|  |  |
| --- | --- |
| 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 |
| 03-05-2024 | Day-13 | 3.5 Hours |
| Activities done during the day:  Learned about data transformation process in Machine Learning.  What is data transformation ?  Data transformation in data mining refers to the process of converting raw data into a format that is suitable for analysis and modeling. The goal of data transformation is to prepare the data for data mining so that it can be used to extract useful insights and knowledge. Data transformation typically involves several steps  Stages in Data Transformation:   * **Data cleaning:** Removing or correcting errors, inconsistencies, and missing values in the data. * **Data integration:** Combining data from multiple sources, such as databases and spreadsheets, into a single format. * **Data normalization:** Scaling the data to a common range of values, such as between 0 and 1, to facilitate comparison and analysis. * **Data reduction:** Reducing the dimensionality of the data by selecting a subset of relevant features or attributes. * **Data discretization:** Converting continuous data into discrete categories or bins. * Data aggregation: Combining data at different levels of granularity, such as by summing or averaging, to create new features or attributes. * Data transformation is an important step in the data mining process as it helps to ensure that the data is in a format that is suitable for analysis and modeling, and that it is free of errors and inconsistencies. Data transformation can also help to improve the performance of data mining algorithms, by reducing the dimensionality of the data, and by scaling the data to a common range of values. | | |
|  | | |