|  |  |
| --- | --- |
| 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 |
| 14-05-2024 | Day-22 | 5.5 Hours |
| **Learn about the accuracy checking of Machine learning models.**  **Classifiers and there types in Machine learning**  **What is a Classifier?**     * In machine learning, a classifier is an algorithm that automatically sorts or categorizes data into one or more "classes." Targets, labels, and categories are all terms used to describe classes. * One of the most prominent instances is an email classifier, which examines emails and filters them according to whether they are spam or not. * The job of estimating a mapping function (f) from input variables (X) to discrete output variables is known as classification predictive modelling (y). * Machine learning algorithms are useful for automating operations that were previously done by hand. They may save a lot of time and money while also increasing the efficiency of enterprises. * Classification is a type of supervised learning in which the input data is also delivered to the objectives. Classification has several uses in a variety of fields, including credit approval, medical diagnosis, and target marketing. * Machine learning classifiers are used to assess consumer comments from social media, emails, online reviews, and other sources to determine what people are saying about your company.   **Difference between classifier and a model:**     * learn states the difference between a classifier and a model. * A classifier is an algorithm - the principles that robots use to categorize data. * The ultimate product of your classifier's machine learning, on the other hand, is a classification model. The classifier is used to train the model, and the model is then used to classify your data. * Both supervised and unsupervised classifiers are available. * Unsupervised machine learning classifiers are fed just unlabeled datasets, which they sort into categories based on pattern recognition, data structures, and anomalies. * Training datasets are provided to supervised and semi-supervised classifiers, which teach them how to categorize data into specified categories.     **Types of classifiers in Machine learning:**    There are six different classifiers in machine learning,  **Perceptron:**   * For binary classification problems, the Perceptron is a linear machine learning technique. It is one of the original and most basic forms of artificial neural networks. * It isn't "deep" learning, but it is a necessary building component. It is made up of a single node or neuron that processes a row of data and predicts a class label. The weighted total of the inputs and a bias is used to achieve this (set to 1). The activation is defined as the weighted sum of the model's input. * A linear classification algorithm is the Perceptron. This implies it learns a decision boundary in the feature space that divides two classes using a line (called a hyperplane). * As a result, it's best for issues where the classes can be easily separated using a line or linear model, sometimes known as linearly separable problems. The stochastic gradient descent optimization procedure is used to train the model's coefficients, which are referred to as input weights.   **Logistic Regression:**     * Under the Supervised Learning approach, one of the most prominent Machine Learning algorithms is logistic regression. It's a method for predicting a categorical dependent variable from a set of independent factors. * A logistic function is used to describe the probability of the probable outcomes of a single trial in this technique. Logistic regression was created for this goal (classification), and it's especially good for figuring out how numerous independent factors affect a single outcome variable. * Except for how they are employed, Logistic Regression is quite similar to Linear Regression. For regression issues, Linear Regression is employed, whereas, for classification difficulties, Logistic Regression is used. * The algorithm's sole drawback is that it only works when the predicted variable is binary, requires that all predictors are independent of one another, and expects that the data is free of missing values.   **Naive Bayes:**     * The Naive Bayes family of probabilistic algorithms calculates the likelihood that every given data point falls into one or more of a set of categories (or not). It is a supervised learning approach for addressing classification issues that are based on the Bayes theorem. It's a probabilistic classifier, which means it makes predictions based on an object's likelihood. * In the text analysis, Naive Bayes is used to classifying customer comments, news articles, emails, and other types of content into categories, themes, or "tags" in order to organise them according to specified criteria   **K-Nearest Neighbours:**   * K nearest neighbours is a straightforward method that maintains all existing examples and categorizes new ones using a similarity metric (e.g., distance functions). * KNN has been utilized as a non-parametric approach in statistical estimates and pattern recognition since the early 1970s. It's a form of lazy learning since it doesn't try to build a generic internal model; instead, it only saves instances of the training data. The classification is determined by a simple majority vote of each point's k closest neighbours. * A case is categorized by a majority vote of its neighbours, with the case being allocated to the class having the most members among its K closest neighbours as determined by a distance function. If K = 1, the case is simply allocated to the nearest neighbour's class.   **Support Vector Machine:**     * The Support Vector Machine, or SVM, is a common Supervised Learning technique that may be used to solve both classification and regression issues. However, it is mostly utilized in Machine Learning for Classification difficulties. * The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space into classes so that additional data points may be readily placed in the proper category in the future. A hyperplane is a name for the optimal choice boundary. * SVM techniques categorize data and train models within supra limited degrees of polarity, resulting in a three-dimensional classification model that extends beyond X/Y predictive axes. The extreme points/vectors that assist create the hyperplane are chosen via SVM. * Support vectors are extreme instances, and the method is called a Support Vector Machine. Consider the picture below, which shows how a decision boundary or hyperplane is used to classify two separate categories   **Random Forest:**   * Random forest is a supervised learning approach used in machine learning for classification and regression. It's a classifier that averages the results of many decision trees applied to distinct subsets of a dataset to improve the dataset's projected accuracy. * It's also known as a meta-estimator since it fits a number of decision trees on different sub-samples of datasets and utilizes the average to enhance the model's forecast accuracy and prevent over-fitting. The size of the sub-sample is always the same as the size of the original input sample, but the samples are generated using replacement.   It produces a "forest" out of a collection of decision trees that are frequently trained using the "bagging" method. The main idea of the bagging approach is that combining many learning models enhances the final result. Rather than relying on a single decision tree, the random forest gathers forecasts from each tree and predicts the ultimate output based on the majority of votes. | | |