Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

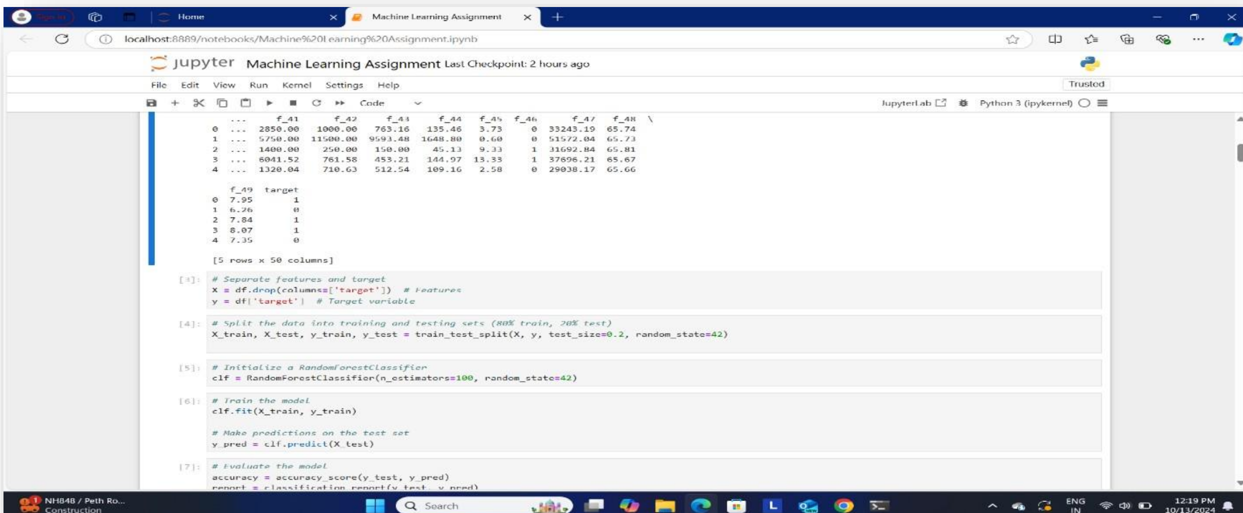
Email:- shrikanttawade17@gmail.com

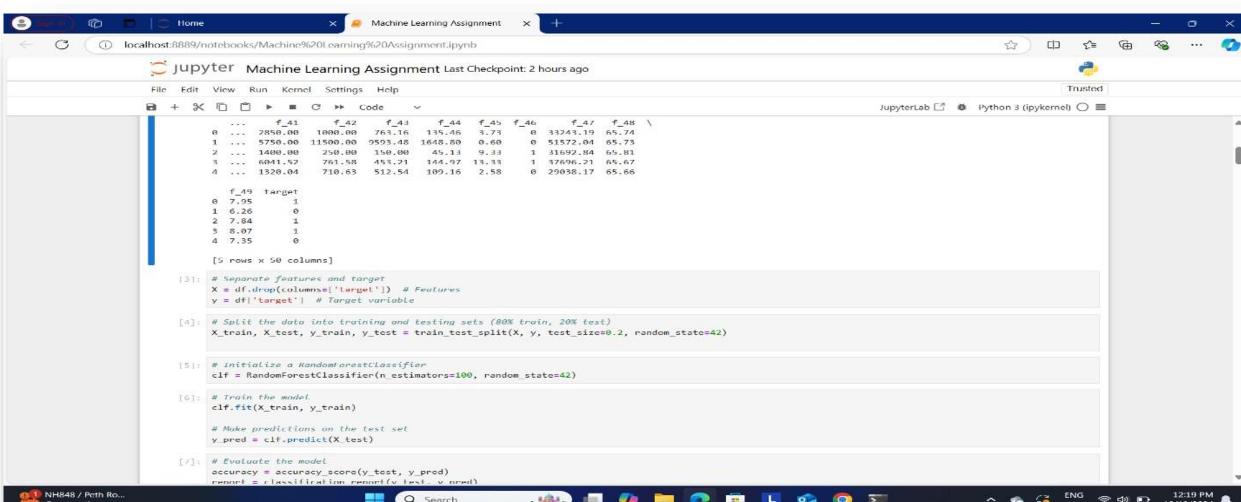
Machine Learning Assignment

Submission – [SHRIKANT VIJAY TAWADE]

shrikanttawade17@gmail.com

Q1 Download the Oil Spill Dataset and perform Data cleaning and Data Pre-Processing if Necessary?

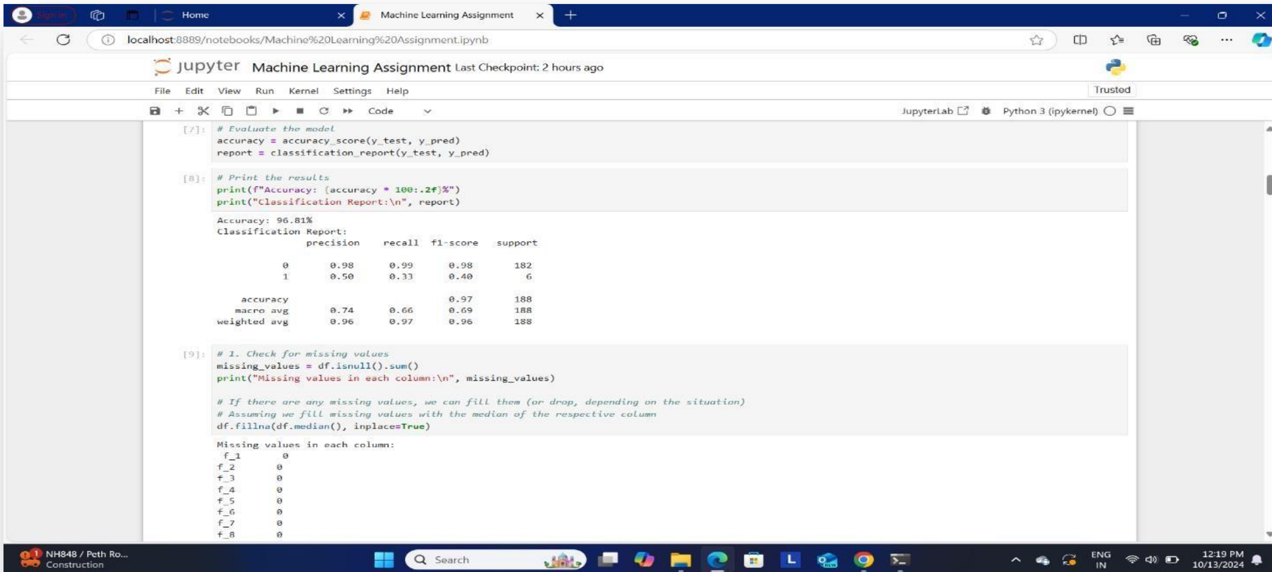


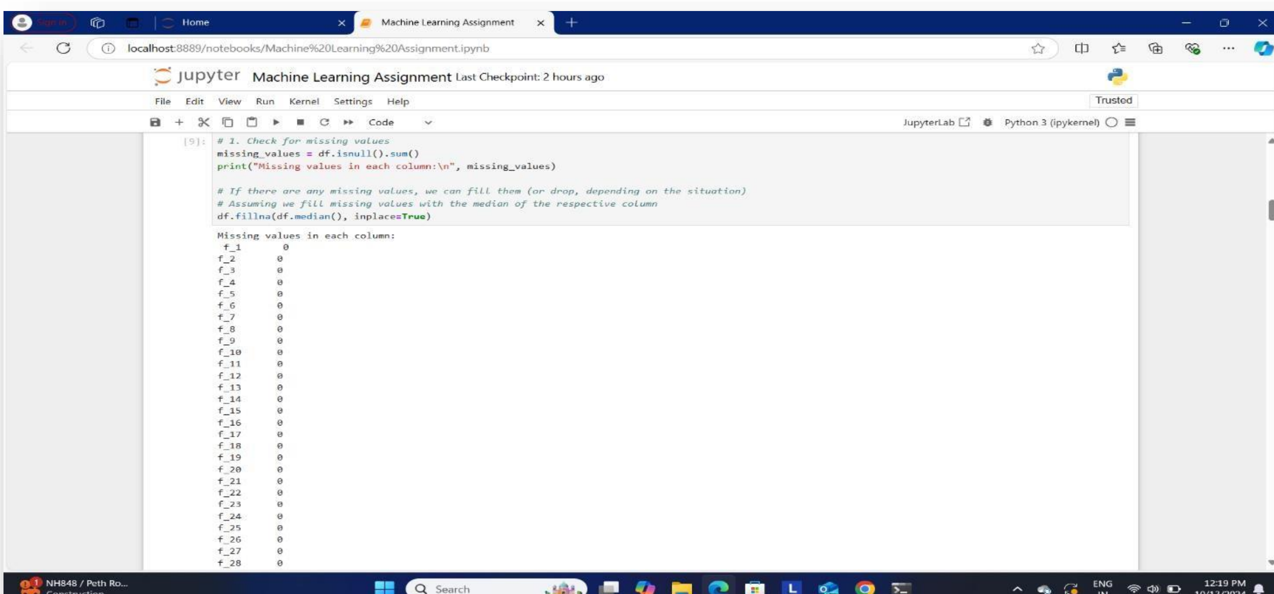


Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

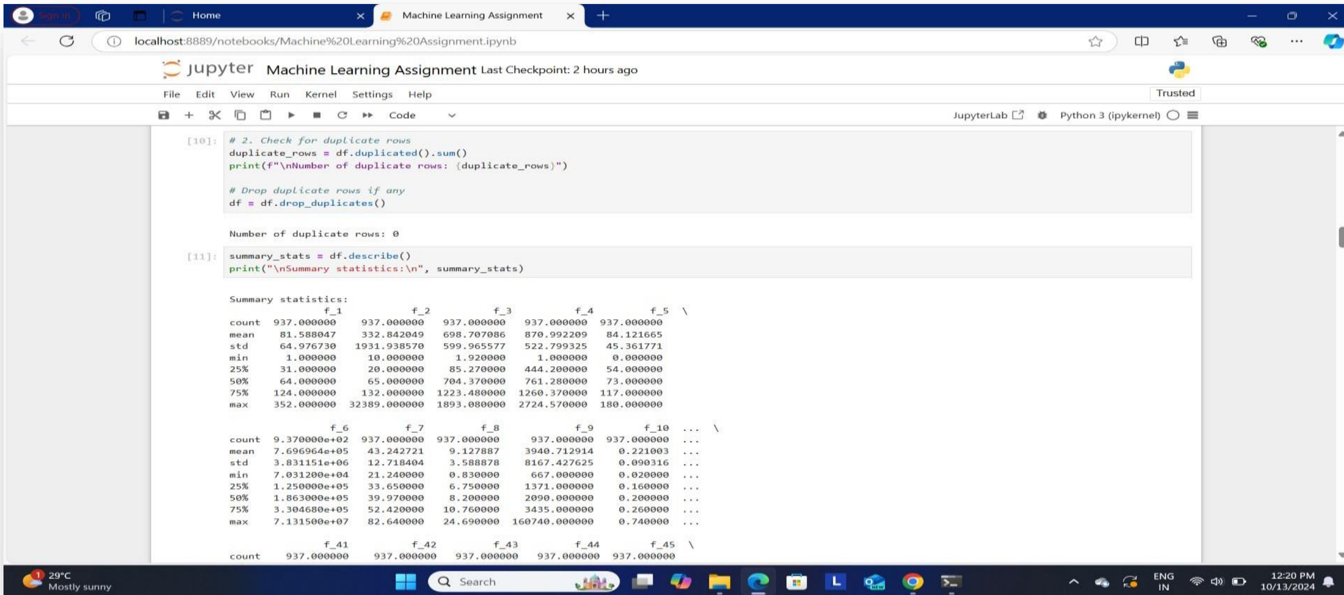




Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

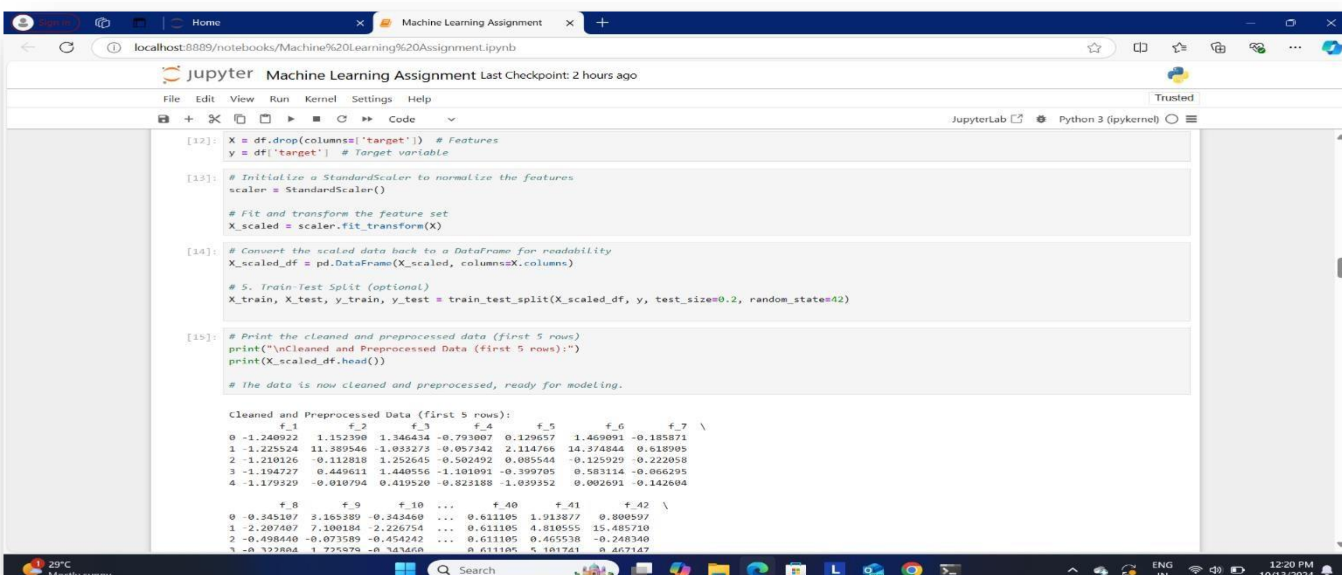


```
[10]: # 2. Check for duplicate rows
      duplicate_rows = df.duplicated().sum()
      print(f"\nNumber of duplicate rows: {duplicate_rows}")

      # Drop duplicate rows if any
      df = df.drop_duplicates()
```

```
Number of duplicate rows: 0
```

```
[11]: summary_stats = df.describe()
      print("\nSummary statistics:\n", summary_stats)
```

```
Summary statistics:
               f_1           f_2          f_3          f_4         f_5  \
count  937.000000    937.000000   937.000000   937.000000  937.000000
mean    81.588047    332.842049   698.707086   870.992209   84.121665
std     64.976730   1931.938570   599.965577   522.799325   45.361771
min      1.000000     10.000000     1.920000     1.000000    0.000000
25%     31.000000     20.000000    85.270000   444.200000   54.000000
50%     64.000000     65.000000   704.370000   761.280000   73.000000
75%    124.000000    132.000000  1223.480000  1260.370000  117.000000
max    352.000000  32389.000000  1893.080000  2724.570000  180.000000

                f_6         f_7         f_8            f_9        f_10  ...  \
count  9.370000e+02  937.000000  937.000000     937.000000  937.000000  ...
mean   7.696964e+05   43.242721    9.127887    3940.712914    0.221003  ...
std    3.831151e+06   12.718404    3.588878    8167.427625    0.090316  ...
min    7.031200e+04   21.240000    0.830000     667.000000    0.020000  ...
25%    1.250000e+05   33.650000    6.750000    1371.000000    0.160000  ...
50%    1.863000e+05   39.970000    8.200000    2090.000000    0.200000  ...
75%    3.304680e+05   52.420000   10.760000    3435.000000    0.260000  ...
max    7.131500e+07   82.640000   24.690000  160740.000000    0.740000  ...

              f_41         f_42         f_43         f_44        f_45  \
count   937.000000   937.000000   937.000000   937.000000  937.000000
```



```
[12]: X = df.drop(columns=['target'])  # Features
      y = df['target']  # Target variable
```

```
[13]: # Initialize a StandardScaler to normalize the features
      scaler = StandardScaler()

      # Fit and transform the feature set
      X_scaled = scaler.fit_transform(X)
```

```
[14]: # Convert the scaled data back to a DataFrame for readability
      X_scaled_df = pd.DataFrame(X_scaled, columns=X.columns)

      # 5. Train-Test Split (optional)
      X_train, X_test, y_train, y_test = train_test_split(X_scaled_df, y, test_size=0.2, random_state=42)
```

```
[15]: # Print the cleaned and preprocessed data (first 5 rows)
      print("\nCleaned and Preprocessed Data (first 5 rows):")
      print(X_scaled_df.head())

      # The data is now cleaned and preprocessed, ready for modeling.
```

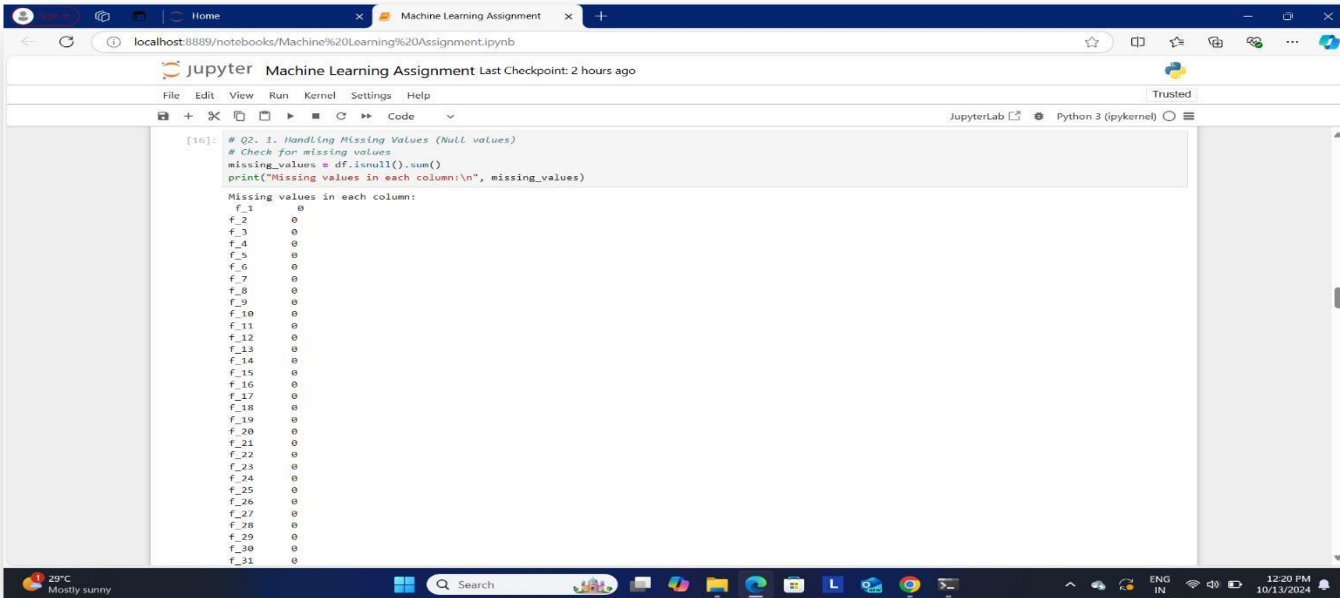
```
Cleaned and Preprocessed Data (first 5 rows):
        f_1        f_2       f_3       f_4       f_5        f_6       f_7  \
0 -1.240922   1.152390  1.346434 -0.793007  0.129657   1.469091 -0.185871
1 -1.225524  11.389546 -1.033273 -0.057342  2.114766  14.374844  0.618905
2 -1.210126  -0.112818  1.252645 -0.502492  0.085544  -0.125929 -0.222058
3 -1.194727   0.449611  1.440556 -1.101091 -0.399705   0.583114 -0.066295
4 -1.179329  -0.010794  0.419520 -0.823188 -1.039352   0.002691 -0.142604

        f_8       f_9      f_10  ...      f_40      f_41       f_42  \
0 -0.345107  3.165389 -0.343460  ...  0.611105  1.913877   0.800597
1 -2.207407  7.100184 -2.226754  ...  0.611105  4.810555  15.485710
2 -0.498440 -0.073589 -0.454242  ...  0.611105  0.465538  -0.248340
3 -0.322804  1.725979 -0.343460       0.611105  5.101741   0.467147
```

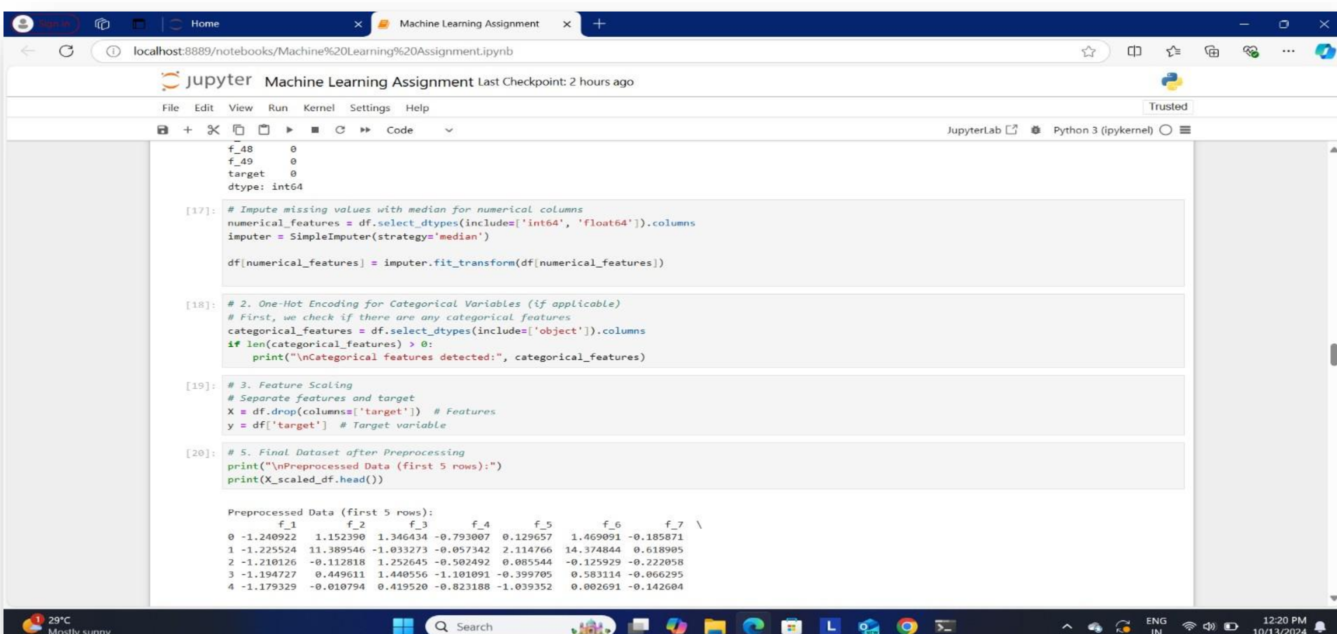
Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Q2) Use various methods such as Handling null values, One-Hot Encoding, Imputation, and Scaling of Data Pre-Processing where necessary?



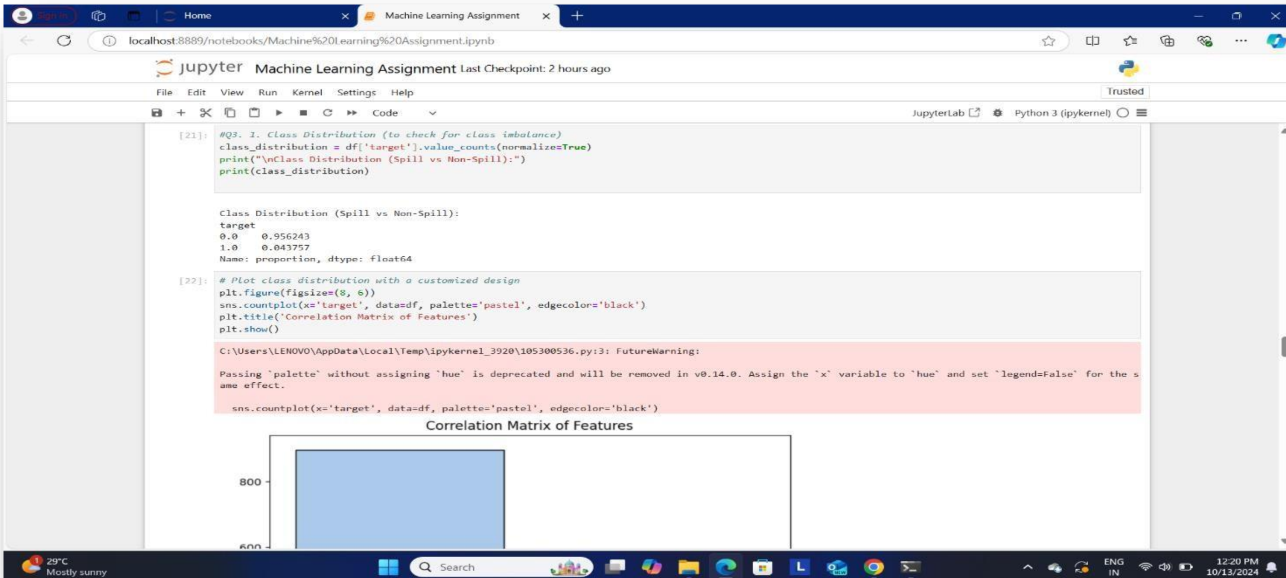


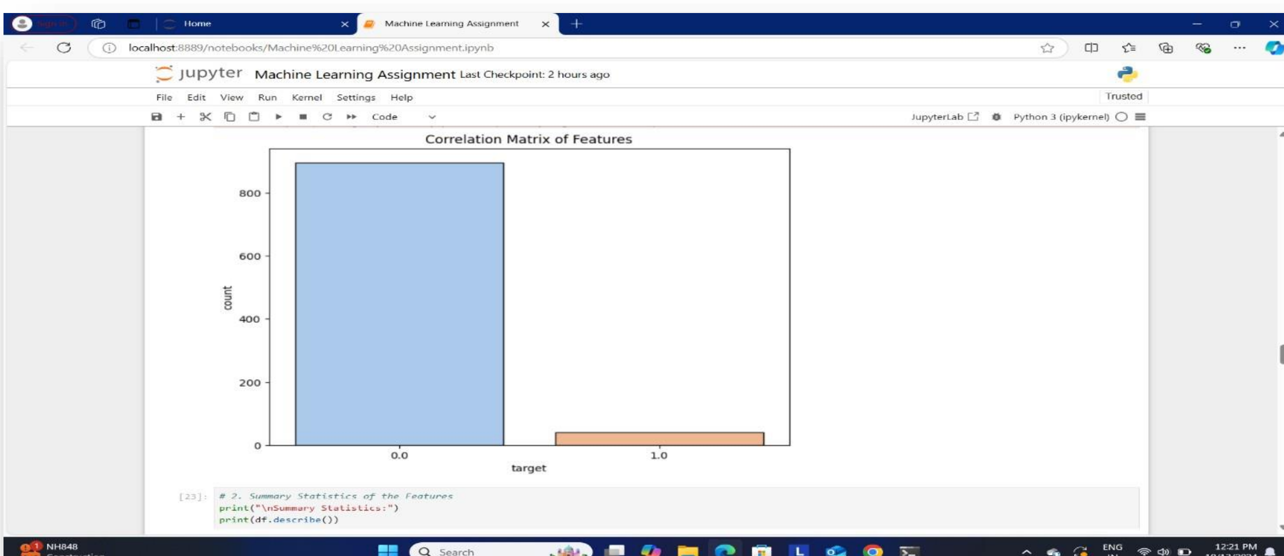
Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Q3) Derive some insights from the dataset ?

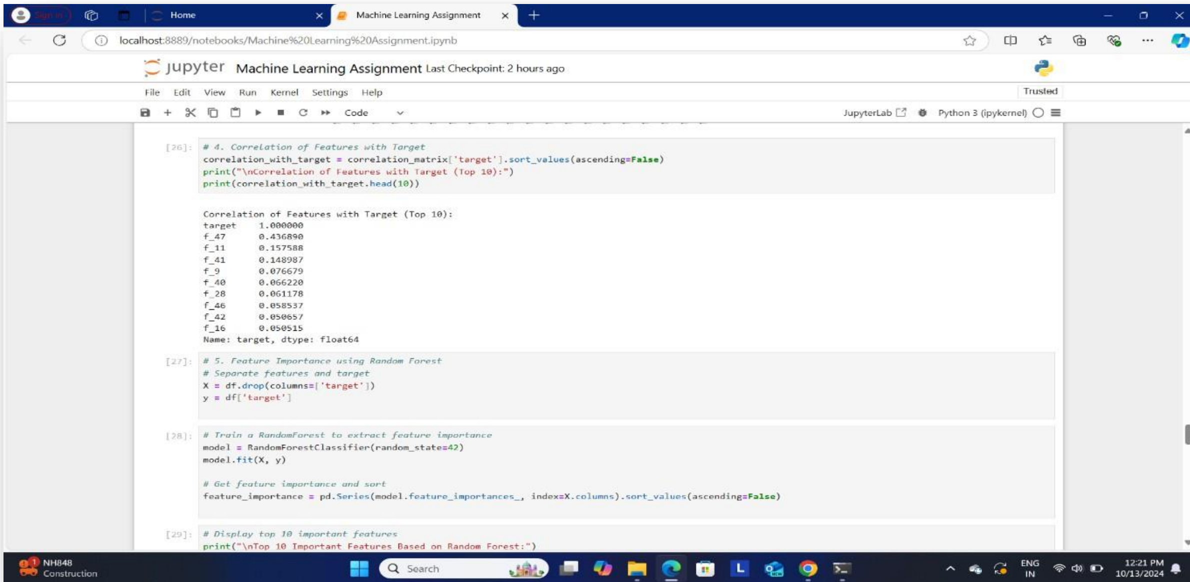


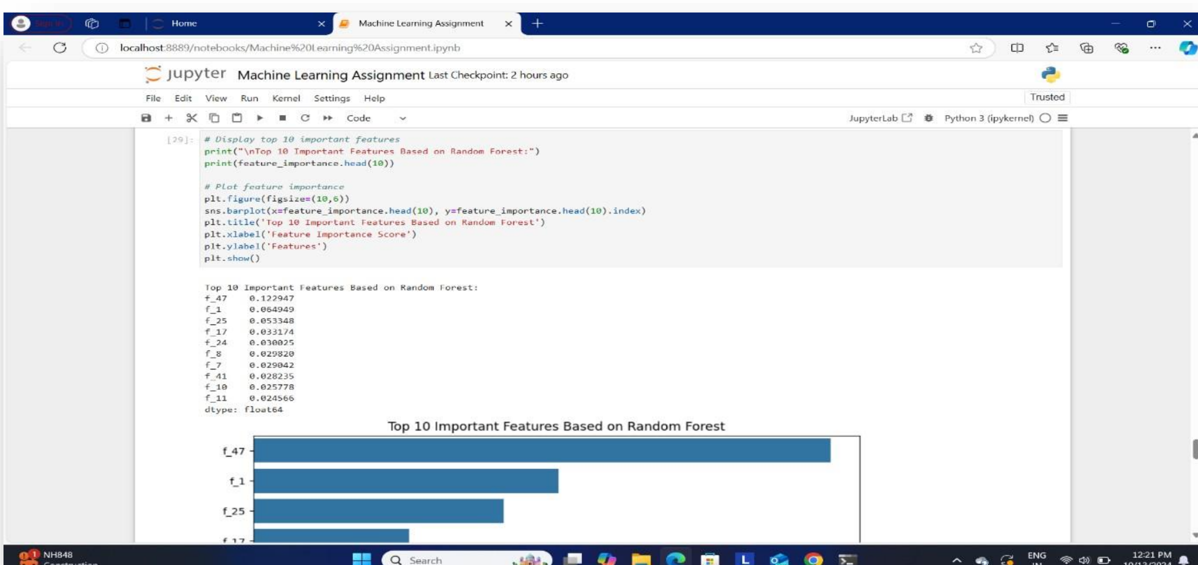


Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

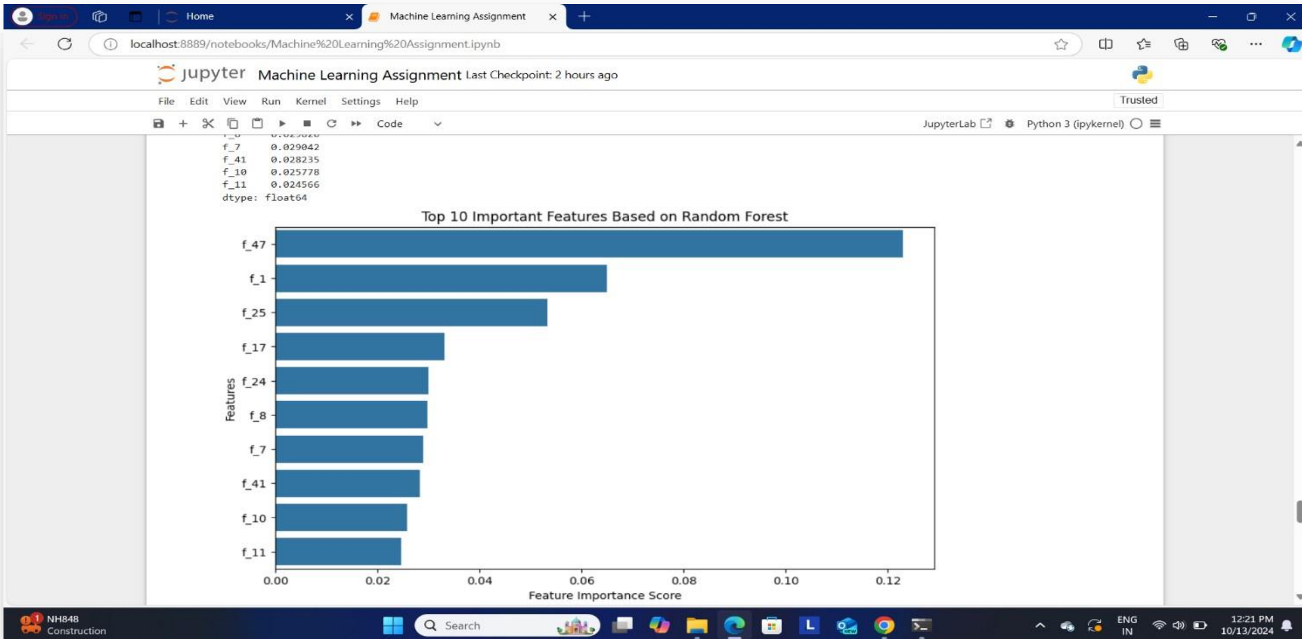




Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

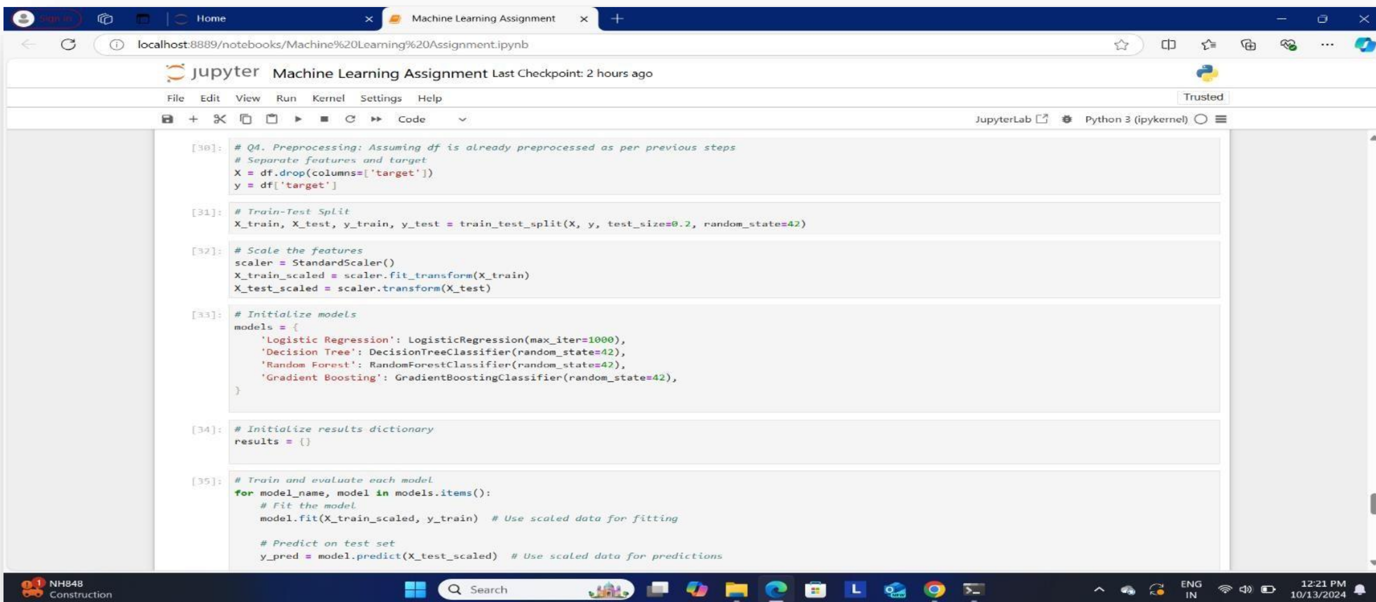


Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Q4) Apply various Machine Learning techniques to predict the output in the target column, make use of Bagging and Ensemble as required, and find the best model by evaluating the model using Model evaluation techniques ?

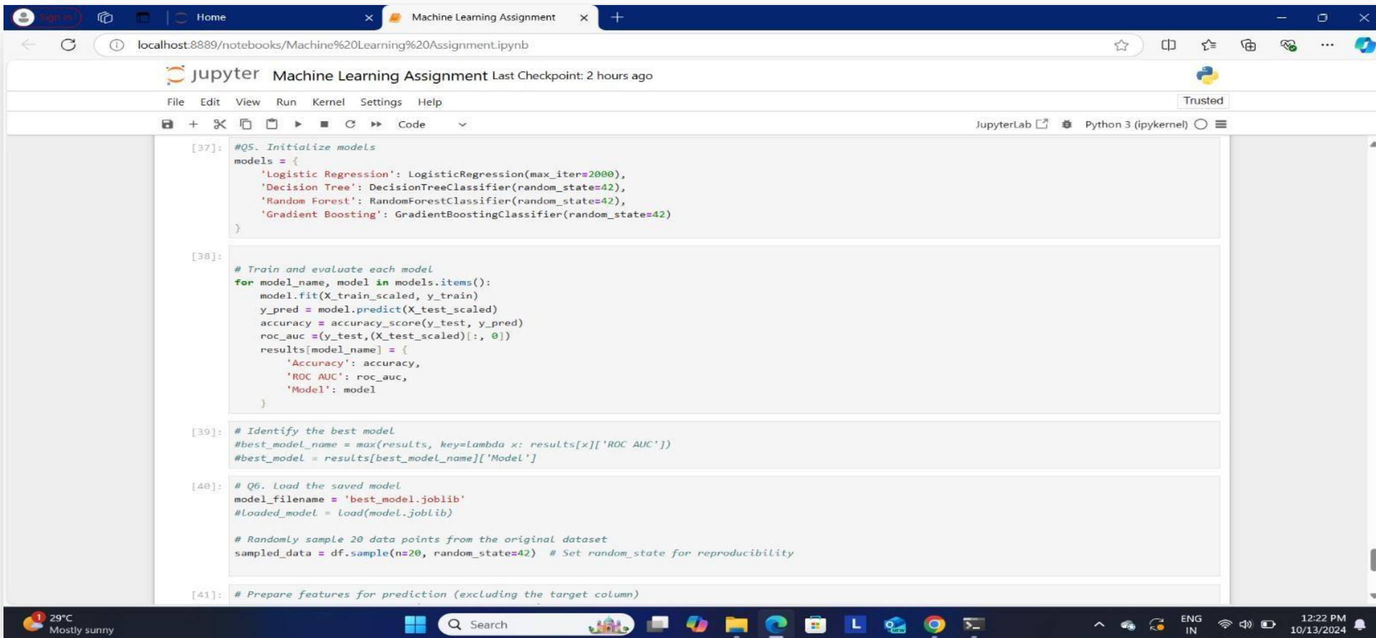


Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Q5) Save the best model and Load the model ?

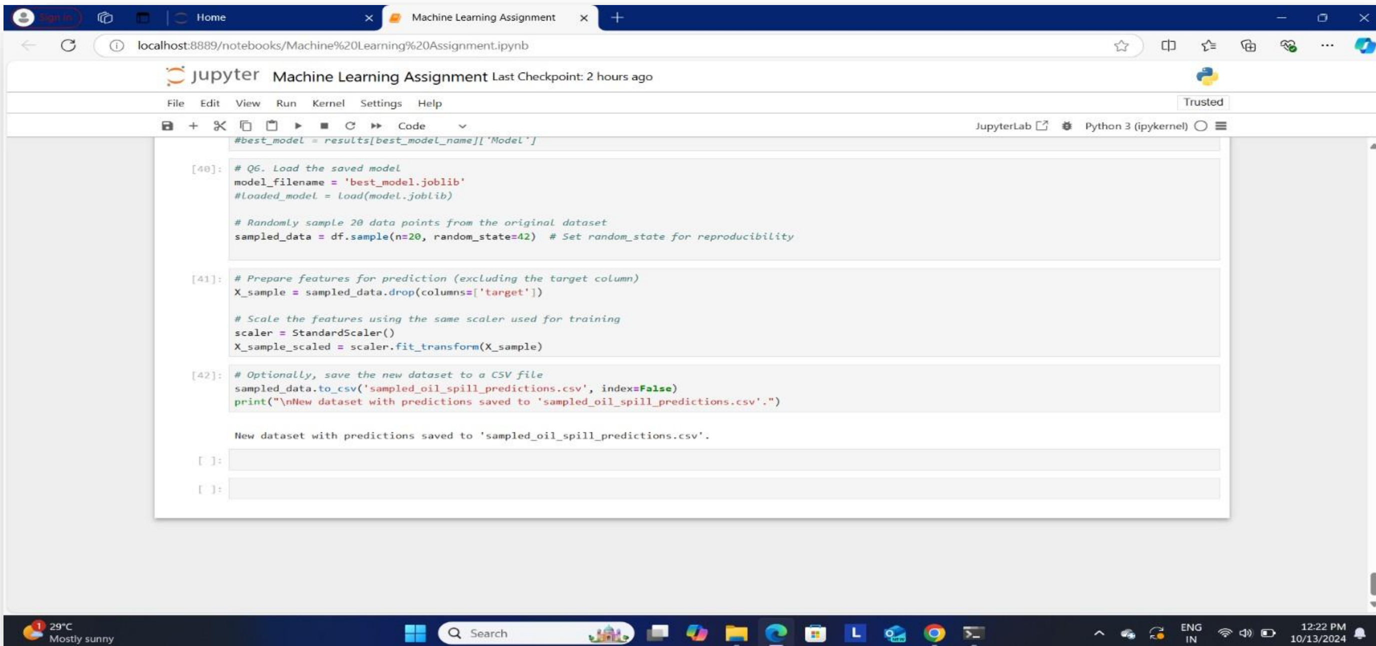


Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Q6) Take the original data set and make another dataset by randomly picking 20 data points from the oil spill dataset and applying the saved model to the same ?



Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com

Name:-SHRIKANT VIJAY TAWADE

Course:- Data Science Machine Learning Assignment (Major)

Email:- shrikanttawade17@gmail.com