Sanket Gore

Internship ID : UMIP27591

```
In [2]:  # Import necessary libraries
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt

         from sklearn.model_selection import train_test_split
         from sklearn.preprocessing import StandardScaler

         from sklearn.metrics import accuracy_score, classification_report
```

```
In [3]:  # Load the dataset
         df = pd.read_csv(r"train.csv")
```

```
In [4]:  df.shape
```

Out[4]: (15120, 56)

EDA

```
In [5]:  df.head(5)
```

Out[5]:

	Id	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrolo
0	1	2596	51	3	258	
1	2	2590	56	2	212	
2	3	2804	139	9	268	
3	4	2785	155	18	242	
4	5	2595	45	2	153	

5 rows × 56 columns

```
In [6]:  df['Cover_Type'].unique()
```

Out[6]: array([5, 2, 1, 7, 3, 6, 4], dtype=int64)

```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15120 entries, 0 to 15119
Data columns (total 56 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Id                                  15120 non-null  int64
 1   Elevation                           15120 non-null  int64
 2   Aspect                              15120 non-null  int64
 3   Slope                               15120 non-null  int64
 4   Horizontal_Distance_To_Hydrology    15120 non-null  int64
 5   Vertical_Distance_To_Hydrology      15120 non-null  int64
 6   Horizontal_Distance_To_Roadways     15120 non-null  int64
 7   Hillshade_9am                       15120 non-null  int64
 8   Hillshade_Noon                      15120 non-null  int64
 9   Hillshade_3pm                       15120 non-null  int64
 10  Horizontal_Distance_To_Fire_Points  15120 non-null  int64
 11  Wilderness_Area1                    15120 non-null  int64
 12  Wilderness_Area2                    15120 non-null  int64
 13  Wilderness_Area3                    15120 non-null  int64
 14  Wilderness_Area4                    15120 non-null  int64
 15  Soil_Type1                          15120 non-null  int64
 16  Soil_Type2                          15120 non-null  int64
 17  Soil_Type3                          15120 non-null  int64
 18  Soil_Type4                          15120 non-null  int64
 19  Soil_Type5                          15120 non-null  int64
 20  Soil_Type6                          15120 non-null  int64
 21  Soil_Type7                          15120 non-null  int64
 22  Soil_Type8                          15120 non-null  int64
 23  Soil_Type9                          15120 non-null  int64
 24  Soil_Type10                         15120 non-null  int64
 25  Soil_Type11                         15120 non-null  int64
 26  Soil_Type12                         15120 non-null  int64
 27  Soil_Type13                         15120 non-null  int64
 28  Soil_Type14                         15120 non-null  int64
 29  Soil_Type15                         15120 non-null  int64
 30  Soil_Type16                         15120 non-null  int64
 31  Soil_Type17                         15120 non-null  int64
 32  Soil_Type18                         15120 non-null  int64
 33  Soil_Type19                         15120 non-null  int64
 34  Soil_Type20                         15120 non-null  int64
 35  Soil_Type21                         15120 non-null  int64
 36  Soil_Type22                         15120 non-null  int64
 37  Soil_Type23                         15120 non-null  int64
 38  Soil_Type24                         15120 non-null  int64
 39  Soil_Type25                         15120 non-null  int64
 40  Soil_Type26                         15120 non-null  int64
 41  Soil_Type27                         15120 non-null  int64
 42  Soil_Type28                         15120 non-null  int64
 43  Soil_Type29                         15120 non-null  int64
 44  Soil_Type30                         15120 non-null  int64
 45  Soil_Type31                         15120 non-null  int64
 46  Soil_Type32                         15120 non-null  int64
 47  Soil_Type33                         15120 non-null  int64
 48  Soil_Type34                         15120 non-null  int64
 49  Soil_Type35                         15120 non-null  int64
 50  Soil_Type36                         15120 non-null  int64
 51  Soil_Type37                         15120 non-null  int64
 52  Soil_Type38                         15120 non-null  int64
 53  Soil_Type39                         15120 non-null  int64
 54  Soil_Type40                         15120 non-null  int64
 55  Cover_Type                          15120 non-null  int64
```

```
dtypes: int64(56)
memory usage: 6.5 MB
```

In [8]: `df.describe()`

Out[8]:

	Id	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrolog
count	15120.00000	15120.000000	15120.000000	15120.000000	15120.00000
mean	7560.50000	2749.322553	156.676653	16.501587	227.19570
std	4364.91237	417.678187	110.085801	8.453927	210.07529
min	1.00000	1863.000000	0.000000	0.000000	0.00000
25%	3780.75000	2376.000000	65.000000	10.000000	67.00000
50%	7560.50000	2752.000000	126.000000	15.000000	180.00000
75%	11340.25000	3104.000000	261.000000	22.000000	330.00000
max	15120.00000	3849.000000	360.000000	52.000000	1343.00000

8 rows × 56 columns

```
In [9]: df.isnull().sum()
```

```
Out[9]: Id                                    0
        Elevation                             0
        Aspect                                0
        Slope                                 0
        Horizontal_Distance_To_Hydrology      0
        Vertical_Distance_To_Hydrology        0
        Horizontal_Distance_To_Roadways       0
        Hillshade_9am                         0
        Hillshade_Noon                        0
        Hillshade_3pm                         0
        Horizontal_Distance_To_Fire_Points    0
        Wilderness_Area1                      0
        Wilderness_Area2                      0
        Wilderness_Area3                      0
        Wilderness_Area4                      0
        Soil_Type1                            0
        Soil_Type2                            0
        Soil_Type3                            0
        Soil_Type4                            0
        Soil_Type5                            0
        Soil_Type6                            0
        Soil_Type7                            0
        Soil_Type8                            0
        Soil_Type9                            0
        Soil_Type10                           0
        Soil_Type11                           0
        Soil_Type12                           0
        Soil_Type13                           0
        Soil_Type14                           0
        Soil_Type15                           0
        Soil_Type16                           0
        Soil_Type17                           0
        Soil_Type18                           0
        Soil_Type19                           0
        Soil_Type20                           0
        Soil_Type21                           0
        Soil_Type22                           0
        Soil_Type23                           0
        Soil_Type24                           0
        Soil_Type25                           0
        Soil_Type26                           0
        Soil_Type27                           0
        Soil_Type28                           0
        Soil_Type29                           0
        Soil_Type30                           0
        Soil_Type31                           0
        Soil_Type32                           0
        Soil_Type33                           0
        Soil_Type34                           0
        Soil_Type35                           0
        Soil_Type36                           0
        Soil_Type37                           0
        Soil_Type38                           0
        Soil_Type39                           0
        Soil_Type40                           0
        Cover_Type                            0
        dtype: int64
```

```
In [10]: X = df.drop(columns = ["Cover_Type"])
```

```
In [11]: X.shape
```
Out[11]: (15120, 55)

```
In [12]: Y = df["Cover_Type"]
```

```
In [13]: Y.shape
```
Out[13]: (15120,)

```
In [14]: scaler = StandardScaler()
```

Train-test split

```
In [15]: x_train,x_test,y_train,y_test = train_test_split(X,Y, test_size = 0.2, random_
         state = 42)
```

```
In [16]: x_train.shape
```
Out[16]: (12096, 55)

```
In [17]: x_train.head(5)
```
Out[17]:

	Id	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To
4679	4680	2364	3	30	404	
6634	6635	2883	62	10	365	
11459	11460	2972	162	15	108	
15053	15054	2395	349	32	95	
8222	8223	3244	64	21	108	

5 rows × 55 columns

```
In [18]: x_train_std = scaler.fit_transform(x_train)
```

```
In [19]: print(x_train_std)
         [[-0.66027127 -0.93192165 -1.39353286 ... -0.22401411 -0.21279501
           -0.17614271]
          [-0.21218549  0.31796942 -0.8590903  ... -0.22401411 -0.21279501
           -0.17614271]
          [ 0.89370397  0.53230526  0.04674455 ... -0.22401411 -0.21279501
           -0.17614271]
          ...
          [-0.49731015  0.50822258 -0.03478058 ... -0.22401411 -0.21279501
           -0.17614271]
          [-1.53558564  0.35409343  0.88011261 ... -0.22401411 -0.21279501
           -0.17614271]
          [-0.06641436  1.50042898 -0.71415672 ...  4.46400444 -0.21279501
           -0.17614271]]
```

```
In [20]: x_test_std = scaler.transform(x_test)
```

```
In [21]: from sklearn.ensemble import RandomForestClassifier
```

```
In [22]: rf_model = RandomForestClassifier(n_estimators = 100, random_state = 42)
```

```
In [23]: rf_model.fit(x_train_std, y_train)
```

Out[23]:
```
  ▼        RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
In [24]: y_pred = rf_model.predict(x_test_std)
```

```
In [25]: accuracy = accuracy_score(y_test, y_pred)
```

```
In [26]: print("Accuracy = ", accuracy)

Accuracy =  0.875
```

```
In [27]: report = classification_report(y_test, y_pred)
```

```
In [28]: print("Classification Report :", report)
```

```
Classification Report :                 precision    recall  f1-score   support

                    1        0.78      0.78      0.78       421
                    2        0.81      0.69      0.75       438
                    3        0.87      0.85      0.86       428
                    4        0.93      0.98      0.96       449
                    5        0.89      0.95      0.92       416
                    6        0.87      0.90      0.88       432
                    7        0.95      0.97      0.96       440

            accuracy                            0.88      3024
           macro avg        0.87      0.87      0.87      3024
        weighted avg        0.87      0.88      0.87      3024
```

```
In [ ]:
```