

SARVESH PRATAP YADAV

2025

**Prediction of Survivals on Titanic Using Machine Learning Model**

**Problem Statement -** The RMS Titanic was a British passenger liner that sank in the North

Atlantic Ocean in the early morning hours of 15 April 1912, after it collided with an iceberg

during its maiden voyage from Southampton to New York City. There were an estimated

2,224 passengers and crew aboard the ship, and more than 1,500 died, making it one of

the deadliest commercial peacetime maritime disasters in modern history. The RMS

Titanic was the largest ship afloat at the time it entered service and was the second of

three Olympic-class ocean liners operated by the White Star Line. The Titanic was built by

the Harland and Wolff shipyard in Belfast. Thomas Andrews, her architect, died in the

disaster.

In this Project, we will analyze the Titanic data set and make two predictions. One

prediction to see which passengers on board the ship would survive and then another

prediction to see if we wouldn’t survive.

**Steps to be taken in the Project is sub-divided into the following sections.**

**These are:**

 Load the necessary libraries such as Numpy , Pandas , sklearn.model etc.

 Loading the dataset as csv file and showing first ten rows.

 Drop the unnecessary columns from the data.

 Calculate statistical values and round them up to 3 decimal places.

 Checking for null values and return their sum of numbers of true values in

each column.

 Handle the null by mean of all values fill into them.

 Visualization of Passenger Survival data using Data Visualization with

Python.

 Data preprocessing or (Data cleaning) performed by the one hot encoding

in this process we change categorical data into numerical data and the

technique is called feature Engineering.

 Splitting the cleaned data into dependent and independent variables.

 Splitting the data into train and test sets with train\_test\_split using sklearn

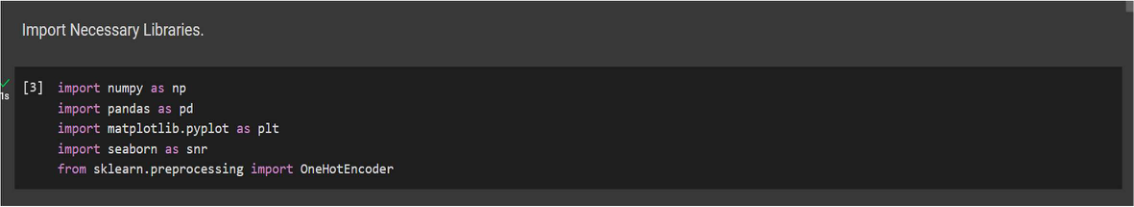
library.

 Import different kind of Classification Models and Train that model with the

help of .fit().

 Predicting the trained models and then checking their accuracy score and

confusion metrics of the model using confusion metrics & accuracy score.



2025

SARVESH PRATAP YADAV

 Then recall the train\_test\_split and split the data into training and testing

set with different models.

 Then predicting the trained models and checking the accuracy of model and

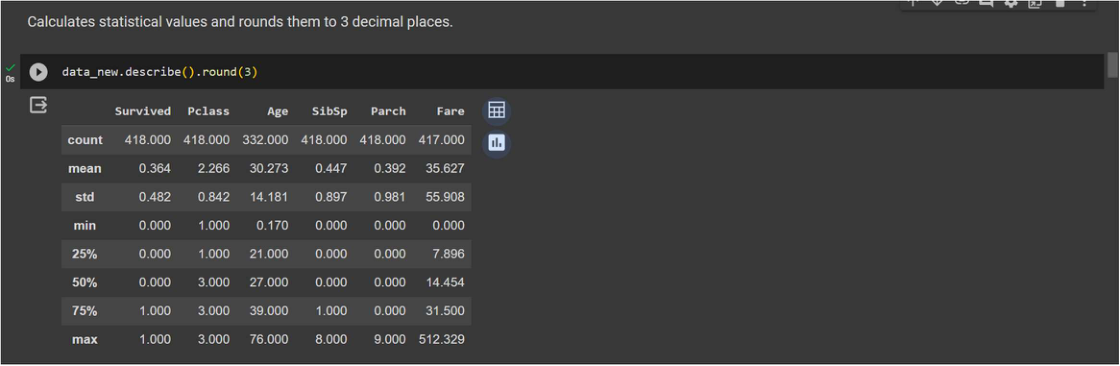
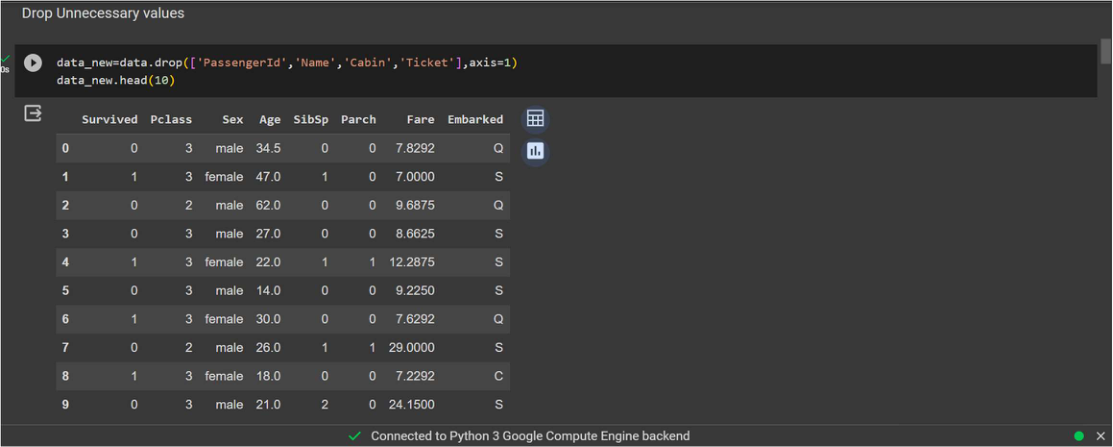
print the accuracy difference.

 And finally predict whether the Titanic Survivals classification generated or

not.

**Step-1 –** Loading Necessary Libraries used in machine learning.

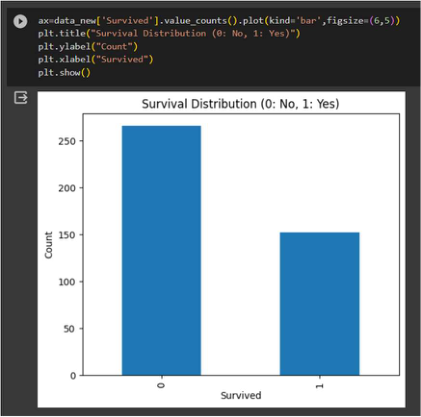
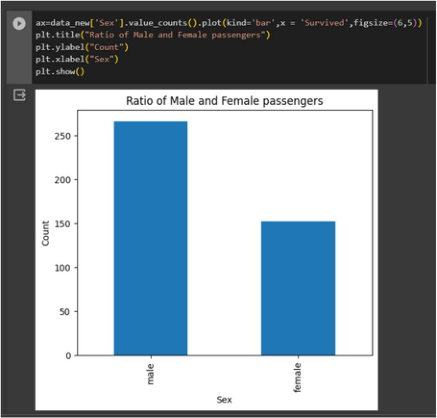
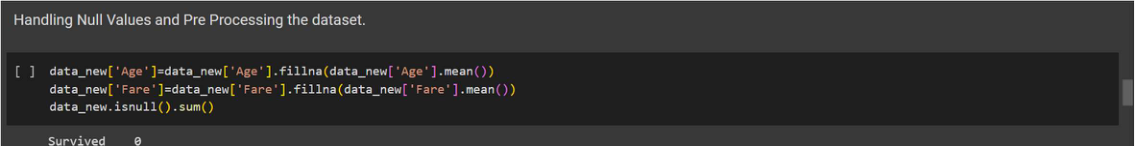
**Step-2 -** Loading the dataset as csv file and showing first ten rows.



SARVESH PRATAP YADAV

**Step-3 -** Drop the unnecessary columns from the data.

**Step-4 -** Calculate statistical values and round them up to 3 decimal places.



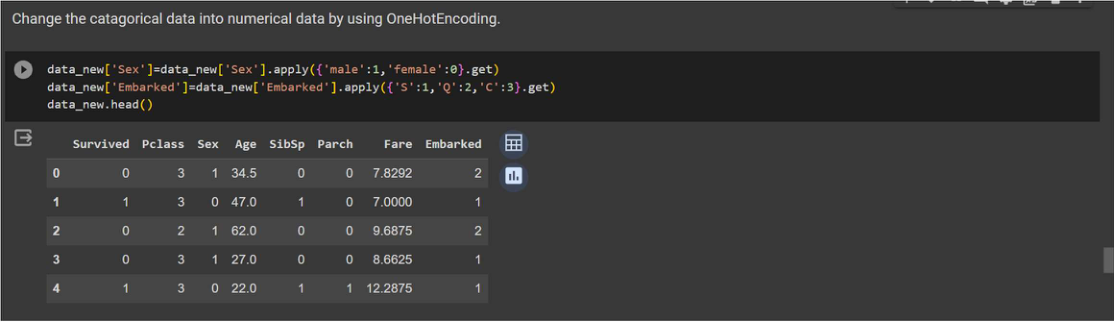
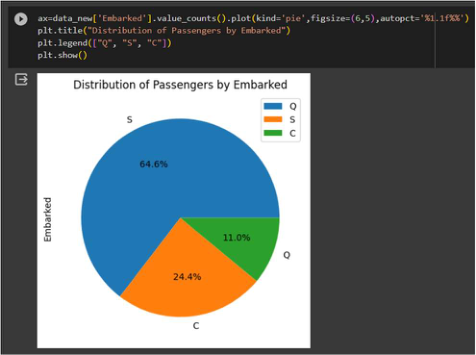
SARVESH PRATAP YADAV

**Step-5 -** Checking for null values and return their sum of numbers of true values

in each column.

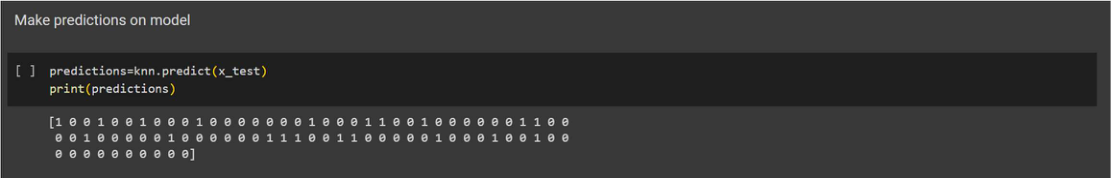
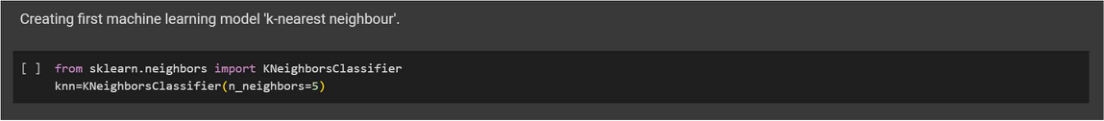
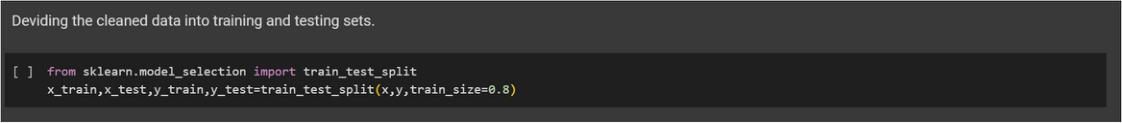
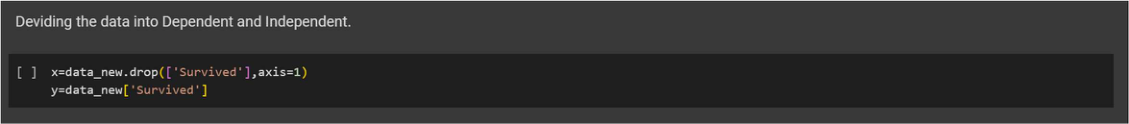
**Step-6 -** Handle the null by mean of all values fill into them.

**Step-7 -** Visualization of Passenger Survival using Data Visualization with Python.



SARVESH PRATAP YADAV

**Step-8 -** Data preprocessing or (Data cleaning) performed by the one hot encoding   
in this process we change categorical data into numerical data and the technique   
is called feature Engineering.



SARVESH PRATAP YADAV

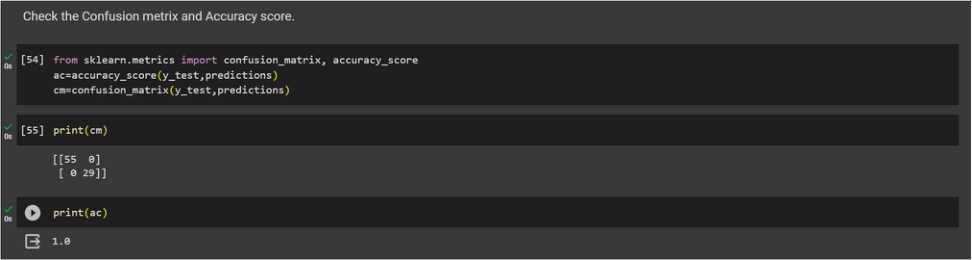
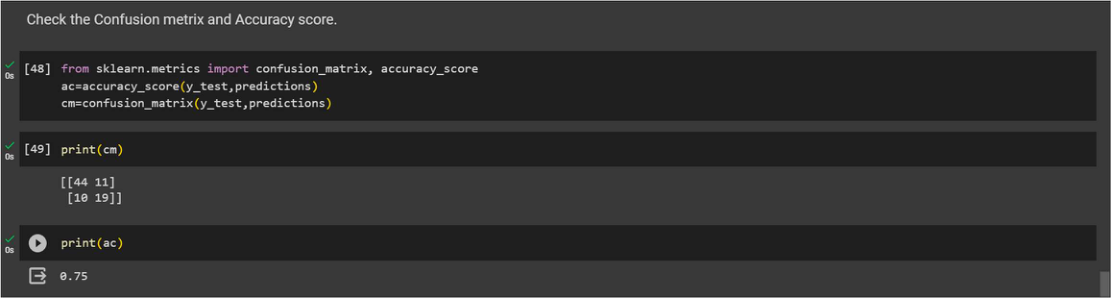
**Step-9 -** Splitting the cleaned data into dependent and independent variables.

**Step-10 -** Splitting the data into train and test sets with train\_test\_split using   
 sklearn library.

**Step-11 -** Import first machine learning model K-Nearest neighbor taking   
 n\_neighbor=5.

**Step-12 -** Train the model using .fit() function.

**Step-13 -** Predict the trained model using .predict() function.



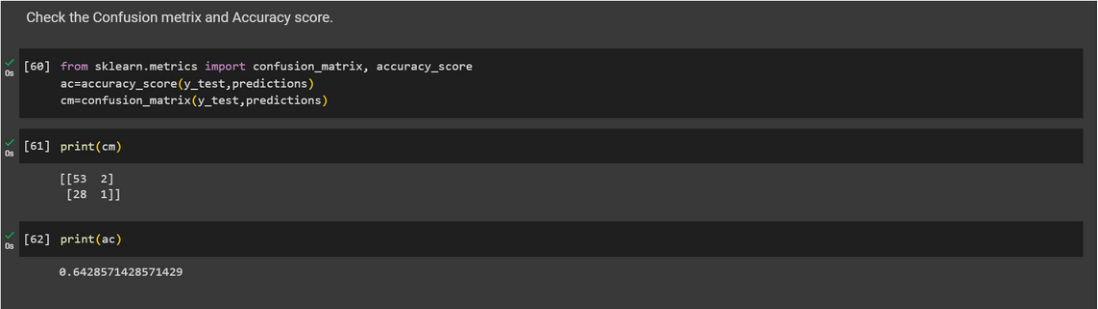
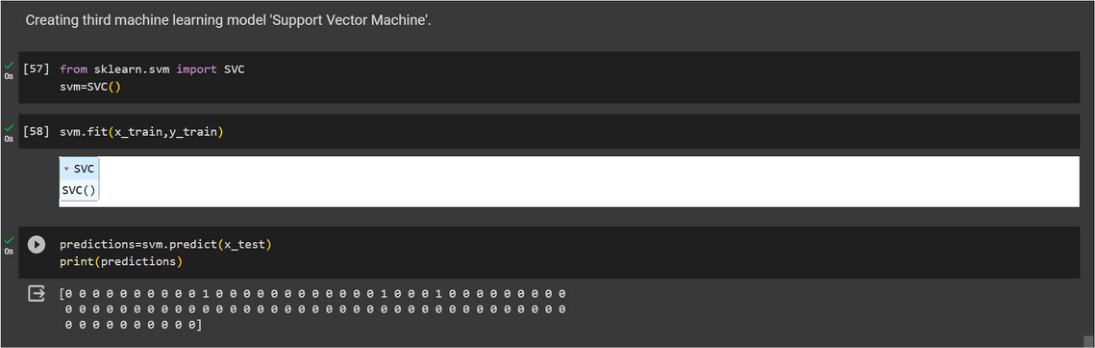
SARVESH PRATAP YADAV

**Step-14 -** Check the accuracy score and print a confusion metrics with   
 confusion metrics & accuracy score.

**Step-15 –** Import the Second Machine Learning Model Decision Tree and train   
 model and then make prediction.

**Step-16 -** Print a confusion metrics and check accuracy score for Decision Tree

Model.



SARVESH PRATAP YADAV

**Step-17 -** Import the Third Machine Learning Model Support Vector Machine and   
 train model and then make prediction.

**Step-18 -** Print a confusion metrics and check accuracy score for Support Vector

Machine Model.

**Conclusion -** The purpose of Project is to use the existing features of passengers onboard   
Titanic as predictors to predict their survival outcome, for 0 being dead and 1 being   
survived from the tragic ship crash. The K-Nearest neighbor is the is first classification   
model performed with k=5, and the I use Random Forest Classifier, and then I use the   
Support Vector Machine (SVM) analysis to improved performance. It is certain through   
the practice of model improvement, the SVM analysis is better performed than the K-  
Nearest Neighbor classification analysis and Random Forest analysis is also performed   
better than K-Nearest Neighbor for prediction accuracy.



2025

SARVESH PRATAP YADAV

However, even from the all three classification model, we can easily see that the Titanic   
survival outcome is highly depended on several predictors, such as sex, age and passenger   
class. In particular, Ratio of survived people are more while keeping other predictors   
conditions constant and lastly, people from a lower class are less likely to survived keeping   
other predictors conditions constant.