|  |  |
| --- | --- |
| 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 |
| 30-04-2024 | Day-10 | 3.5 Hours |
| Activities done during the day:  learned about the Types of Classification Algorithms in detail  Learned about data pre-processing steps for Machine Learning & Data analytics.  **Why data processing is Important and what is the use of data preprocessing ?**   * Data in the real world is quite dirty and corrupted with inconsistencies, noise, incomplete information, and missing values. * The majority of the real-world datasets for machine learning are highly susceptible to be missing, inconsistent, and noisy due to their heterogeneous origin. * Applying data mining algorithms on this noisy data would not give quality results as they would fail to identify patterns effectively. Data Processing is, therefore, important to improve the overall data quality. * Quality decisions must be based on quality data. Data Preprocessing is important to get this quality data, without which it would just be a Garbage In, Garbage Out scenario.   **How data processing can be done ?**   * Data preprocessing is a step in the data mining and data analysis process that takes raw data and transforms it into a format that can be understood and analyzed by computers and machine learning.   Steps in data processing:-   1. Data quality assessment 2. Data cleaning 3. Data transformation 4. Data reduction   Learned in detail about the 4 data processing steps in the below reference link.  <https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing>  <https://www.sciencedirect.com/topics/engineering/data-preprocessing#:~:text=Data%20preprocessing%2Fpreparation%2Fcleaning%20is,or%20coarse%20data%20%5B6%5D>. | | |
|  | | |