Business Analytics with R – Assignment 2

Name – Amruth Pai Thukaram

Net ID – AXT230147

Date – 31st October 2024

9.1]

A]

Rule 1: root

Rule 2: root -> OpenPrice>=3.685

Rule 3: root -> OpenPrice>=3.685 -> ClosePrice< 20.2

Rule 4: root -> OpenPrice>=3.685 -> ClosePrice>=20.2

Rule 5: root -> OpenPrice>=3.685 -> ClosePrice>=20.2 -> OpenPrice>=20.5

Rule 6: root -> OpenPrice>=3.685 -> ClosePrice>=20.2 -> OpenPrice< 20.5

Rule 7: root -> OpenPrice< 3.685

Rule 8: root -> OpenPrice< 3.685 -> ClosePrice< 3.685

Rule 9: root -> OpenPrice< 3.685 -> ClosePrice< 3.685 -> OpenPrice>=2.445

Rule 10: root -> OpenPrice< 3.685 -> ClosePrice< 3.685 -> OpenPrice< 2.445

Rule 11: root -> OpenPrice< 3.685 -> ClosePrice>=3.685

OpenPrice, ClosePrice seems to be best variables to choose as majority of the splits are based on these variables and importance function also indicates high value for these variables

A diagram of a graph

Description automatically generated

B]

After running the model on validation data we get an accuracy of 74-77% which is not ideal for real world scenarios but practical enough for this small dataset.

C]

interesting patterns

open price – rule 2 Auctions with an OpenPrice of 3.685 or higher influence competitive outcomes.

Close price – (rule 3,4) The ClosePrice differentiates outcomes for auctions with an OpenPrice >= 3.685.

Within the high ClosePrice group, further categorization by OpenPrice above or below 20.5 refines understanding of competitive behavior.

Auctions with an OpenPrice below 3.685 can still be competitive, with further distinctions based on ClosePrice.

uninteresting patterns

Indicates that even with low ClosePrice, there's little differentiation based on a minor range of OpenPrice.

D]

Based on the above scenario since the splits are majorly on openprice and closeprice also the importance function gives high values to these variables and very low to others. We will use only these 2 variables to create a tree

Rule 1: root

Rule 2: root -> OpenPrice>=3.615

Rule 3: root -> OpenPrice>=3.615 -> ClosePrice< 10

Rule 4: root -> OpenPrice>=3.615 -> ClosePrice>=10

Rule 5: root -> OpenPrice>=3.615 -> ClosePrice>=10 -> OpenPrice>=10.34

Rule 6: root -> OpenPrice>=3.615 -> ClosePrice>=10 -> OpenPrice< 10.34

Rule 7: root -> OpenPrice< 3.615

Rule 8: root -> OpenPrice< 3.615 -> ClosePrice< 3.645

Rule 9: root -> OpenPrice< 3.615 -> ClosePrice< 3.645 -> OpenPrice>=1.035

Rule 10: root -> OpenPrice< 3.615 -> ClosePrice< 3.645 -> OpenPrice< 1.035

Rule 11: root -> OpenPrice< 3.615 -> ClosePrice>=3.645

A diagram of a graph

Description automatically generated

E]

A graph with red and blue dots

Description automatically generated

The splitting seems effective for low prices auctions but not fully account for auctions with higher openprice or close price

split is reasonable for the main cluster of auctions but may need refinement for better separation across the entire range

f]

A graph with a line

Description automatically generated

Area under the curve is comparatively less and cannot say the model is good enough for the dataset

A number and numbers on a white background

Description automatically generated

Accuracy is 80.01% which is reasonably good but there are many misclassifications for non-competitive

G]

A lower opening price could attract more bidders, especially if buyers perceive they might get a good deal. Conversely, a higher opening price may deter bidders

Sellers should consider starting their auctions with an opening price above 3.615. This price point is crucial for enhancing the chances of receiving at least two bids, as lower opening prices tend to lead to less competitive outcomes.

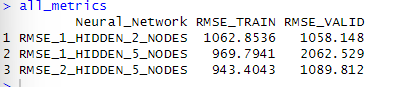
Aim for a closing price that exceeds 10, as this has shown to be a threshold for attracting competitive bids. This may involve setting a reserve price or considering the perceived value of the item.

The factors like duration, currency, category and end date don’t play that huge of a role or very insignificant role in splitting the decision tree therefore I would suggest anything close by to others should do.

11.3]

A diagram of a network

Description automatically generated



i] As the number of layers and nodes increases in the neural network, the RMSE for the training data decreases

ii] The validation RMSE initially rises significantly when increasing from 2 to 5 nodes with 1 hidden layer (indicating possible overfitting). However, when adding a second hidden layer, the RMSE for validation decreases, although it remains higher than the training RMSE

iii]

1 Hidden Layer, 2 Nodes offers a simpler model with relatively similar RMSE values for both training and validation, suggesting less overfitting.

2 Hidden Layers, 5 Nodes reduces the training RMSE while maintaining a reasonable validation RMSE, so it could be suitable if the application requires more flexibility in capturing patterns.

An Appropriate number to choose from the above would be 2 hidden layers and 5 nodes but I would choose using 2 hidden layers and lesser nodes for better performance