|  |  |
| --- | --- |
| 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 |
| 8/11/2022 | 28 | 5 hours |
| Activities done during the day:  **Random Forest classifier in Machine learning:**  **What is random forest?**  Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.  **Why use the Random Forest algorithm?**     * It can be used for both classifications as well as regression tasks. * Overfitting problem that is censorious and can make results poor but in case of the random forest the classifier will not overfit if there are enough trees. * It can handle missing values. * It can be used for categorical values as well.   **Random forest Vs Decision Tree**    Random forest is nothing but a set of many decision trees. Decision trees are faster. Extensive decision trees might get troubled by overfitting, but random forest prevents that by generating more trees on random subsets. Random forests are complex and not easy to explain but decision trees are easy and can be converted to certain rules.  **Applications of Random Forest**  There are mainly four sectors where Random forest mostly used:  **Banking:** Banking sector mostly uses this algorithm for the identification of loan risk.  **Medicine:** With the help of this algorithm, disease trends and risks of the disease can be identified.  **Land Use:** We can identify the areas of similar land use by this algorithm.  **Marketing:** Marketing trends can be identified using this algorithm.  **What are the advantages and disadvantages of the Random forest algorithm?**    **Advantages:**   * It overcomes the problem of overfitting. * It is fast and can deal with missing values data as well. * It is flexible and gives high accuracy. * Can be used for both classifications as well as regression tasks. * Using random forest you can compute the relative feature importance. * It can give good accuracy even if the higher volume of data is missing.   **Disadvantages:**     * Random forest is a complex algorithm that is not easy to interpret. * Complexity is large. * Predictions given by random forest takes many times if we compare it to other algorithms * Higher computational resources are required to use a random forest algorithm.   **How does Random Forest algorithm work?**  Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.  The Working process can be explained in the below steps and diagram:  Step-1: Select random K data points from the training set.  Step-2: Build the decision trees associated with the selected data points (Subsets).  Step-3: Choose the number N for decision trees that you want to build.  Step-4: Repeat Step 1 & 2.  Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.  Reference:  <https://www.javatpoint.com/machine-learning-random-forest-algorithm/>  <https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm/> | | |