Innovaccer Assignment Solutions

By :-
Shrikant Temburwar
Email: shrikant.temburwar@gmail.com

Qa: Build a classifier to predict the outcome of a new patient with high accuracy.

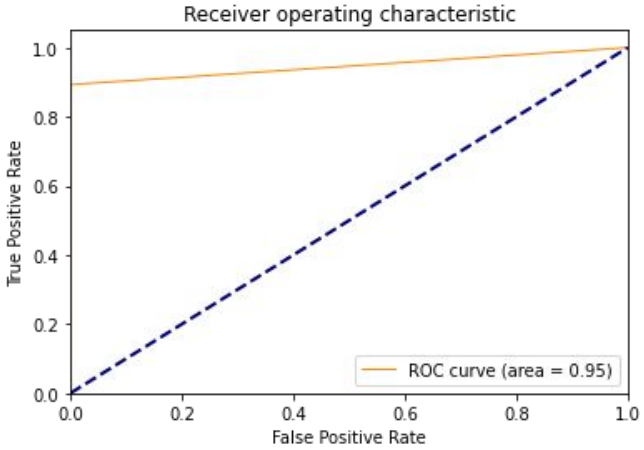
- Result:

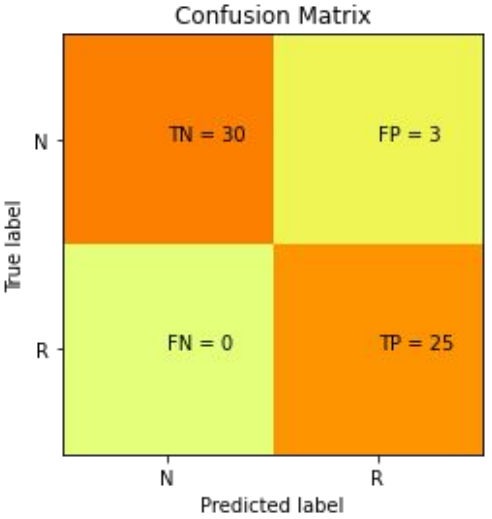
  Accuracy = 94.83%

  F1 Score = 94.85%

  Recall = 1.0

  False Negative = 0





Model = ExtraTrees (Extremely Randomized Trees) Classifier with Hyperparameter tuning.

Approach

- First data is cleaned by replacing '?' with maximum value of that row.
- Data was imbalanced. Tried different model on this data.
- To further improve result, selected better feature and resampled data using SMOTE (Synthetic Minority Oversampling Technique) and further improved accuracy by selecting better model with Hyperparameter tuning using AutoML tool TPOT (Tree-Based Pipeline Optimization Tool).
- Algorithm: **ExtraTrees (Extremely Randomized Trees) Classifier**

1. ExtraTreesClassifier is an ensemble learning method based primarily on decision trees. ExtraTreesClassifier, like RandomForest, randomizes certain data subsets and decisions to minimize data overlearning and overfitting.
2. Builds multiple trees with **bootstrap = False** by default, which means it samples without replacement.
3. The nodes are divided based on random divisions between a random subset of the characteristics selected at each node.

Qb: Build a regression model to predict the recurrence time for patients whose outcome is R.

- Result:

  RMSE = 4.973

  R^2 Score = 0.953

  Model = XGBoost (eXtreme Gradient Boosting) Regressor with Hyperparameter tuning.

Approach

- First data is cleaned by replacing '?' with maximum value of that row.
- Data was imbalanced. Tried different model on this data.
- To further improve result, selected better feature and resampled data using SMOTE (Synthetic Minority Oversampling Technique) and further improved accuracy by selecting better model with Hyperparameter tuning using AutoML tool TPOT (Tree-Based Pipeline Optimization Tool).
- Algorithm: **XGBoost (eXtreme Gradient Boosting) Regressor**

1. XGBoost is a decision tree-based Machine Learning algorithm that uses a gradient reinforcement framework.
2. XGBoost focuses on your speed and the efficiency of your model.
3. It supports parallelization by creating decision trees.
4. XGBoost minimizes a regularized objective function (L1 and L2) that combines a convex loss function and a penalty term for the complexity of the model (in other words, the regression tree functions).