Modelos Y Simulación De Sistemas 1

Proyecto De Semestre - Entrega 2

UNIVERSIDAD
DE ANTIOQUIA
1 8 0 3

Responsables

Eliana Janneth Puerta Morales

Valentina Muñoz Rincón

Juan Fernando Lopera Muñoz

Medellín - Antioquia 2023

Preprocesamiento del DataSet

Antes de avanzar con el conjunto de datos, lo primero que se hizo fue asegurar que se cumplieran los requisitos establecidos. Uno de estos requisitos no se cumplió, ya que el conjunto de datos no tenía ningún valor faltante. Lo que se necesitaba era introducir intencionadamente valores nulos en ese conjunto de datos para poder comenzar con el proceso de entrenamiento.

```
for col in columns_to_simulate:
  for k in range(0,900):
    random_num = np.random.randint(0,high=15120)
    df_used.loc[random_num,col] = np.nan
```

Además de la generación de datos, se reconstruyó la información para el proceso de entrenamiento. El dataset consta de 56 columnas, de las cuales 1 es inmutable debido a ser un identificador único (ID), 10 son de tipo entero y 45 son de tipo booleano, es decir, son categóricas con valores que únicamente pueden ser 0 o 1.

Para abordar esta tarea, se usó la librería “numpy” para obtener el listado de las medias de cada una de las columnas sin sus valores nulos, luego se implementó un bucle para rellenar los valores nulos con el promedio obtenido de las columnas sin valores nulos.

```
means_list = np.mean(df_without_null[columns_to_simulate], axis=0)
```

```
for column in columns_to_simulate:
  df_used[column].fillna(means_list[column], inplace=True)
```

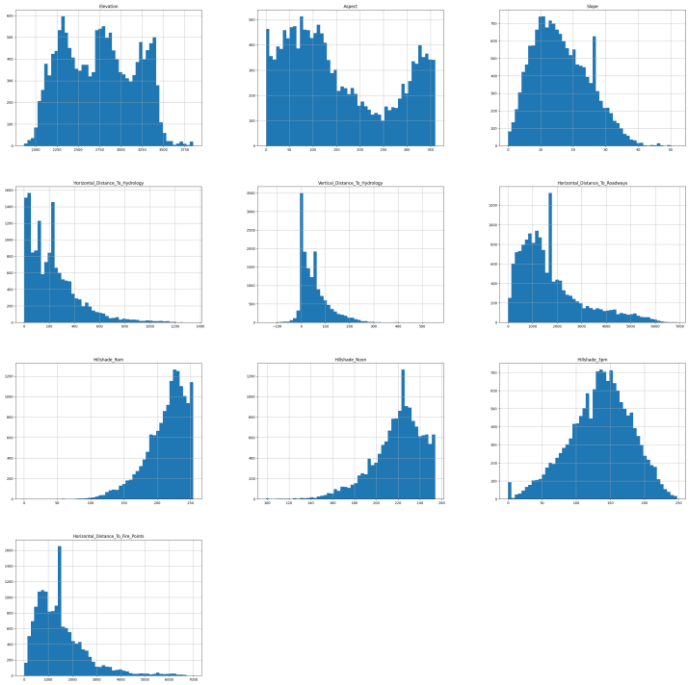
Análisis de Datos

Después de completar la tarea de llenar los valores nulos en el dataframe, se procedió a analizar la información presentada. Se evidenció que ya no existían datos nulos en la lista.

Como se mencionó anteriormente, las columnas desde la 11 hasta la 55 son variables categóricas:

```
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  float6
5   Vertical_Distance_To_Hydrology      15120 non-null  float6
6   Horizontal_Distance_To_Roadways     15120 non-null  float6
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  float6
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
```

Además, se realizó un histograma que representaba la frecuencia con la que los datos se mostraban en pequeños grupos de 50.



Se realizó una función que mostraba la relación de todas las columnas con respecto al Cover_Type:

```
abs(df_used.corr()['Cover_Type'])
```

```
Soil_Type24                           0.100797
Soil_Type25                           0.008133
Soil_Type26                           0.017184
Soil_Type27                           0.023109
Soil_Type28                           0.012202
Soil_Type29                           0.218564
Soil_Type30                           0.001393
Soil_Type31                           0.079882
Soil_Type32                           0.132312
Soil_Type33                           0.078955
Soil_Type34                           0.003470
Soil_Type35                           0.114327
Soil_Type36                           0.025726
Soil_Type37                           0.071210
Soil_Type38                           0.257810
Soil_Type39                           0.240384
Soil_Type40                           0.205851
Cover_Type                            1.000000
```

En las filas de "Soil_Type7" y "Soil_Type15" están con nulo ya que no existe una correlación entre el tipo de cobertura (Cover_Type) y el tipo de suelo de esas categorías.

Se puede observar que el tipo de suelo 38 y 39 son los que más correlación tienen con respecto al tipo de cobertura.

```
Id                                    0.108363
Elevation                             0.016090
Aspect                                0.008015
Slope                                 0.087722
Horizontal_Distance_To_Hydrology      0.010515
Vertical_Distance_To_Hydrology        0.075647
Horizontal_Distance_To_Roadways       0.105662
Hillshade_9am                         0.010286
Hillshade_Noon                        0.098905
Hillshade_3pm                         0.053399
Horizontal_Distance_To_Fire_Points    0.089389
Wilderness_Area1                      0.230117
Wilderness_Area2                      0.014994
Wilderness_Area3                      0.122146
Wilderness_Area4                      0.075774
Soil_Type1                            0.015069
Soil_Type2                            0.022627
Soil_Type3                            0.016393
Soil_Type4                            0.027816
Soil_Type5                            0.027692
Soil_Type6                            0.006521
Soil_Type7                                 NaN
Soil_Type8                            0.008133
Soil_Type9                            0.027012
Soil_Type10                           0.128972
Soil_Type11                           0.010228
Soil_Type12                           0.129985
Soil_Type13                           0.040528
Soil_Type14                           0.022019
Soil_Type15                                NaN
Soil_Type16                           0.008793
Soil_Type17                           0.042453
Soil_Type18                           0.006312
Soil_Type19                           0.031824
Soil_Type20                           0.053013
Soil_Type21                           0.024410
Soil_Type22                           0.195993
Soil_Type23                           0.158762
```