Name : Vaishnavi Eknath Avhad
Class : D15C
Roll No. : 41

Practical 6

Aim : To implement Classification algorithms (Decision tree and Naïve Bayes' algorithms) using Rapid Miner tool

Theory :

1. Decision Tree

A Decision Tree is a flowchart-like structure used to make decisions based on feature values. It splits the dataset based on the best possible splits according to a criterion like Gini Index or Information Gain.

- RapidMiner Implementation : The "Decision Tree" operator is used to train a decision tree, where it splits nodes based on features to predict the class (fraud vs non-fraud).

2. Naïve Bayes

Naïve Bayes is a probabilistic classifier based on Bayes' Theorem, assuming feature independence. It works well for large datasets and is commonly used in text classification and fraud detection.

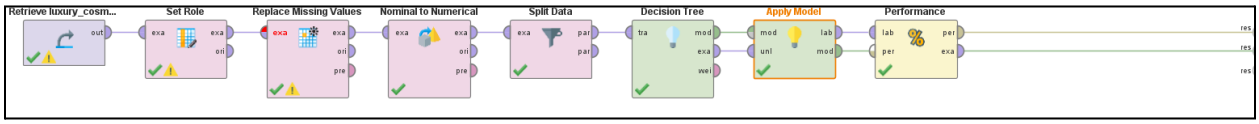
- RapidMiner Implementation: The "Naive Bayes" operator is used to train a probabilistic model to predict whether a transaction is fraudulent.

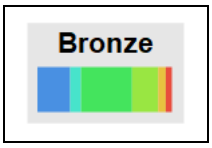
3. Evaluation Metrics

- Accuracy : The percentage of correctly predicted instances out of all predictions.
- Precision : The ratio of correctly predicted positive observations to the total predicted positives.
- Recall : The ratio of correctly predicted positive observations to all observations in actual class.
- F1-Score : The weighted average of precision and recall.

- Confusion Matrix : A matrix used to evaluate the performance of the classification model.

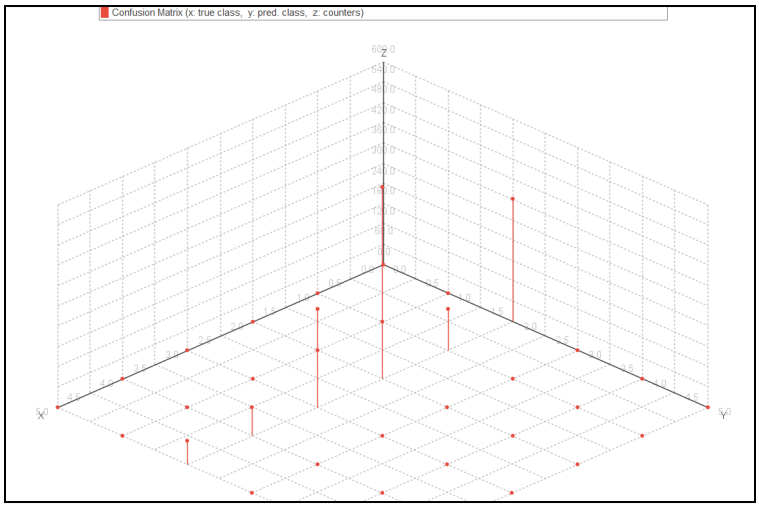
Code with output :

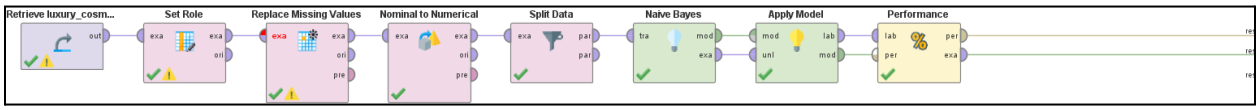




accuracy: 37.98%

	true Silver	true Platinum	true Bronze	true Gold	true VIP	true None	class precision
pred. Silver	0	0	0	0	0	0	0.00%
pred. Platinum	0	0	0	0	0	0	0.00%
pred. Bronze	364	120	567	290	84	68	37.98%
pred. Gold	0	0	0	0	0	0	0.00%
pred. VIP	0	0	0	0	0	0	0.00%
pred. None	0	0	0	0	0	0	0.00%
class recall	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	



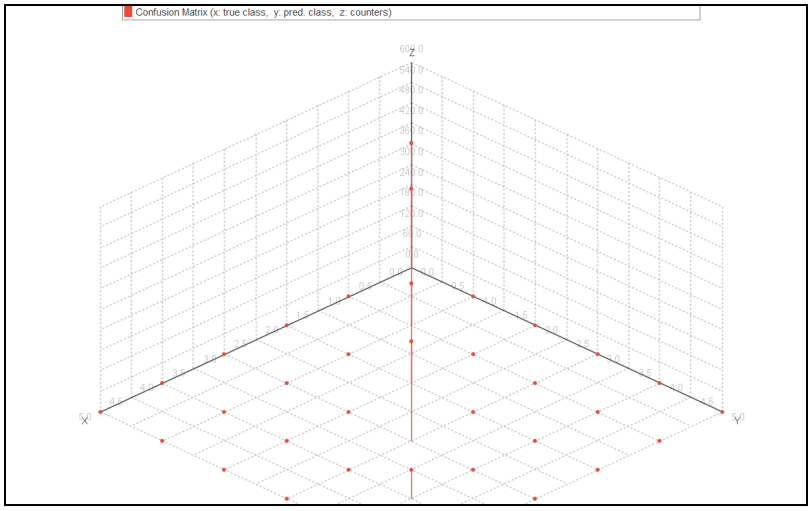


SimpleDistribution

```
Distribution model for label attribute Customer_Loyalty_Tier


Class Silver (0.244)
6609 distributions

Class Platinum (0.080)
6609 distributions

Class Bronze (0.380)
6609 distributions

Class Gold (0.194)
6609 distributions

Class VIP (0.056)
6609 distributions

Class None (0.046)
6609 distributions
```

accuracy: 100.00%

	true Silver	true Platinum	true Bronze	true Gold	true VIP	true None	class precision
pred. Silver	364	0	0	0	0	0	100.00%
pred. Platinum	0	120	0	0	0	0	100.00%
pred. Bronze	0	0	567	0	0	0	100.00%
pred. Gold	0	0	0	290	0	0	100.00%
pred. VIP	0	0	0	0	84	0	100.00%
pred. None	0	0	0	0	0	68	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	



Conclusion :

This experiment demonstrates how to implement two important classification algorithms, Decision Tree and Naïve Bayes, using RapidMiner for fraud detection. The process includes data preprocessing, model training, evaluation, and visualization, making it a comprehensive approach for building predictive models in the business context.