|  |  |
| --- | --- |
| Name Of The Student | Himanshu |
| Internship Project Topic | TCS iON RIO-210: Build a Classification Model for Drug Trials Dataset |
| Name of the Organization | TCS iON |
| Name of the Industry Mentor | Himdweep Walia |
| Name of the Institute | Amity University |

|  |  |  |
| --- | --- | --- |
| Date | Day # | Hours Spent |
| 02-06-2024 | Day-40 | 6 Hours |
| Activities done during the day:  **Project Hands-on – Preprocessing the data**  **Link of the google drive google Colab file :-**  <https://colab.research.google.com/drive/1VQRq0l6oc9Uj4cOOqiuhkfS1JmpKr3fU?usp=sharing>  Data preprocessing is a crucial step in the data analysis pipeline where raw data is transformed into a format that is suitable for analysis and modeling. Here are some common techniques used in data preprocessing:  **Handling missing values**  **Data cleaning**  **Feature scaling**  **Feature encoding**  **Feature selection**  **Data transformation**  **Normalization**  **Data splitting**  data = data.drop('Age Range', axis=1)  # Filter out ages outside the range of 100  filtered\_data = data[data['AGE'] > 90]  print(len(filtered\_data))  # Print the output  filtered\_data.head(20)     Define a function to process the age values  def process\_age(age):      if age > 100:          return int(age / 10)      return age  # Apply the process\_age function to the 'AGE' column in data  data['AGE'] = data['AGE'].apply(process\_age)  # Print the corrected DataFrame  data.head()    # Filter out ages outside the range of 100  filtered\_data = data[data['AGE'] > 90]  print(len(filtered\_data))  # Calculate the mean age  mean\_age = data['AGE'].mean()  # Replace null values with the mean age  data['AGE'].fillna(mean\_age, inplace=True)  data.isnull().sum() | | |
|  | | |