

Assignment 3

1. Describe how a decision tree works and explain the concepts of nodes, branches, splits, and leaf nodes? How do you overcome overfitting in regression trees? (Expecting not more than 200 words)
2. Explain why Random Forest work better than decision trees and how do you choose between a decision tree and random forests?
3. Using the diabetes dataset implement both decision tree and Random Forest to create baseline models (Use all predictors to predict the outcome). Compare accuracy in both implementations and write your observations.
4. Use the feature_importance_values to determine the most critical variables and use them to implement boosting algorithm such as gradient boosting on the diabetes dataset. Compare the performance of boosting model with Decision tree and Random Forest models.