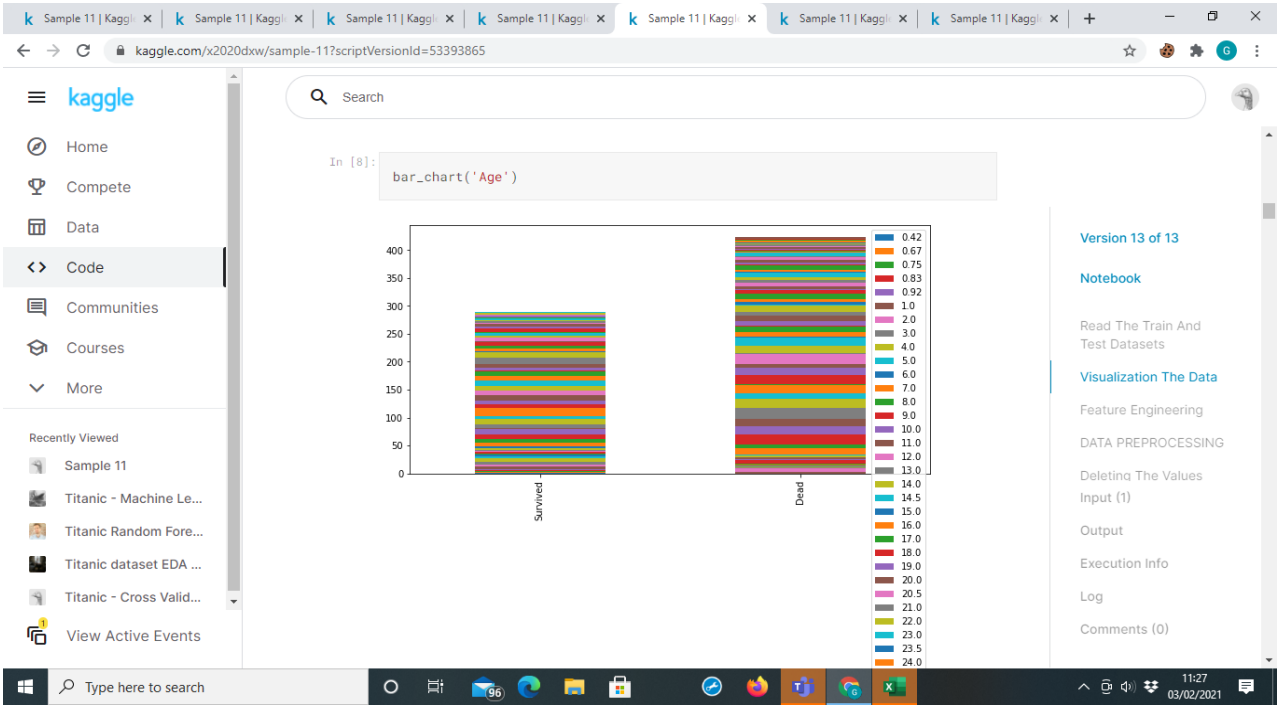
REPORT ON SUBMISSION

Steps involved in my code

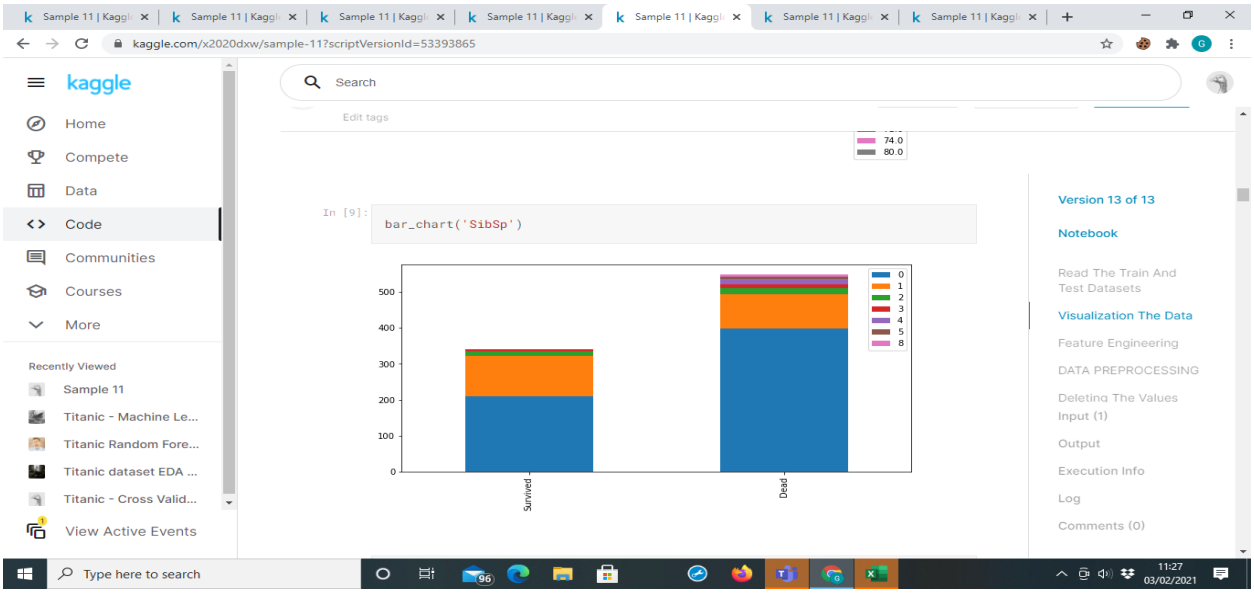
- **I** wrote my code in Kaggle Notebook and used the Dataset and File location from the Kaggle itself.
- READ THE TRAIN AND TEST DATASET
  1. I have read the train and test dataset using read_csv from the dataset provided in the Kaggle.
  2. I used info() function to check the summary of data frame then I found few missing values.

VISULIAZATION OF DATA

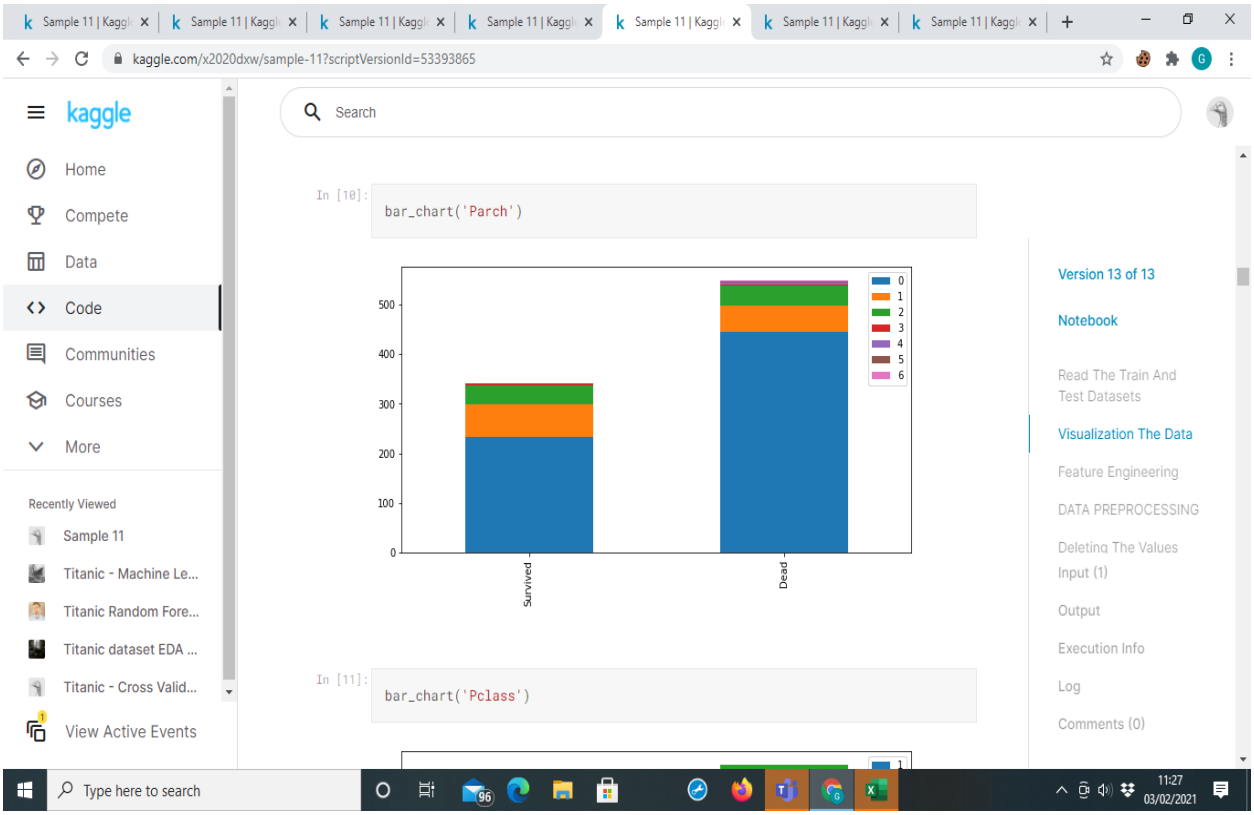
- **To** get visualization of data I used bar plot which was imported from matplotlib library.
- **I** divided the train dataset in to Survived and dead with assigning values from 'Survived' column.
- **Plotted bar gra**ph for different columns like
  1. **AGE**



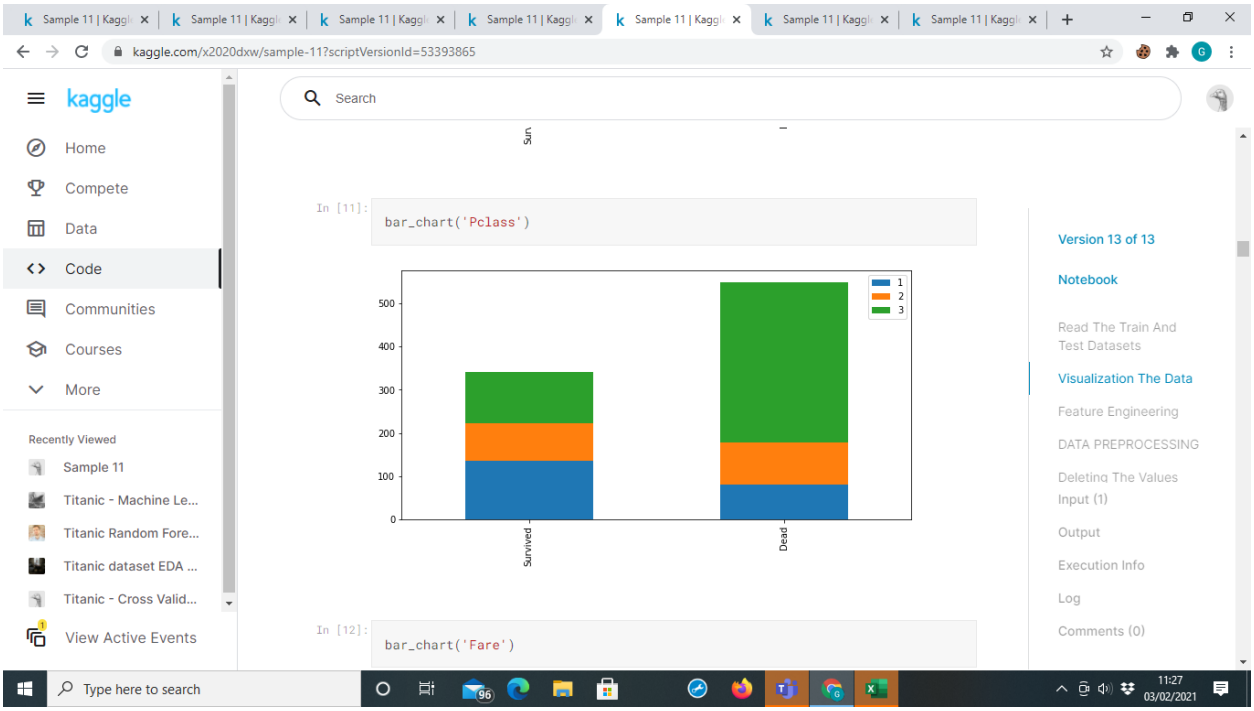
2▪Sibsp



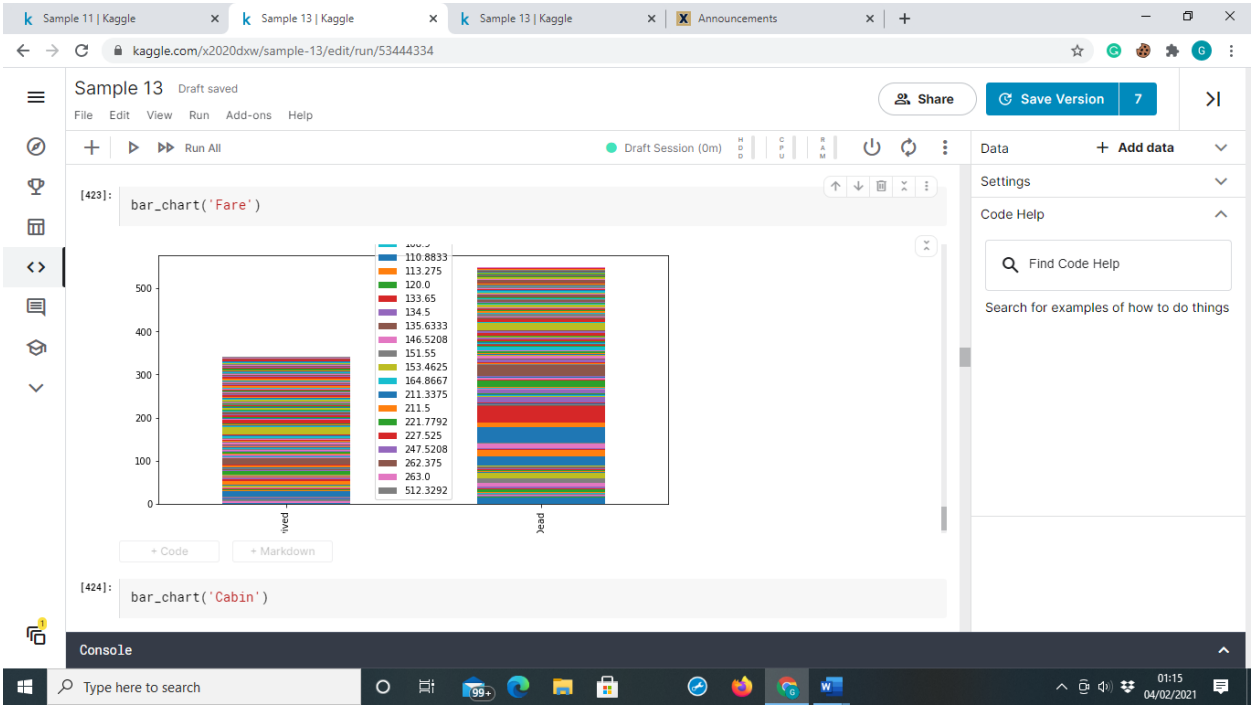
3.Parch



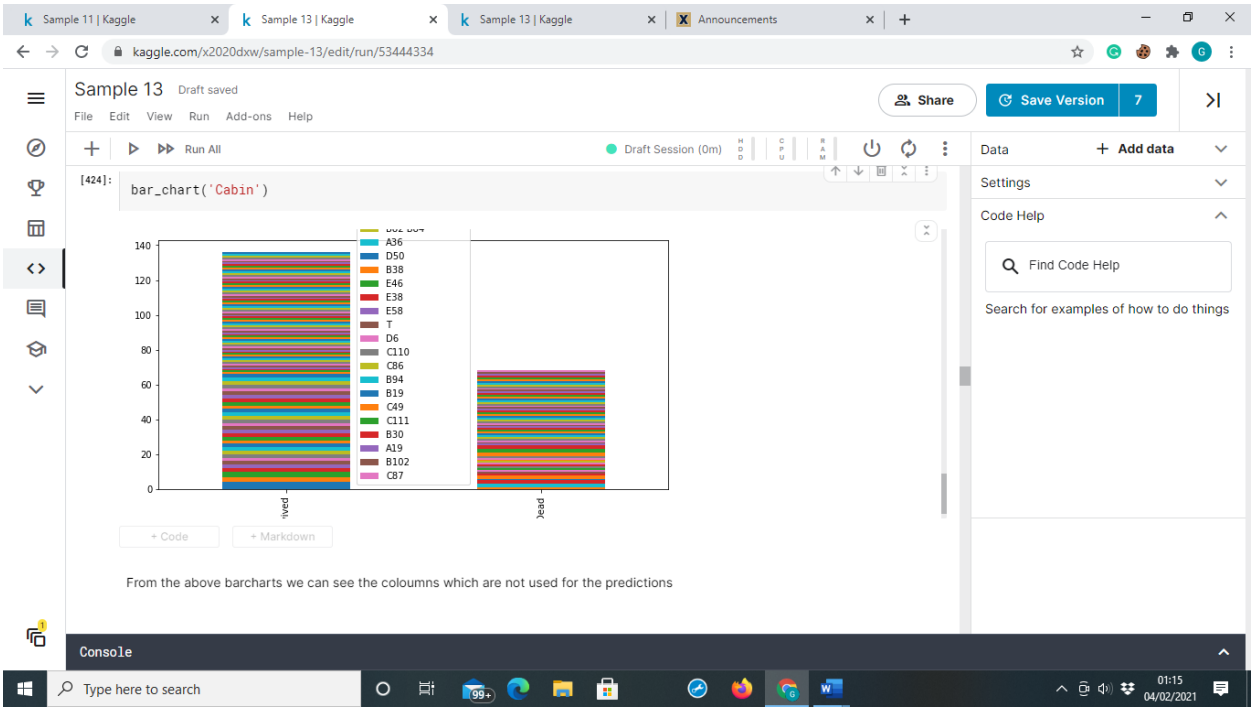
4.Pclass



5.Fare



6.Cabin



- From the above 6 visualization we can notice that Age, Fare, Cabin columns are not predicting the most and they also contain a large number of unique values. whereas the remaining providing the accurate result about the survived and dead people by their classification.

Filling the missing

- To get clean dataset we need to treat missing values.
- fillna() function is used to fill missing value in both train and test datasets.

Feature Engineering

- I created a column name ‘FamilySize’ which contains a Family with the combination of “SibSP”, “Parch” columns and one more person.
- I have used ‘Name’ columns which contains a large number of unique columns. So, I Performed some feature engineering techniques like extracting Title name from the ‘Name’ columns and giving corresponding values to them by checking their value_counts. The most repeated one got different values and remaining got the same value which is equal to 3 by using Title mapping.
- To get prediction we can observe from the train and test data set that age group of 16 are more survived. So, to get more predictions we divide Age column into four different Age groups.
- Similarly, the same techniques is applied on the Fare columns .it is divided in to four different fare categories based on the price.
- We divided Age, Fare columns by using cut() function with parameters like bins and labels where the values for the parameters are metioned.

Deleting the columns

- Dropped few columns like [Name, Age, Fare, Cabin, Ticket] these columns don’t make any effect on test dataset to get predictions about survival.
- From the above visualization we notice this columns are not suitable for predictions.
- The above dropped columns also consist few null values. So, for any prediction model it should contain clean data to predict.

Converting the datatype

- We are using sklearn model for prediction about the survival rate. So sklearn doesn’t accept string datatype. It needs the input should present in the Numerical Format.
- get_dummies() are used for converting of categorical variables into dummy or indicator variables with specified column name.

Train_test_split

- Both independent and Dependent variables are present in train dataset. So, we need to separate Independent and Dependent variable by creating (X, Y variables) for dataset.
- For splitting the train and test dataset we use train_test_split function in sklearn model selection.

SUPPORT VECTOR MACHINE

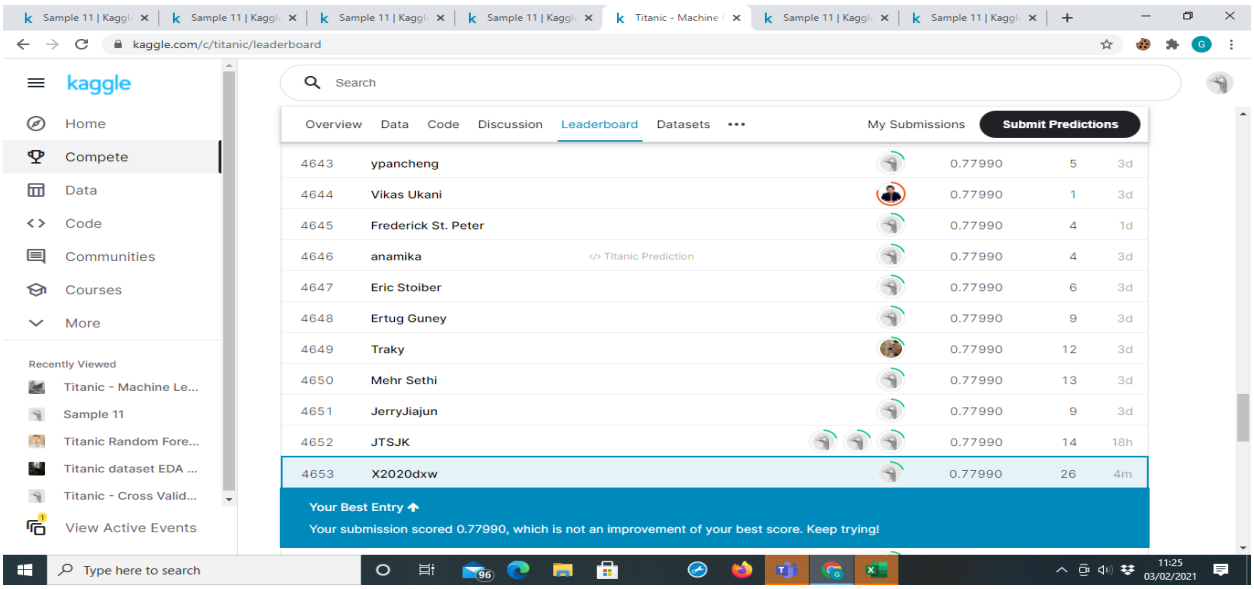
- Support vector machine is a Supervised machine learning algorithm which is mainly used in Classification challenges for predictions.
- It is a sklearn model imported from SVC library
- We use fit() function for training dataset to adjust weights according to data values for better accuracy
- Score() function is used to check the performance of train and test dataset and return the accuracy score.

Grid search and cross validation

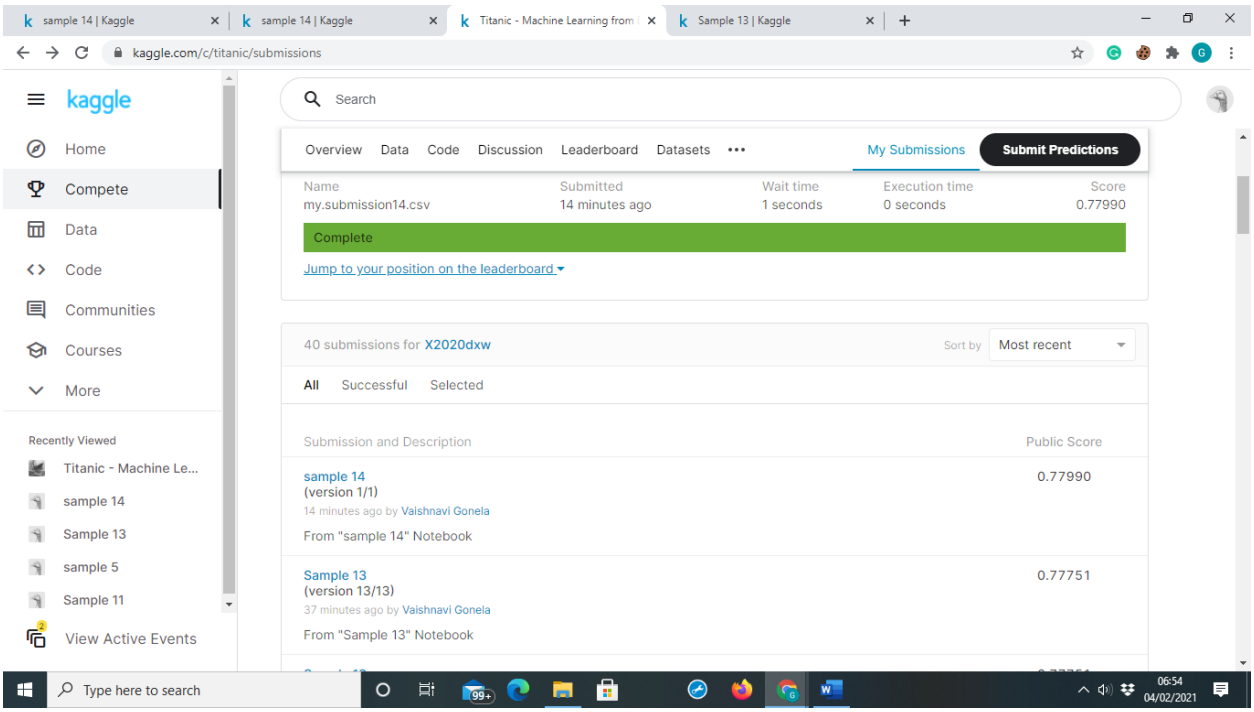
- We Grid search and hyper parameter tunning to get best accuracy.
- Grid search is also a sklearn model imported from GridResearchCV.
- We used some parameter selection method. Firstly, we defined some parameter and selected the best parameters from them by using GridSearch.
- Used these parameters for SVM Model and calculated the accuracy score.

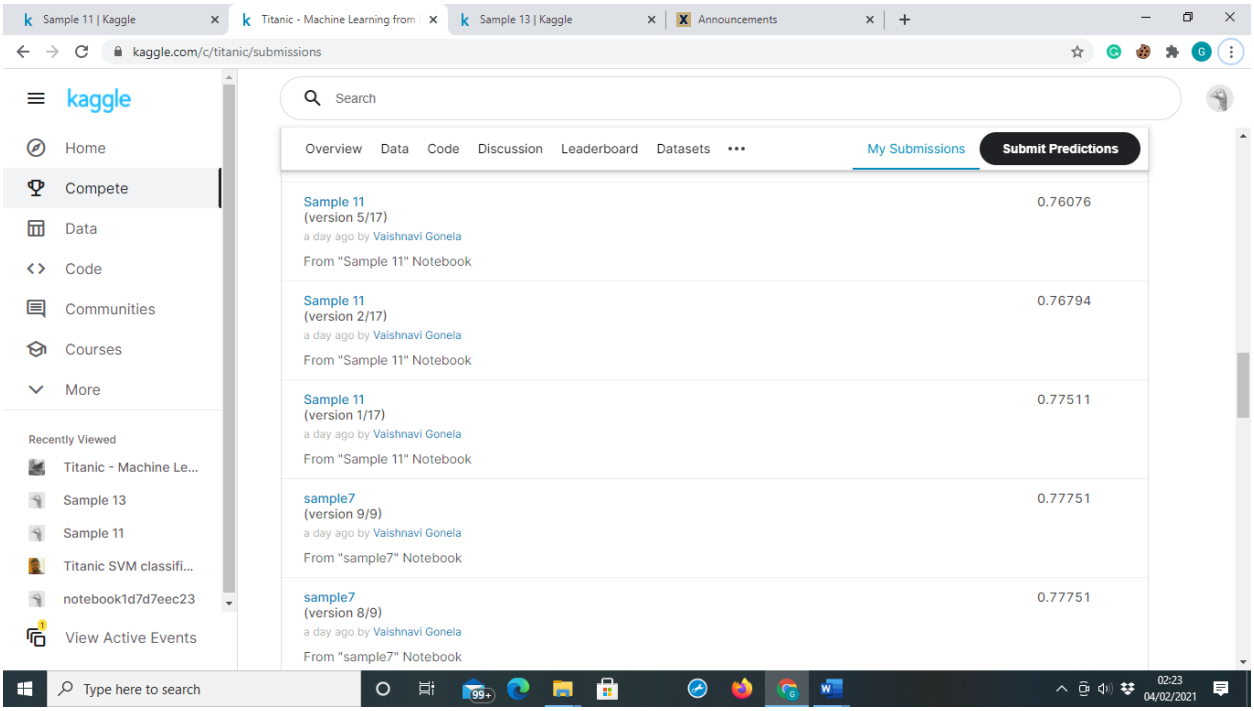
Predictions

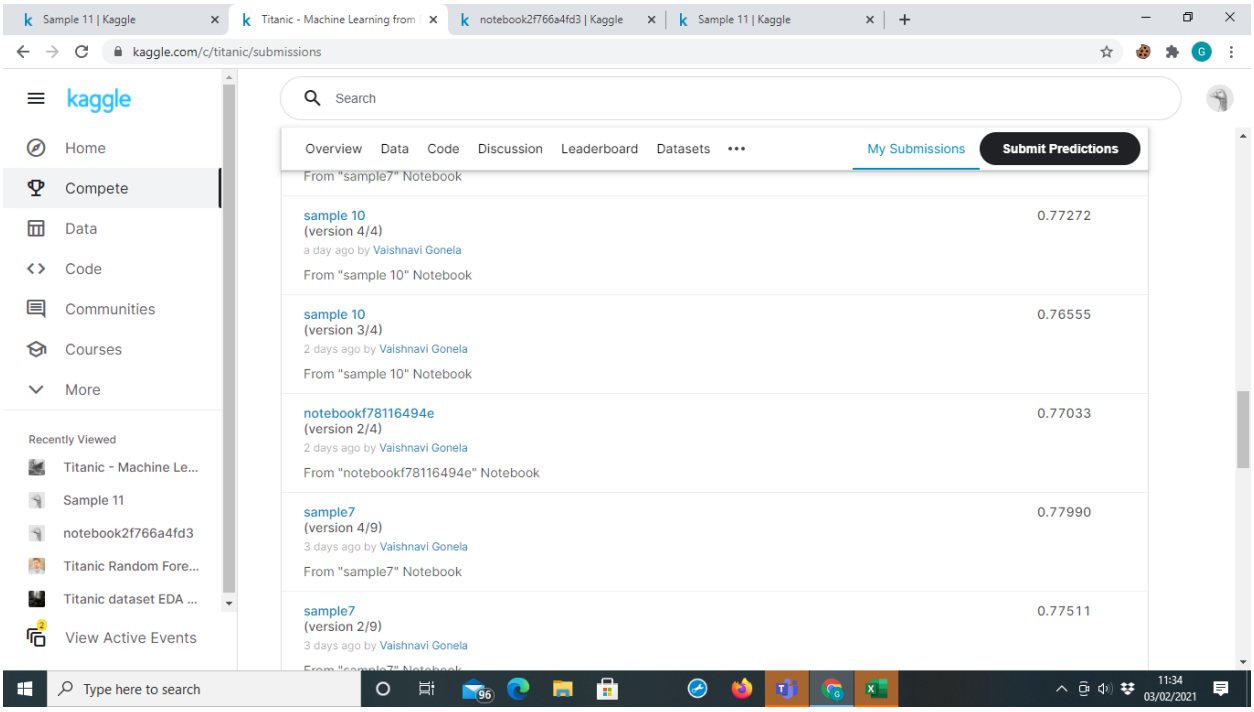
- For predictions we have created a new variable and dropped “PassengerId” from the test data set and stored in the newly created variable.
- We perform prediction by using pred() function.
- To get the output file we have concatenated the “PassengerId” column from the test dataset and “Survived “column from the prediction variable.
- The output file is stored with (.csv extension).
- Submitted the result to the Kaggle and scored 0.7799 accuracy with position of 4,653 on the leader board.

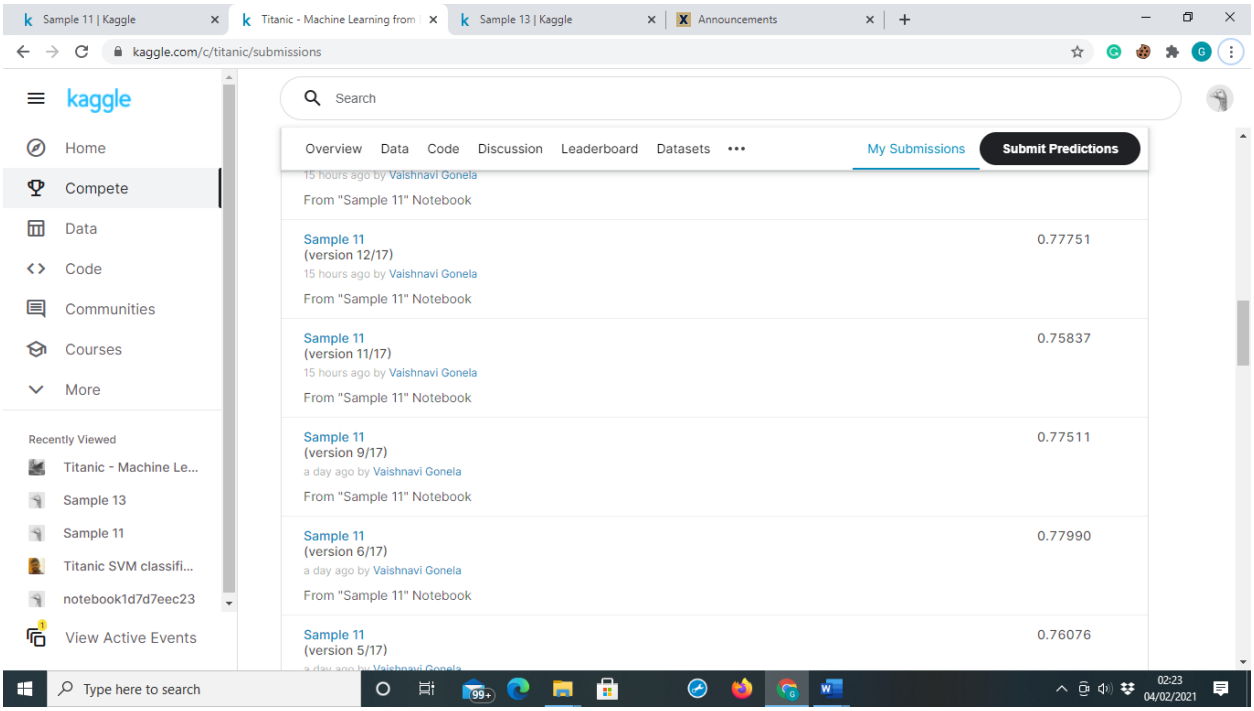


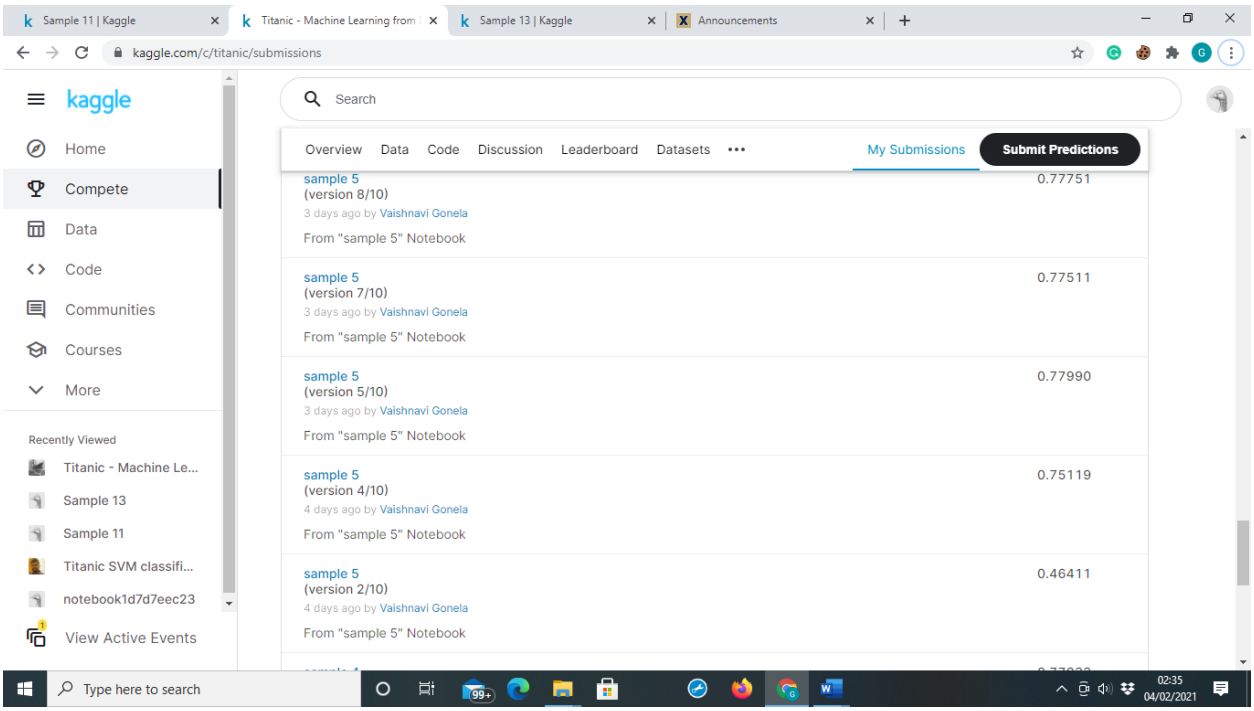
- I have also tried few other Model like Logistic Regressions, Decision tree classifier, RandomForestClassifeirs etc.…, but I got high accuracy by using SVM algorithm.
- I did around 40 submission in Kaggle by using different models and different approaches to get the better accuracy for my predictions.
- Below are few of my submissions score screenshots in Kaggle platform.
-




Sample 11 | Kaggle
Titanic - Machine Learning from
Sample 13 | Kaggle
Announcements
kaggle.com/c/titanic/submissions
kaggle
Search
Home
Compete
Data
Code
Communities
Courses
More
Recently Viewed
Titanic - Machine Le...
Sample 13
Sample 11
Titanic SVM classifi...
notebook1d7d7eec23
View Active Events
Overview
Data
Code
Discussion
Leaderboard
Datasets
My Submissions
Submit Predictions
Sample 11
(version 5/17)
a day ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.76076
Sample 11
(version 2/17)
a day ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.76794
Sample 11
(version 1/17)
a day ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.77511
sample7
(version 9/9)
a day ago by Vaishnavi Gonela
From "sample7" Notebook
0.77751
sample7
(version 8/9)
a day ago by Vaishnavi Gonela
From "sample7" Notebook
0.77751
Type here to search
02:23
04/02/2021


Sample 11 | Kaggle
Titanic - Machine Learning from
notebook2f766a4fd3 | Kaggle
Sample 11 | Kaggle
kaggle.com/c/titanic/submissions
kaggle
Search
Home
Compete
Data
Code
Communities
Courses
More
Recently Viewed
Titanic - Machine Le...
Sample 11
notebook2f766a4fd3
Titanic Random Fore...
Titanic dataset EDA ...
View Active Events
Overview
Data
Code
Discussion
Leaderboard
Datasets
My Submissions
Submit Predictions
From "sample7" Notebook
sample 10
(version 4/4)
a day ago by Vaishnavi Gonela
From "sample 10" Notebook
0.77272
sample 10
(version 3/4)
2 days ago by Vaishnavi Gonela
From "sample 10" Notebook
0.76555
notebookf78116494e
(version 2/4)
2 days ago by Vaishnavi Gonela
From "notebookf78116494e" Notebook
0.77033
sample7
(version 4/9)
3 days ago by Vaishnavi Gonela
From "sample7" Notebook
0.77990
sample7
(version 2/9)
3 days ago by Vaishnavi Gonela
0.77511
Type here to search
11:34
03/02/2021


Sample 11 | Kaggle
Titanic - Machine Learning from
Sample 13 | Kaggle
Announcements
kaggle.com/c/titanic/submissions
kaggle
Search
Home
Compete
Data
Code
Communities
Courses
More
Recently Viewed
Titanic - Machine Le...
Sample 13
Sample 11
Titanic SVM classifi...
notebook1d7d7eec23
View Active Events
Overview
Data
Code
Discussion
Leaderboard
Datasets
My Submissions
Submit Predictions
15 hours ago by Vaishnavi Gonela
From "Sample 11" Notebook
Sample 11
(version 12/17)
15 hours ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.77751
Sample 11
(version 11/17)
15 hours ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.75837
Sample 11
(version 9/17)
a day ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.77511
Sample 11
(version 6/17)
a day ago by Vaishnavi Gonela
From "Sample 11" Notebook
0.77990
Sample 11
(version 5/17)
0.76076
Type here to search
02:23
04/02/2021


Sample 11 | Kaggle
Titanic - Machine Learning from
Sample 13 | Kaggle
Announcements
kaggle.com/c/titanic/submissions
kaggle
Search
Home
Compete
Data
Code
Communities
Courses
More
Recently Viewed
Titanic - Machine Le...
Sample 13
Sample 11
Titanic SVM classifi...
notebook1d7d7eec23
View Active Events
Overview
Data
Code
Discussion
Leaderboard
Datasets
My Submissions
Submit Predictions
sample 5
(version 8/10)
3 days ago by Vaishnavi Gonela
From "sample 5" Notebook
0.77751
sample 5
(version 7/10)
3 days ago by Vaishnavi Gonela
From "sample 5" Notebook
0.77511
sample 5
(version 5/10)
3 days ago by Vaishnavi Gonela
From "sample 5" Notebook
0.77990
sample 5
(version 4/10)
4 days ago by Vaishnavi Gonela
From "sample 5" Notebook
0.75119
sample 5
(version 2/10)
4 days ago by Vaishnavi Gonela
From "sample 5" Notebook
0.46411
Type here to search
02:35
04/02/2021