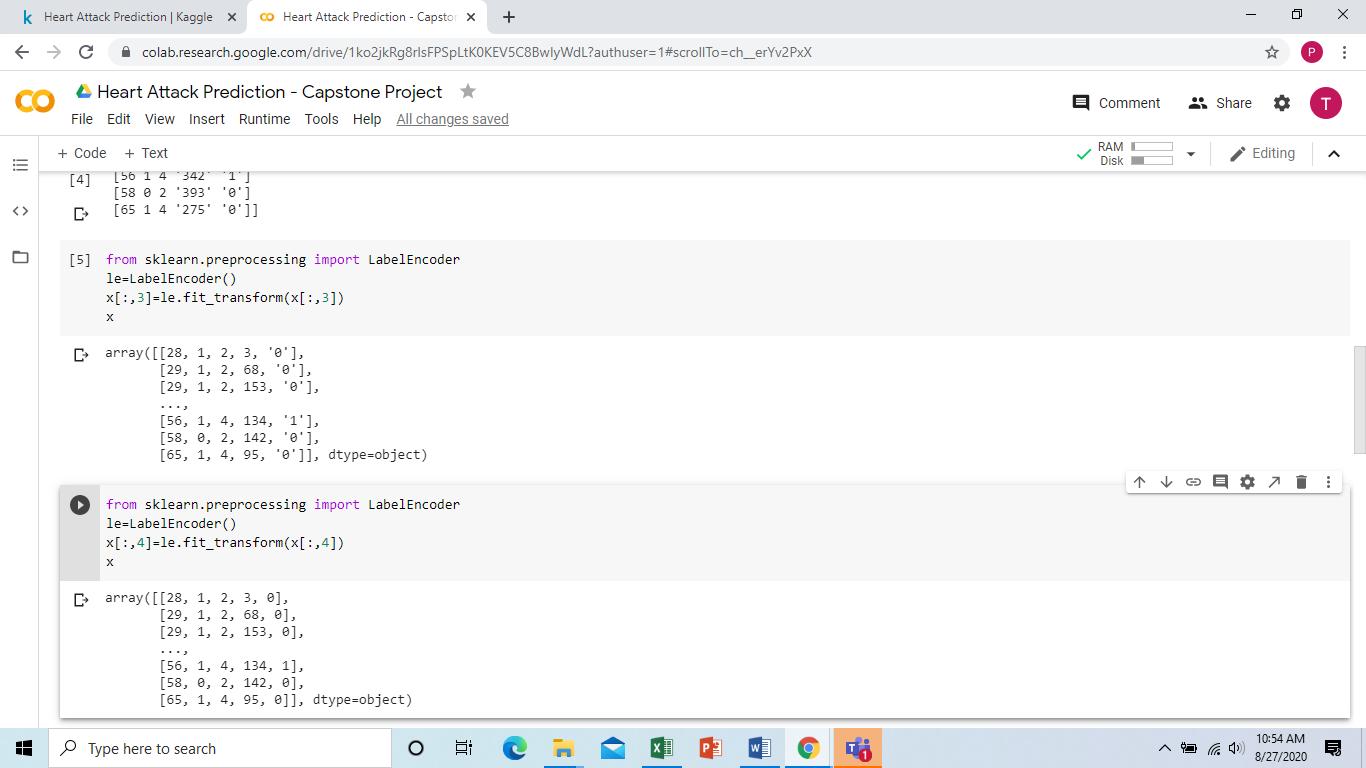
1. **Explain the code**
2. The first line code is to import pandas which is a library that helps import data sets or files. Then the next code is to load the dataset as a csv file. Next one is to print the columns.



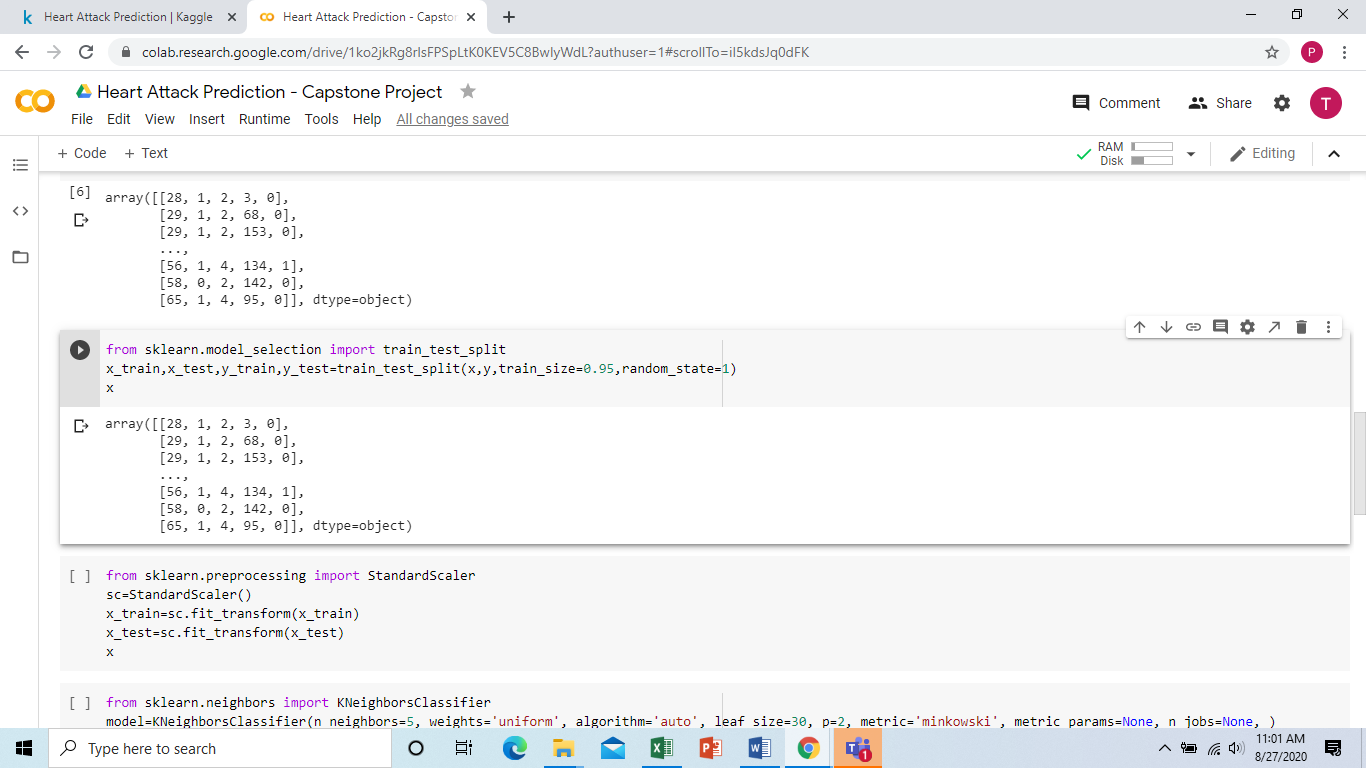
1. This code is to replace all question marks ‘?’ with zero ‘0’. As ‘?’ would not give the output.
2. This code is to state the independent and dependent variables from the dataset. My independent variable or features is age, sex, chest pain type, cholesterol level, fasting blood sugar. My dependent variable(target) or the variable I am testing is diagnosis of heart disease



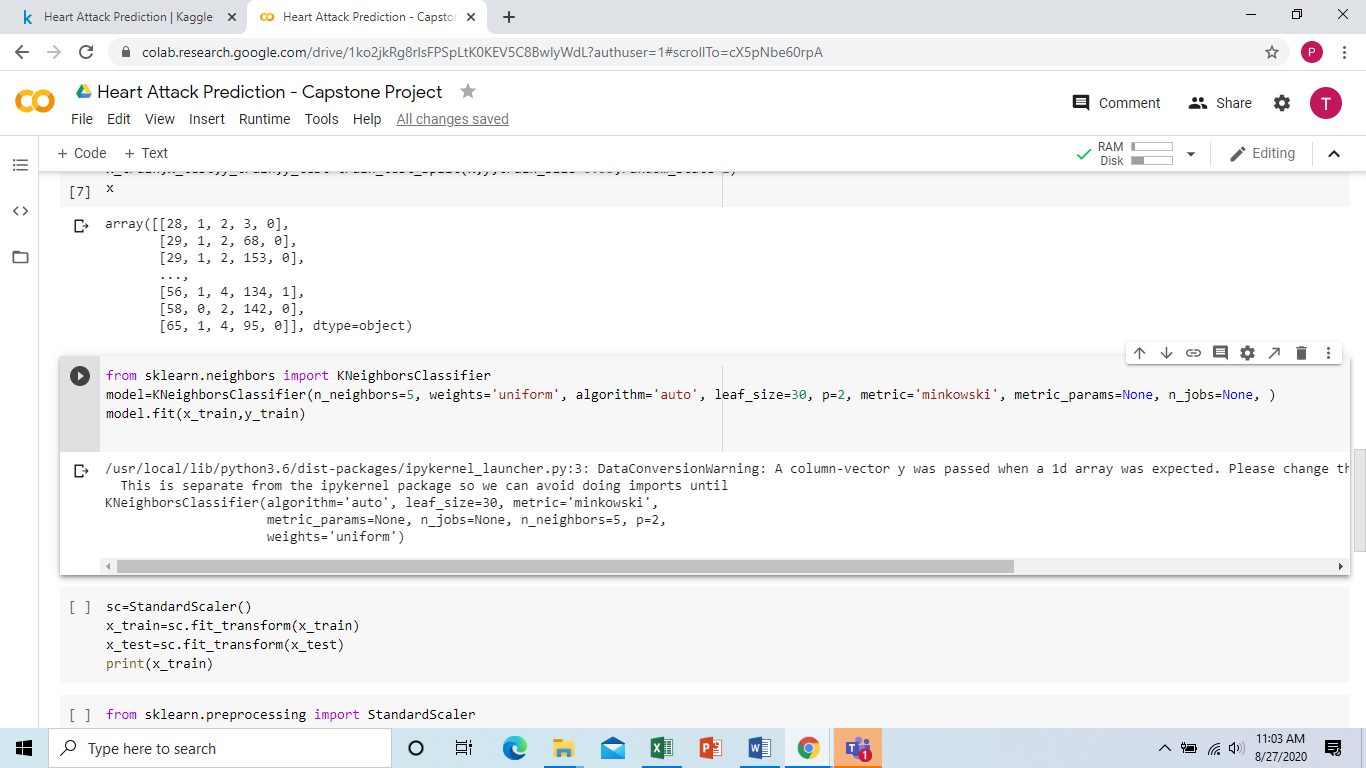
1. This is the code to transform all words into numbers so that the model is able to calculate via mathematics.



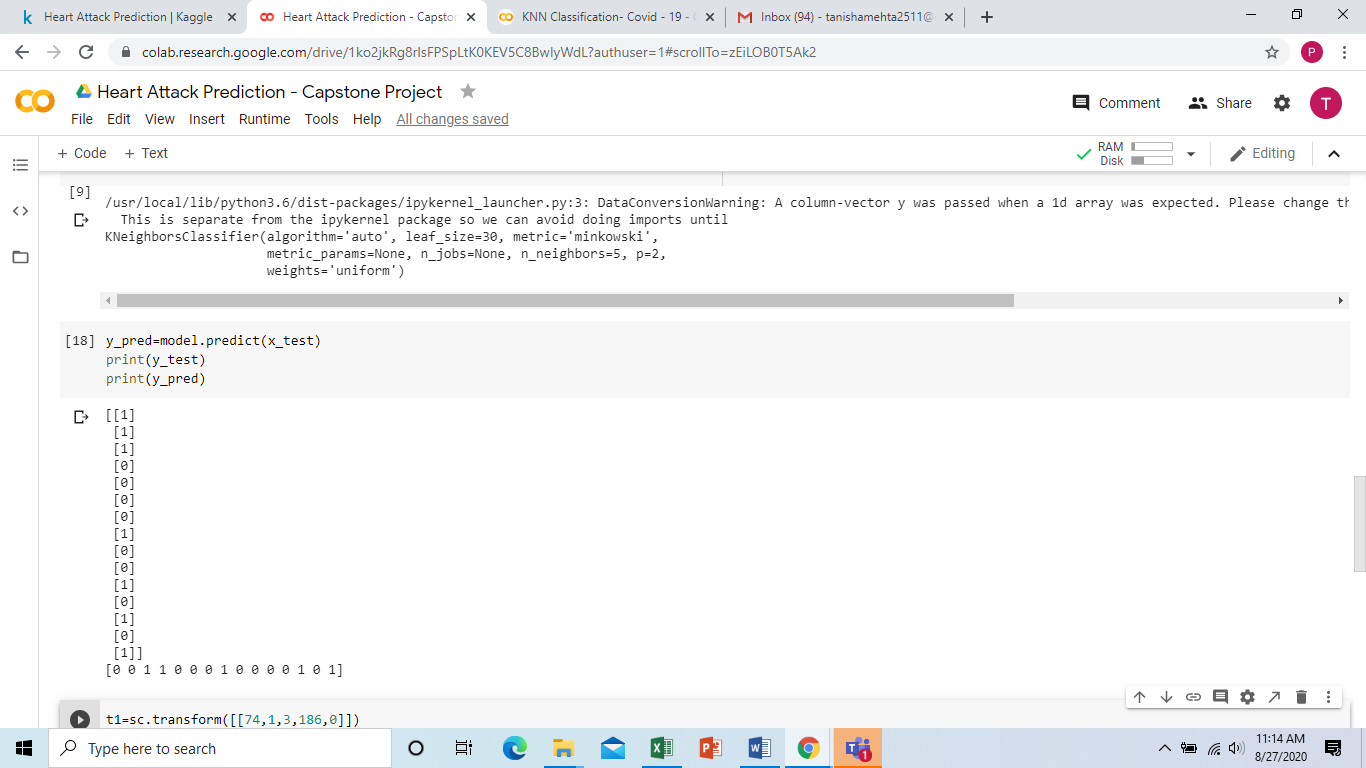
1. This code is to train , test and split the model.



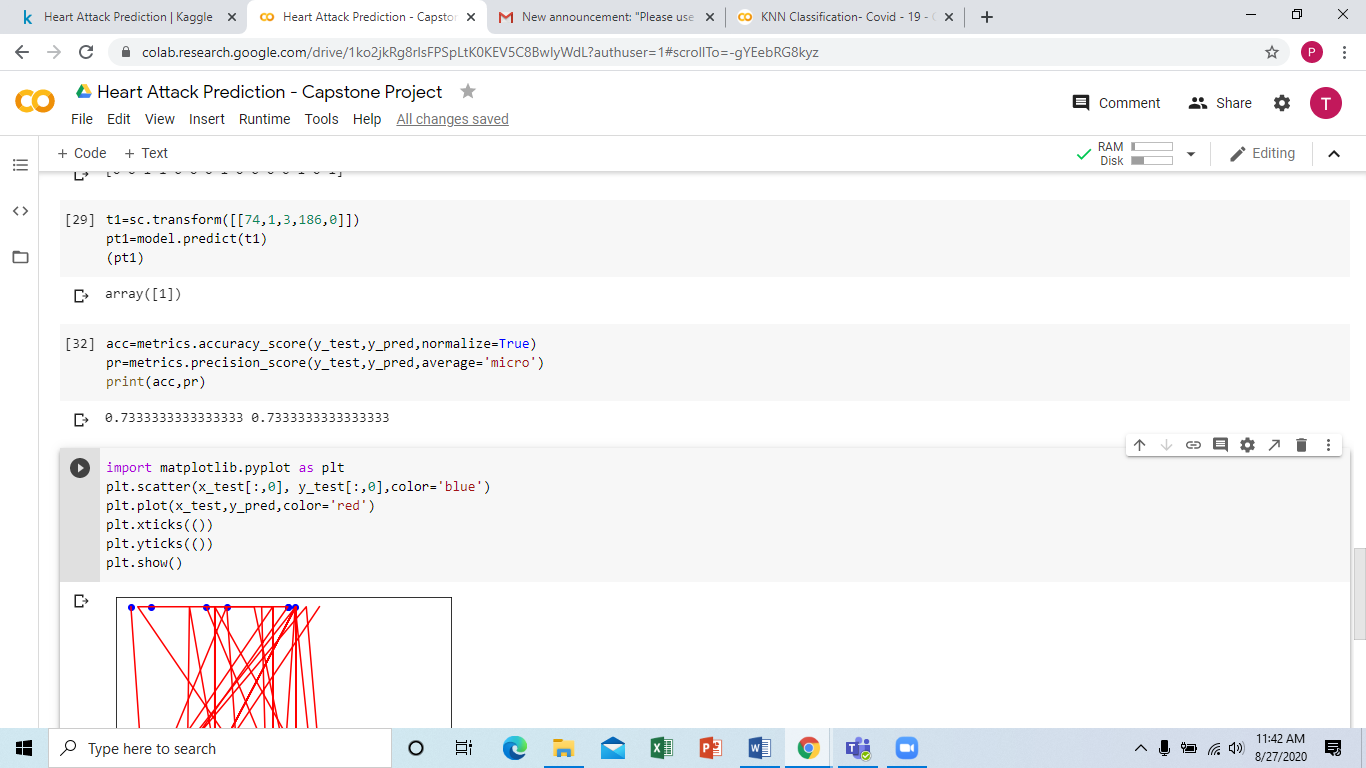
1. This code is to load the KNN classifier algorithm and give specific parameters so that the model is more efficient.



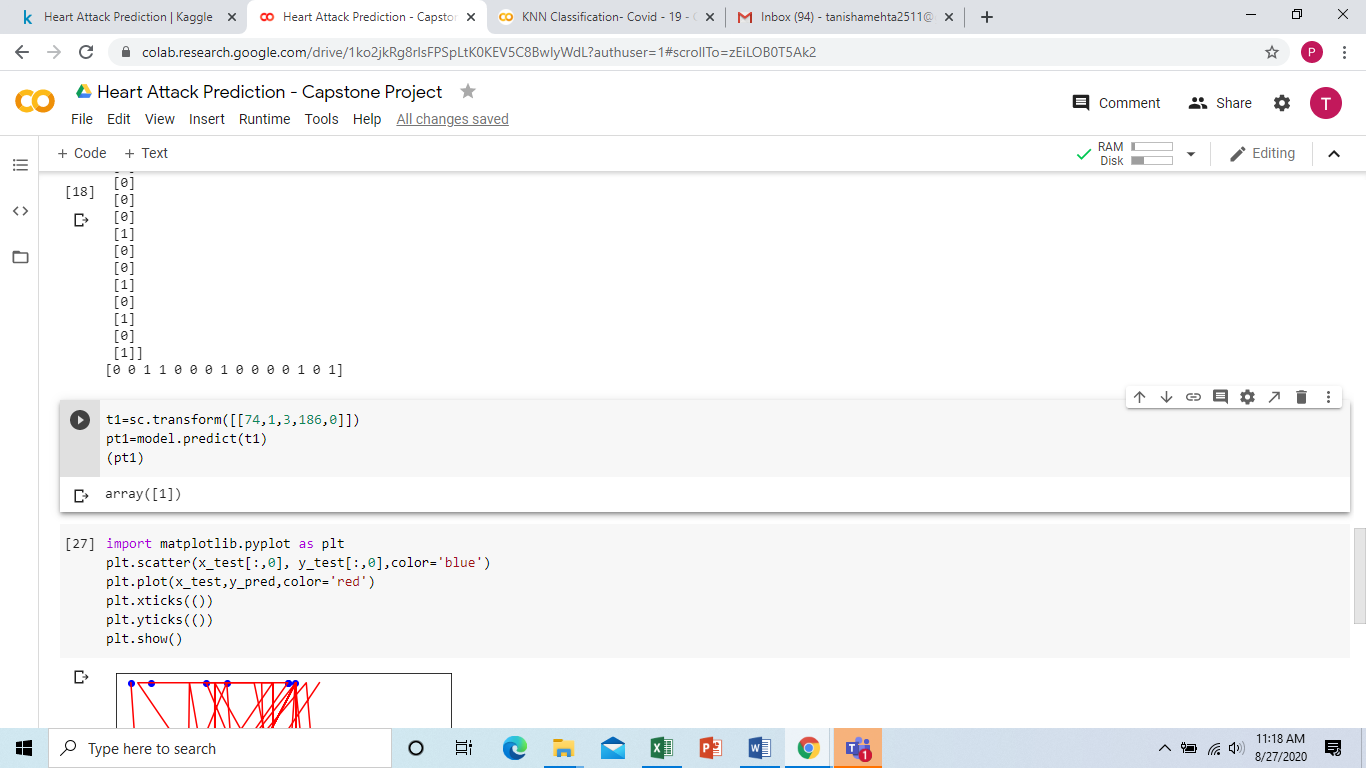
1. This code is to help the model predict the output



1. This code is to show how accurate and precise the model is. In this code it means that my output is o.73 accurate and precise.



1. This code is to predict outcome with the random input you have given. 1: means that > 50% diameter narrowing meaning the risk of getting heart attack is more.



1. This code is to plot a graph based on the data.

