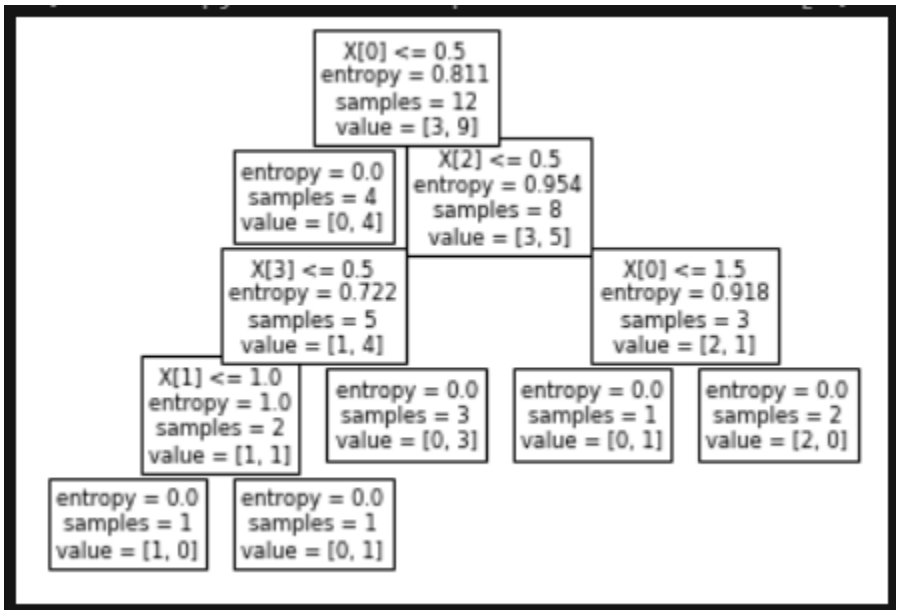
LAB REPORT-2

Mayank Raj (B19CSE053)

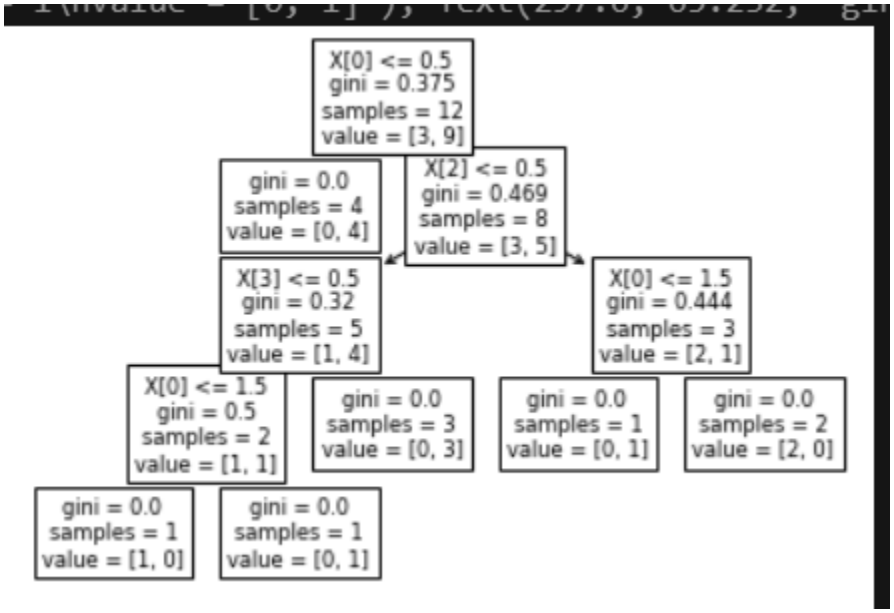
Question 1)

Procedure:

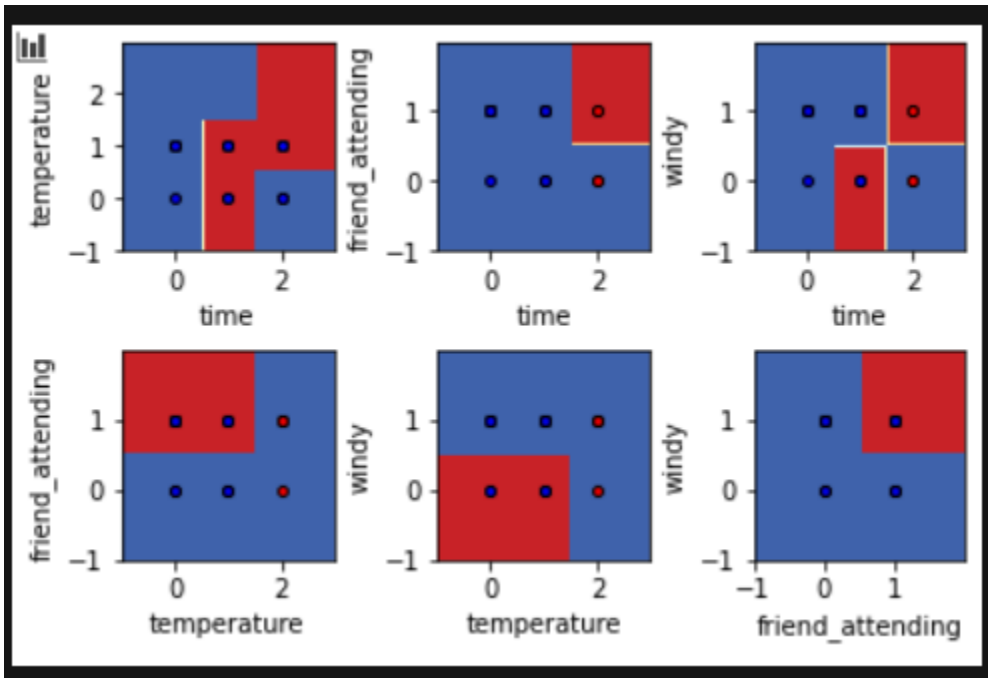
- Preprocessing - we converted data in integers using LabelEncoder in order to apply our model.
- Cross_validation: Cross validation was done using 5 fold with cross_val_score function.
- Training : we trained the model both on gini and entropy both and checked their accuracy.
- Plotting decision tree and decision boundary: We then printed the tree using plot_tree and decision surface was printed using the reference 4 through matplotlib.



Entropy Model



Gini model



Conclusion

- Depth of gini model<=Depth of entropy model
- The Gini model has better accuracy than the entropy model on test data.
- The cross validation score for the gini model is also higher.

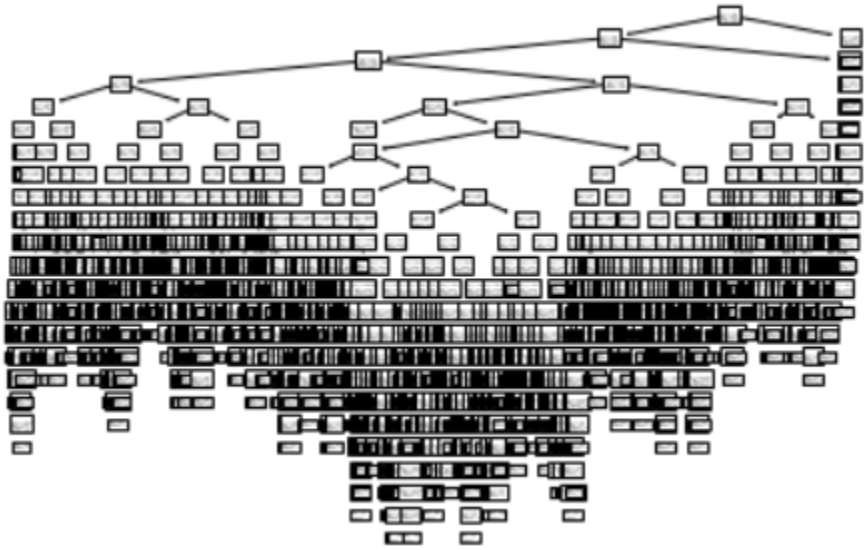
Question 2)

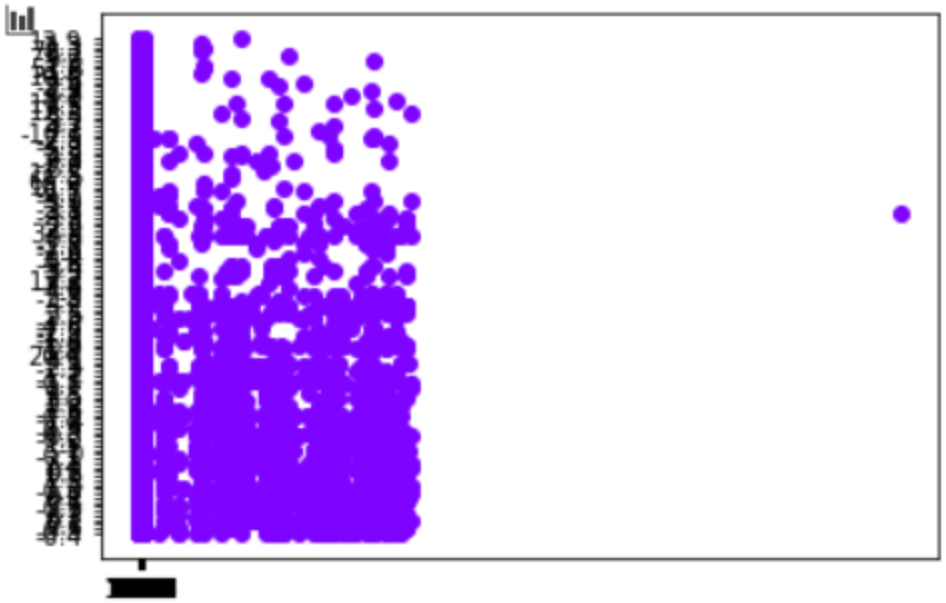
Procedure:

- Preprocessing: This was done by using Label Encoder for strings and the numbers with ‘ , ’ in between were converted to string and ‘ , ‘ was removed and all columns with null values were removed.
- The data was separated into training and testing sets.
- The cross validation was done with 5 folds and mean of the

score was printed

- Model Training: Then the model was trained on DecisionTreeRegression .
- The ground truth and mean squared error was calculated Testing accuracy was 0.5188 and mean squared error was 1.18e-33 And 3.77 for training and testing respectively.
- Decision tree and Decision was plotted using the function used in question 1, since this is large data set ,The tree will have many nodes and the decision surface will be highly dense.





Conclusion:

- The testing error was very high in this case because of using decision regression trees. Random forests or linear regression was a better choice as they will yield better results.