Decision Tree And Random Forest Classifier Models

Decision Tree Classifier

What is The Decision Tree Classifier?

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.

It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

Random Forest Classifier

What is The Random Forest Classifier?

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique.

It can be used for both Classification and Regression problems in ML.

It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

Evaluation Classification Models

What is The Evaluation Classification Models?

Evaluating the performance of a Machine learning model is one of the important steps while building an effective ML model.

To evaluate the performance or quality of the model, different metrics are used, and these metrics are known as performance metrics or evaluation metrics.

Confusion Matrix

What is The Confusion Matrix?

A confusion matrix is a tabular representation of prediction outcomes of any binary classifier, which is used to describe the performance of the classification model on a set of test data when true values are known.

Beginner Friendly CATBOOST with OPTUNA

Catboost HyperParameter Tuning with Optuna

Parameters:

- Objective: Supported metrics for overfitting detection and best model selection
- colsample_bylevel: this parameter speeds up the training and usually does not affect the quality.
- depht : Depth of the tree.
- boosting_type : By default, the boosting type is set to for small datasets. This prevents overfitting but it is expensive in terms of computation. Try to set the value of this parameter to to speed up the training.
- bootstrap_type : By default, the method for sampling the weights of objects is set to . The training is performed faster if the method is set and the value for the sample rate for bagging is smaller than 1.

Conclusion

1. decide which metric to use.
2. analyze both target and features in detail.
3. transform categorical variables into numeric so we can use them in the model.
4. use pipeline to avoid data leakage.
5. look at the results of the each model and selected the best one for the problem on hand.
6. look in detail Catboost
7. make hyperparameter tuning of the Catboost with Optuna to see the improvement
8. look at the feature importance.
9. After this point it is up to me to develop and improve the models.