ML-2 Classification

Course Description

The primary role of a data scientist is to predict solution to different real time problems and draw actionable insights. In this module given a problem the solution to problem is predicted into classes. Given an example **classify** if the patient has cancer or not. The model will use training data which the classification algorithms learn patterns from training data and predict with the given testing data into. This course teaches you everything you need to know to begin working with different Classification algorithms.

Course Objectives

The participants will be able to

- Understand the Supervised Classification concepts, Evaluation Parameters
- Classification Algorithm of Logistic Regression and Odd's Probability with performance measures.
- Imbalanced datasets and how to manage using SMOTE and Random under sampling along with Feature Importance.Decision Tree Algorithm using Entropy and Gini with usage of different Hyper-parameters to tune the model.
- K Nearest Neighbors and Naïve Bayes with Hyper parameters
- Bagging Classifier and Ensemble algorithms like Random Forest with different Hyper parameters
- .Boosting, Stacking and Voting along with AdaBoost, GradientBoost and XG Boost with Hyper parameters

Pedagogy

The objectives envisaged in this course will be met through Interactive Lectures and hands on exercises and practice sessions.

Course Content

Session Id	Course Outline
1	•Standard Process of DS projects •Visiting Basics •Odds/Probability •Binomial logistic regression
2	•Logistic regression •Assumptions of LR •Model evaluation metrics •Model Performance metrics •Imbalanced data
3	•Bayes theorem •Business Problem •Naïve Bayes • KNN Basics •KNN algorithm
4	•Understanding terminologies •Entropy, •Shannons entropy, •Conditional entropy •Information gain •Decision Tree •Measure of purity of node •Gini index •Classification error •Construction of Decision Tree •Decision Tree Algorithm
5	•Model Evaluation •Model Performance measures •Overfitting in decision tree •Hyperparameters •Ensemble Learning •Random Forest •Bootstrap sample •Random Forest Hyperparameters •Feature importance
6	• Introduction to Boosting, Adaboost • Gradient boosting • Adaboost vs GBM • XGBoost model discussion • Stacking • Voting
7	Case study