**QUESTION 1:**

For each classifier, various parameters used are mentioned in the pdf (screenshots.pdf) included in the code itself check in page number 11 in the pdf.

**INFERENCES FROM THE QUESTION 2,3,4 :**

*Effect of Dataset Size on Model Performance:*

• As dataset size increases (D1 → D2 → D3 → D), model performance generally improves.

• Larger datasets provide better generalization, reducing overfitting seen in smaller datasets.

• The difference in accuracy between D3 (75%) and D (100%) is relatively small, suggesting diminishing returns after a certain dataset size.

*Comparison of Classifiers:*

• Decision Tree: Performs well on large datasets but overfits on smaller ones.

• Naïve Bayes: Consistently lower accuracy due to its strong independence assumption.

• KNN: Performs moderately but is computationally expensive as data size increases.

• ANN: Achieves the best balance between performance and scalability, adapting well to dataset size changes.

*Impact of SMOTE on Balancing Data:*

• Before applying SMOTE, the model was biased towards majority price categories.

• SMOTE improved recall, ensuring minority classes were better recognized.

• However, slight noise was introduced due to synthetic data, which slightly impacted precision.

• The performance boost from SMOTE was significant in Decision Tree and ANN but minimal for Naïve Bayes.

**Description about assignment:**

**How I cleaned my dataset:**

**•** Check for blank rows and remove them

**•** Check for spelling mistakes and redundancy and edit those cells

Facet->text facet->blank

**•** Checkwhether they are in the range of the data if not edit those cells

**•** Cluster and editrequired cells

**•** Changing the data type of the columns if required

Like text to numbers for attributes like price.

**•** Handling missing values with mean of that column.

**•** Checking repetitions by grouping columns and clustering and using text filters if required based on the columns.

**•** rounding off the values given up to 2 decimal places for all the attributes for better comparison.

***#FOR DOING THESE OPERATIONS I USED GREL COMMANDS***

***•*** After cleaning the dataset I created a file to split my data set and run the code in the terminal which divides the data set into 80% training set and 20% test dataset by random splitting.

• Then write the codes and run them it will display the desired output indicating the required parameters if mentioned and also displays the image overview.