|  |  |
| --- | --- |
| Name Of The Student | Vaishnavi G |
| Internship Project Topic | 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 | SRM Institute of Science and Technology |

|  |  |  |
| --- | --- | --- |
| Date | Day # | Hours Spent |
| 22/11/2022 | 42 | 5 hours |
| Activities done during the day:  **Project Hands on - Transform Scale data**  **Transform Scale data:**   * When your data has different values, and even different measurement units, it can be difficult to compare them. * In order for the model to treat each and every column equally and fairly it is important to transform or scale the data in each column to a similar range. This transformation or scaling of data is also known as Feature Scaling. * some scaling techniques that can be used to scale numeric features. * MinMax Scaler * Standard Scaler * MaxAbs Scaler * Robust Scaler * Quantile Transform Scaler * Power Transformer Scaler * Unit Vector Scaler/Normalizer * All these scalers can be imported from the sklearn.preprocessing Python library. It is an open-source library that is used for processing the data for machine learning and Data Science.  |  | | --- | | cols = ['urlDrugName', 'effectiveness', 'condition']  for x in cols:  df[x] = pd.factorize(df[x])[0]  target = pd.factorize(target)[0] |   **df** - A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns.  **target -** The target variable of a dataset is the feature of a dataset about which you want to gain a deeper understanding. A supervised machine learning algorithm uses historical data to learn patterns and uncover relationships between other features of your dataset and the target.  **pandas.factorize()** method helps to get the numeric representation of an array by identifying distinct values. This method is available as both pandas.factorize() and Series.factorize().  **head() -** The head function in Python displays the first five rows of the dataframe by default. It takes in a single parameter. the number of rows. We can use this parameter to display the number of rows of our choice.  **Code:**   |  | | --- | |  |   **Output:**   |  | | --- | |  | | | |