|  |  |
| --- | --- |
| 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 |
| 14/11/2022 | 34 | 5 hours |
| Activities done during the day:  **Underfitting and Overfitting**  **Underfitting:**   * A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data, i.e., it only performs well on training data but performs poorly on testing data. * Underfitting destroys the accuracy of our machine learning model. Its occurrence simply means that our model or the algorithm does not fit the data well enough. * It usually happens when we have fewer data to build an accurate model and also when we try to build a linear model with fewer non-linear data. * In such cases, the rules of the machine learning model are too easy and flexible to be applied to such minimal data and therefore the model will probably make a lot of wrong predictions. * Underfitting can be avoided by using more data and also reducing the features by feature selection. * In a nutshell, Underfitting refers to a model that can neither performs well on the training data nor generalize to new data.   **Reasons for Underfitting:**   * High bias and low variance * The size of the training dataset used is not enough. * The model is too simple. * Training data is not cleaned and also contains noise in it.   **Techniques to reduce underfitting:**   * Increase model complexity * Increase the number of features, performing feature engineering * Remove noise from the data. * Increase the number of epochs or increase the duration of training to get better results.   **Overfitting:**   * A statistical model is said to be overfitted when the model does not make accurate predictions on testing data. * When a model gets trained with so much data, it starts learning from the noise and inaccurate data entries in our data set. * And when testing with test data results in High variance. Then the model does not categorize the data correctly, because of too many details and noise. * The causes of overfitting are the non-parametric and non-linear methods because these types of machine learning algorithms have more freedom in building the model based on the dataset and therefore they can really build unrealistic models. * A solution to avoid overfitting is using a linear algorithm if we have linear data or using the parameters like the maximal depth if we are using decision trees. * In a nutshell, Overfitting is a problem where the evaluation of machine learning algorithms on training data is different from unseen data.   **Reasons for Overfitting are as follows:**   * High variance and low bias * The model is too complex * The size of the training data   Examples:  IMG_256    IMG_256  **Techniques to reduce overfitting:**   * Increase training data. * Reduce model complexity. * Early stopping during the training phase (have an eye over the loss over the training period as soon as loss begins to increase stop training). * Ridge Regularization and Lasso Regularization * Use dropout for neural networks to tackle overfitting.   Reference:  [https://ml-concepts.com/2022/03/04/everything-you-need-to-know-about-model-fitting-in-machine-learning//](https://ml-concepts.com/2022/03/04/everything-you-need-to-know-about-model-fitting-in-machine-learning/) | | |