Examining Customer Purchase Intent

Predictive Analytics Project: Wong Wen Bing Y2 DAAA

TABLE OF CONTENTS

01

About

Examining the retail industry

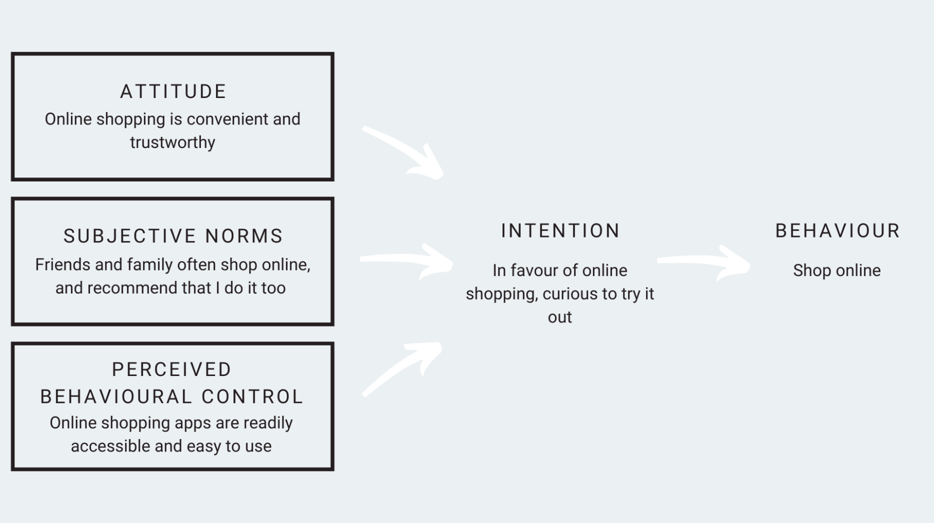
Theory of Planned Behaviour (TPB)

"The theory of planned behaviour assumes that, when an individual **intends to perform a specific behaviour, they will actually do so**. According to the theory of planned behaviour, behavioural intentions are formed on the basis of attitudes, subjective norms and perceived behavioural control."

Gu, Shengyu. (2019). Using the Theory of Planned Behaviour to Explain Customers' Online Purchase Intention. 5. 226-249. 10.6911/WSRJ.201909_5(9).0026.

Applying TPB in Retail

- A frequently used technique by marketers to predict and understand consumer's behaviour.
- Argument for the theory's usage in retail is that behaviours are **solely determined on behavioural intentions** which are determined by **behavioural attitude, subjective norms and perceived behaviour**



Example on Online Shopping (Blogs @NTU)

Problem for retailers

“Despite the growing attention on enhancing purchase intent, there is no coherent framework for determining retail customers’ purchase intent.”

Menidjel, Choukri & Bilgihan, Anil. (2022). The determinants of retail customers’ purchase intent. International Journal of Consumer Studies. 46. 10.1111/ijcs.12802.

Retailers

What do they

THINK AND FEEL?

- Concerned about the unpredictability in sales and customer behaviour
- Wants to tailor marketing strategies to increase customer engagement and loyalty

What does he/she

SEE?

- Sales data, transaction history, customer demographics (Data)
- Purchasing patterns,in-store traffic paterrns, seasonal spikes

What do they

HEAR?

- Customer reviews, complaints and suggestions
- Social media conversations
- Industry trends, best practices
- Competitor activities, promotional strategies

What does he/she

SAY AND DO?

- Implement predictive analytics tools to forecast sales trends.
- Adjust pricing, promotions, and inventory based on predictive insights.

PAIN	GAIN
• Risk of over-reliance on historical data, leading to inaccurate predictions. • Balancing short-term sales targets with long-term customer relationship management. • Unable to keep up with industry trends.	• Enhanced customer satisfaction through personalized experiences. • Increased profitability by optimizing marketing spend and reducing waste.

"Empowering retailers to make MORE informed decisions about their business through **predictive modelling research on purchase intent**."

~Project Business Objective

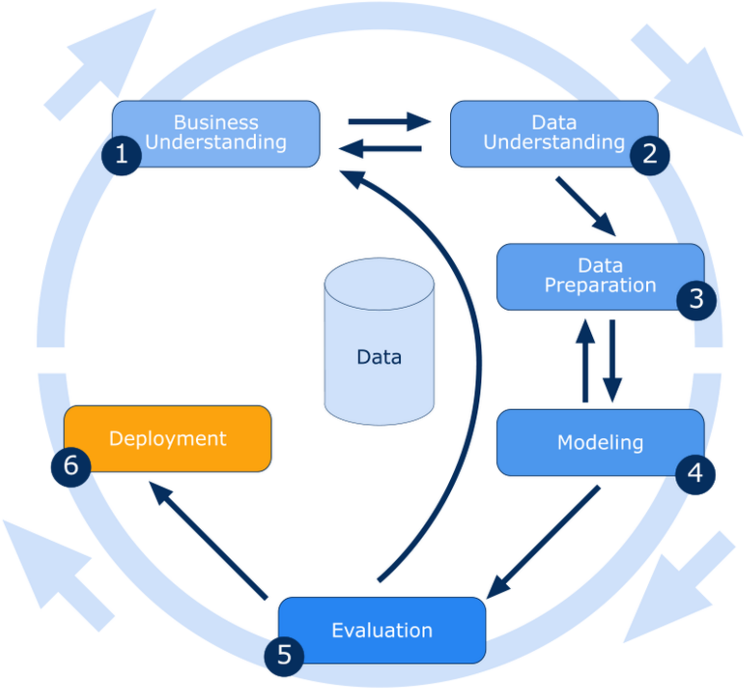
02

Process

CRISP-DM

CRISP-DM

- Comprehensive process of conducting a data mining project
- Structured: ensuring all aspects of this investigation are aspects of the project are thoroughly addressed
- Effective: ensures business goals are met and thorough data preparation and evaluation is conducted



03

Data Understanding and Preparation

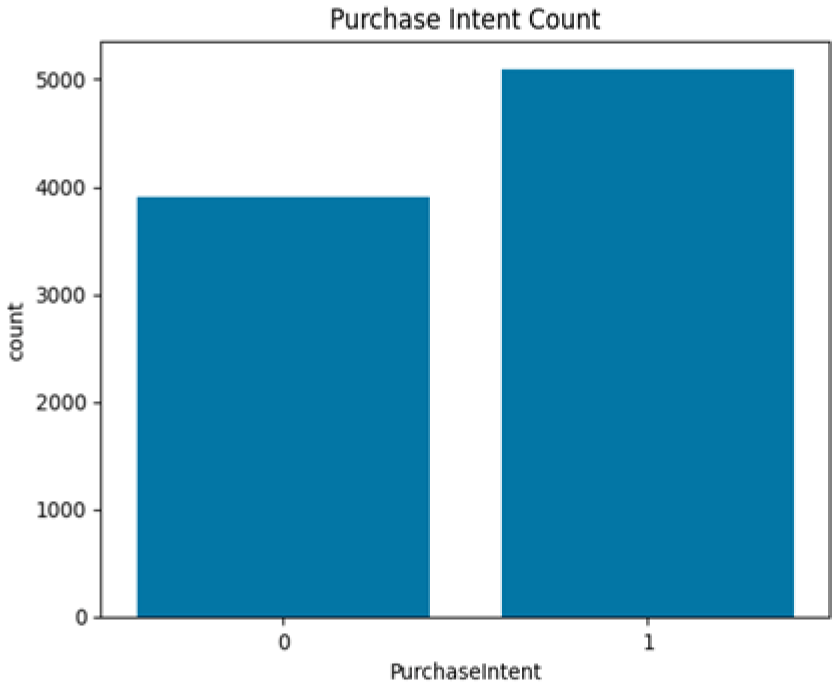
Dataset Information – Electronic Sales Dataset

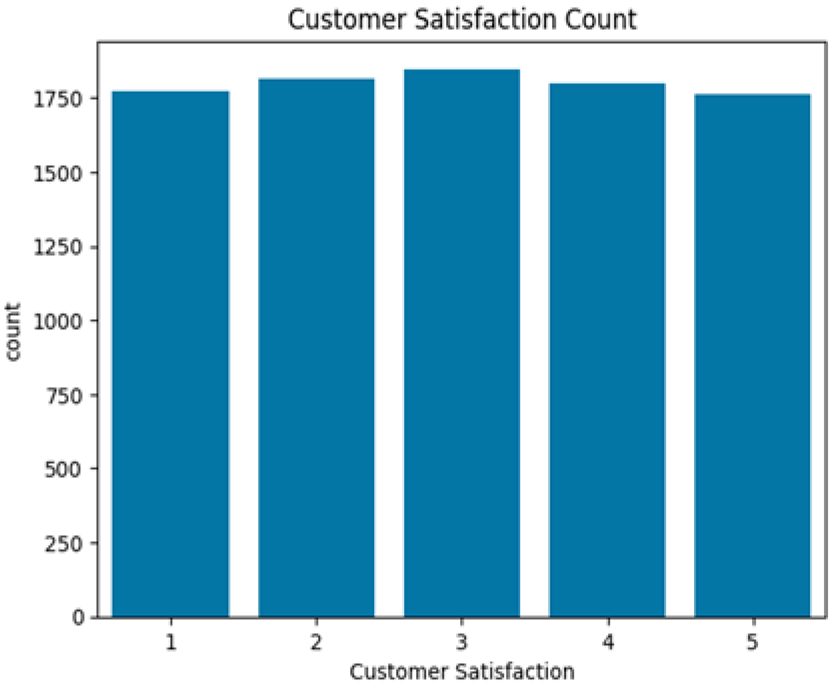
- Source: https://www.kaggle.com/datasets/rabieelkharoua/consumer-electronics-sales-dataset/data
- Examine customer purchase intent based on an Electronics store
- Data Information:
  - Total Data: 9000 rows
  - Total Number of Columns: 9

Data Dictionary

ProductID	Unique identifier for each product.
ProductCategory	Category of the consumer electronics product (e.g., Smartphones, Laptops).
ProductBrand	Brand of the product (e.g., Apple, Samsung).
ProductPrice	Price of the product ($).
CustomerAge	Age of the customer.
CustomerGender	Gender of the customer (0 - Male, 1 - Female).
PurchaseFrequency	Average number of purchases per year.
CustomerSatisfaction	Customer satisfaction rating (1 - 5).
PurchaseIntent	Intent to purchase. (0 - no intent, 1 - intent)

Balance of Dataset





Data Preparation – KNIME

- Checking and removing necessary null values
- Removing unnecessary columns – ProductID
- Changing values in CustomerGender (0 to Male, 1 to Female)

Data Preparation – Reducing Overfitting

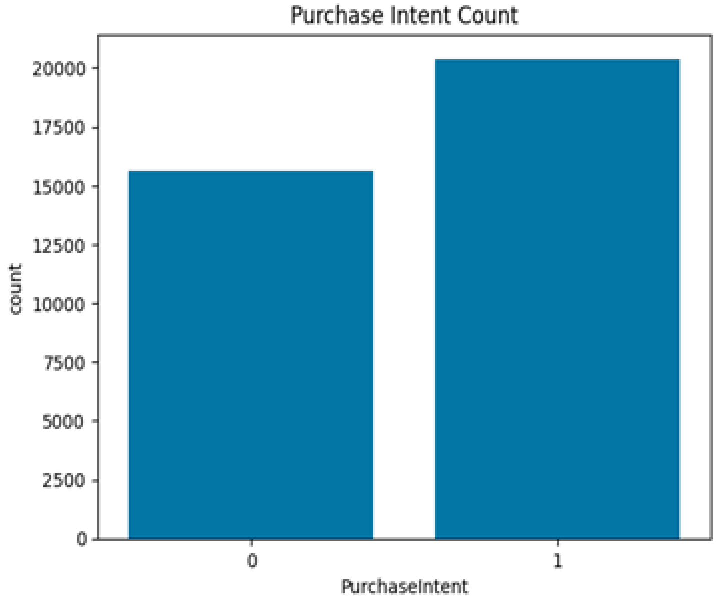
- During Data preparation: Even though dataset is balanced but there is overfitting due to data count and imbalance
- In KNIME: SMOTE node to create synthetic data
- By doing so: reduces the possibility of overfitting

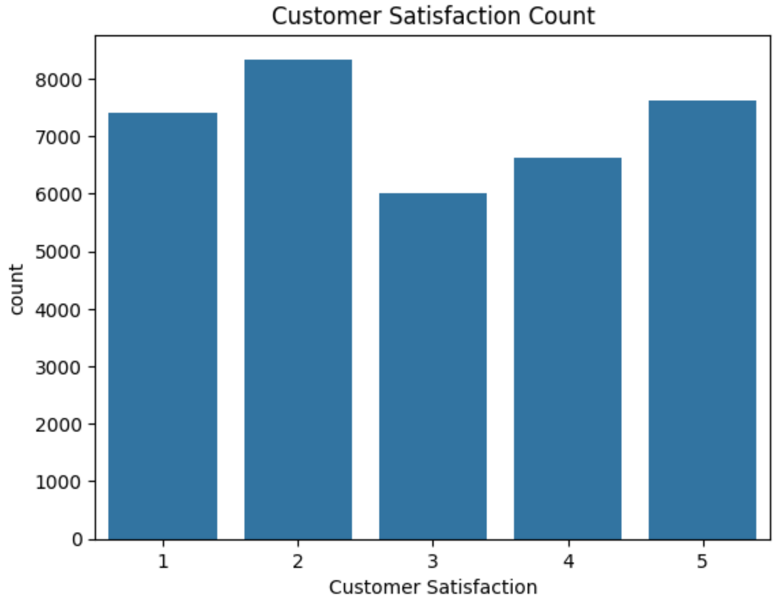
Final Dataset

- Total Number of Rows: 36000
- Total Number of Features: 8

Name	Type of Variable	Description
ProductCategory	Category	Category of the consumer electronics product (e.g., Smartphones, Laptops).
ProductBrand	Category	Brand of the product (e.g., Apple, Samsung).
ProductPrice	Numeric (Aggregated as Average)	Price of the product ($).
CustomerAge	Numeric	Age of the customer.
CustomerGender	Category	Gender of the customer (Male, Female).
PurchaseFrequency	Numeric (Aggregated as Average)	Average number of purchases per year.
CustomerSatisfaction	Category	Customer satisfaction rating (1 - 5).
PurchaseIntent	Category	Intent to purchase.

Data Balance





Partioning

- Given the relatively well balance of the dataset, a 70-30 partition was done with SAS Viya using simple random sampling

WHY Simple Random Sampling:

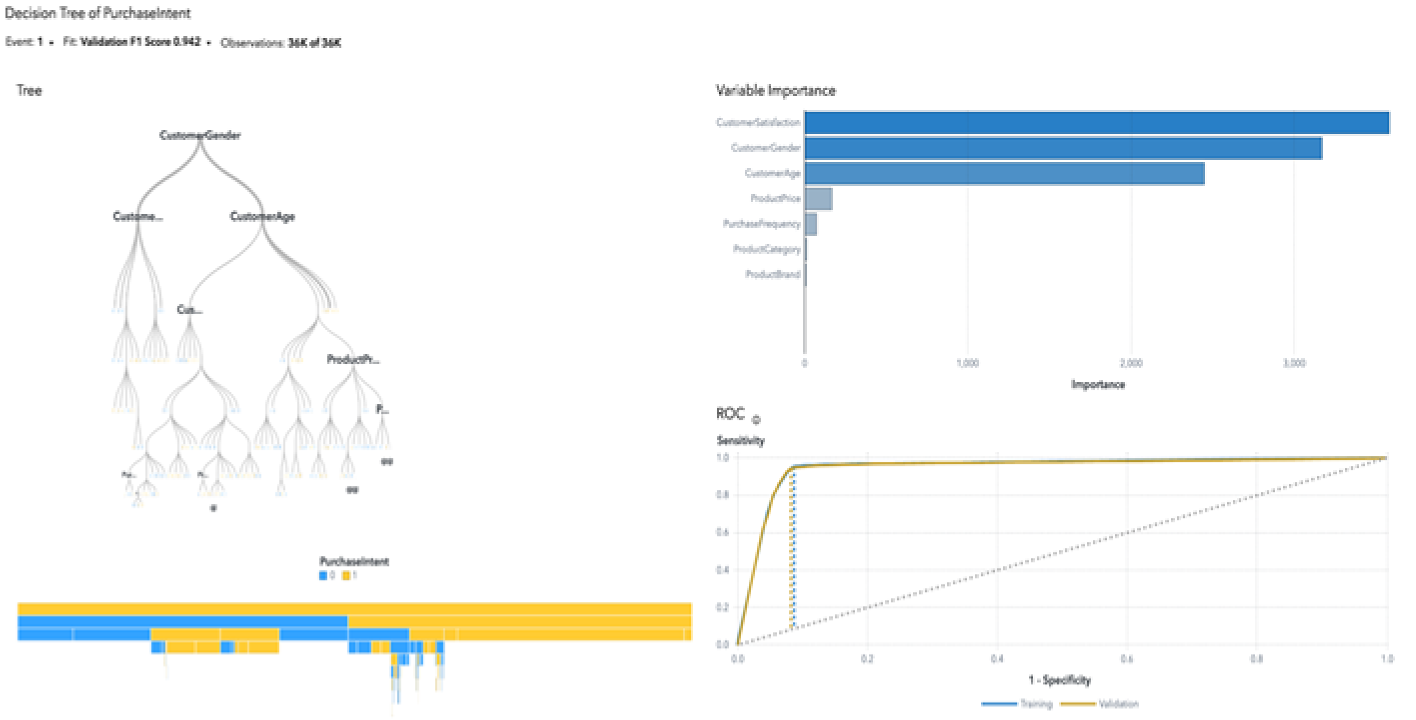
- Both types of variable can have an equal and fair chance of being used as compared to stratified sampling

03

Modelling

Finding the best model possible

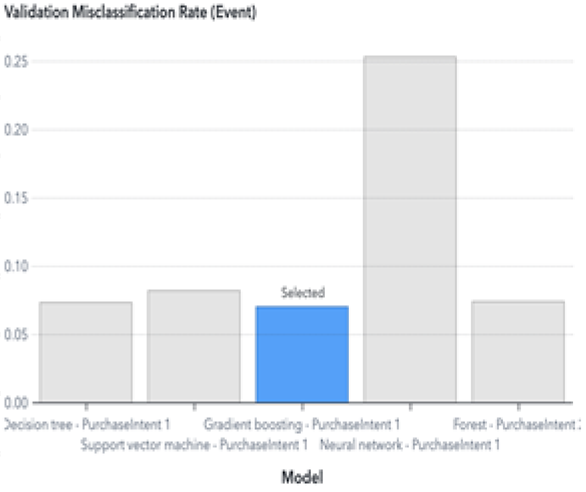
Base Model

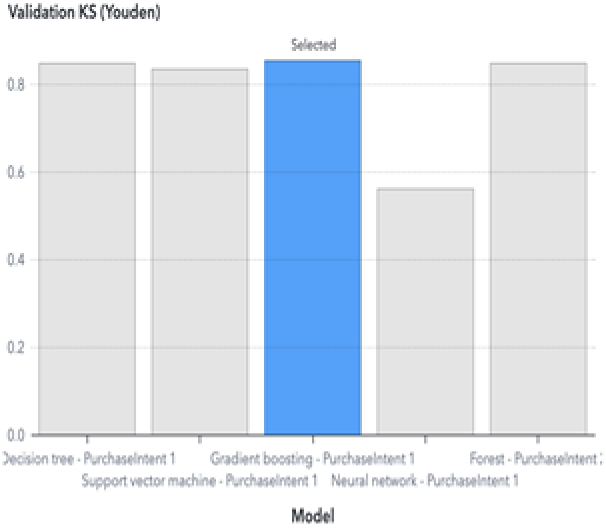


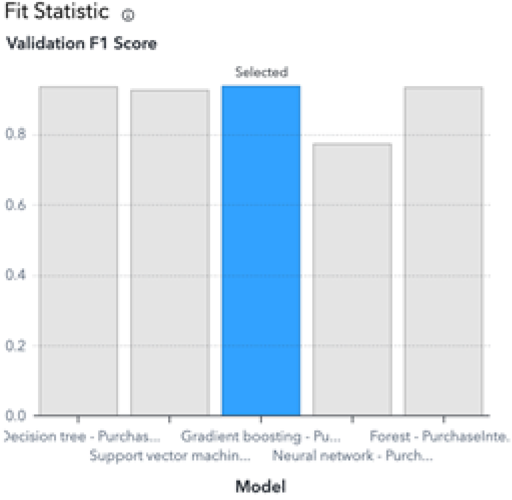
- F1 Score: 0.942
- KS-Youden: 0.8611
- Misclassification Rate: 0.0669

Model Comparison - Statistics

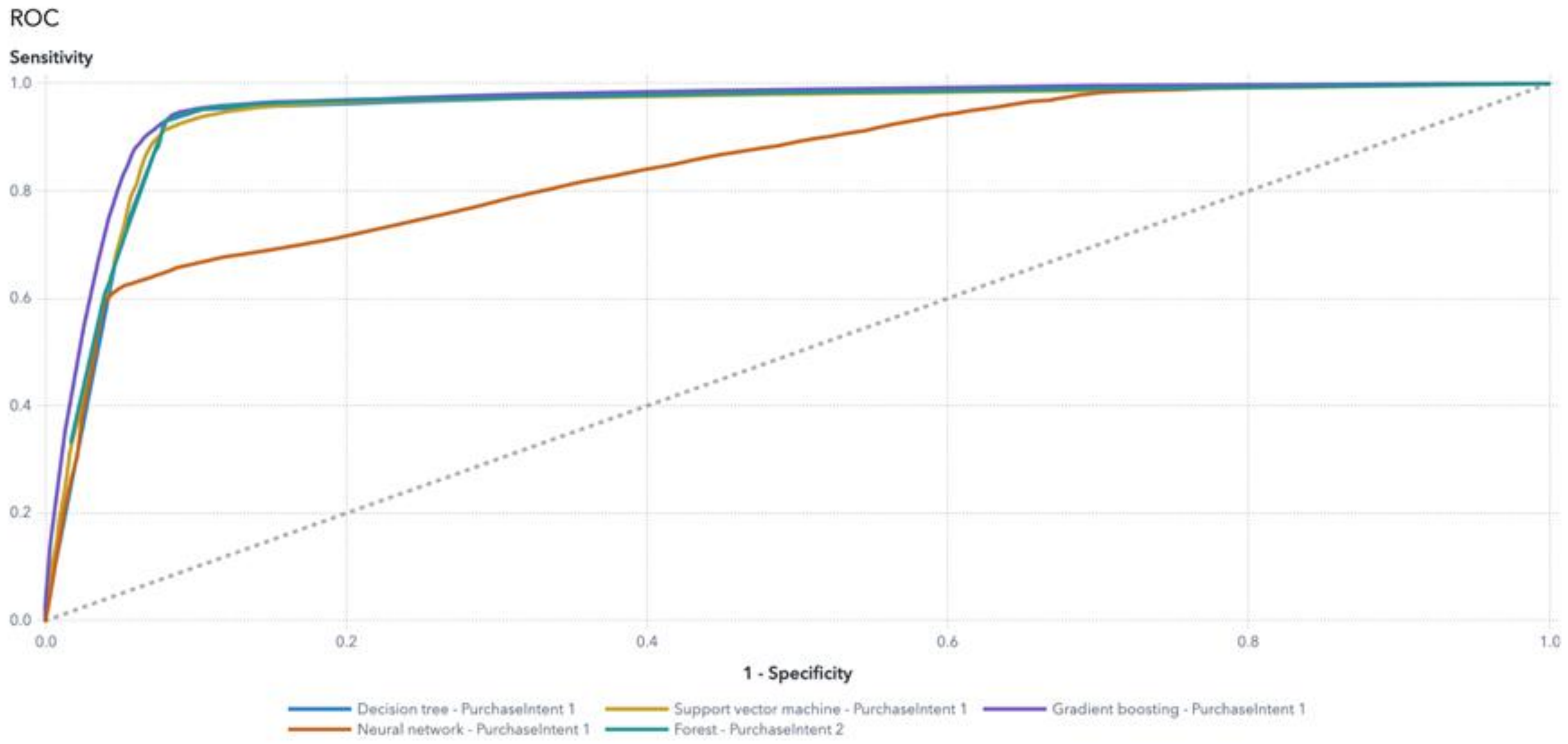
- Variety of models used to model the data: Decision Tree, Support Vector Machine, Gradient Boosting, Neural Network Random Forest



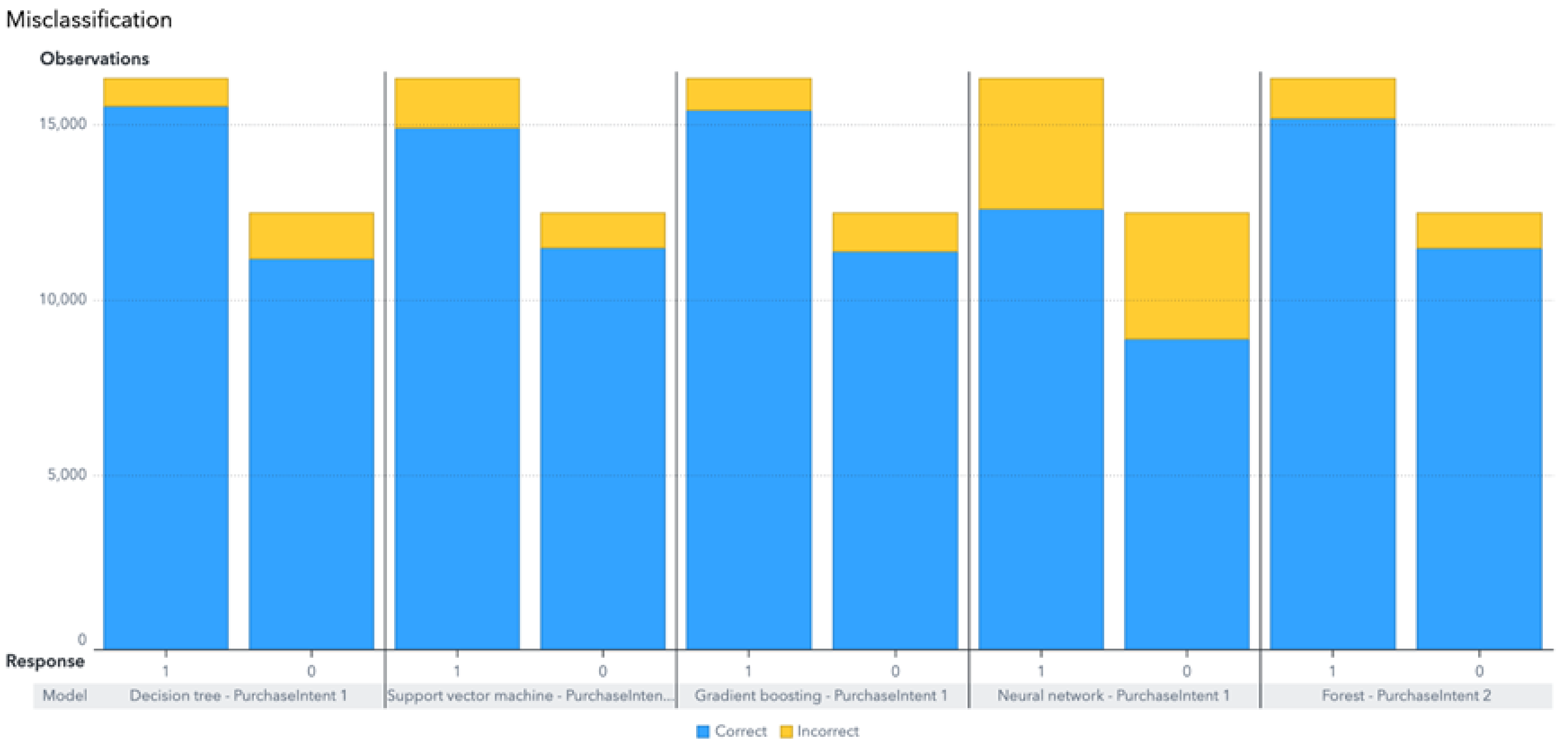




Model Comparison – ROC Curve



Model Comparison - Misclassification



Final Decision

Final Decision: Gradient Boosting

Model Statistics:

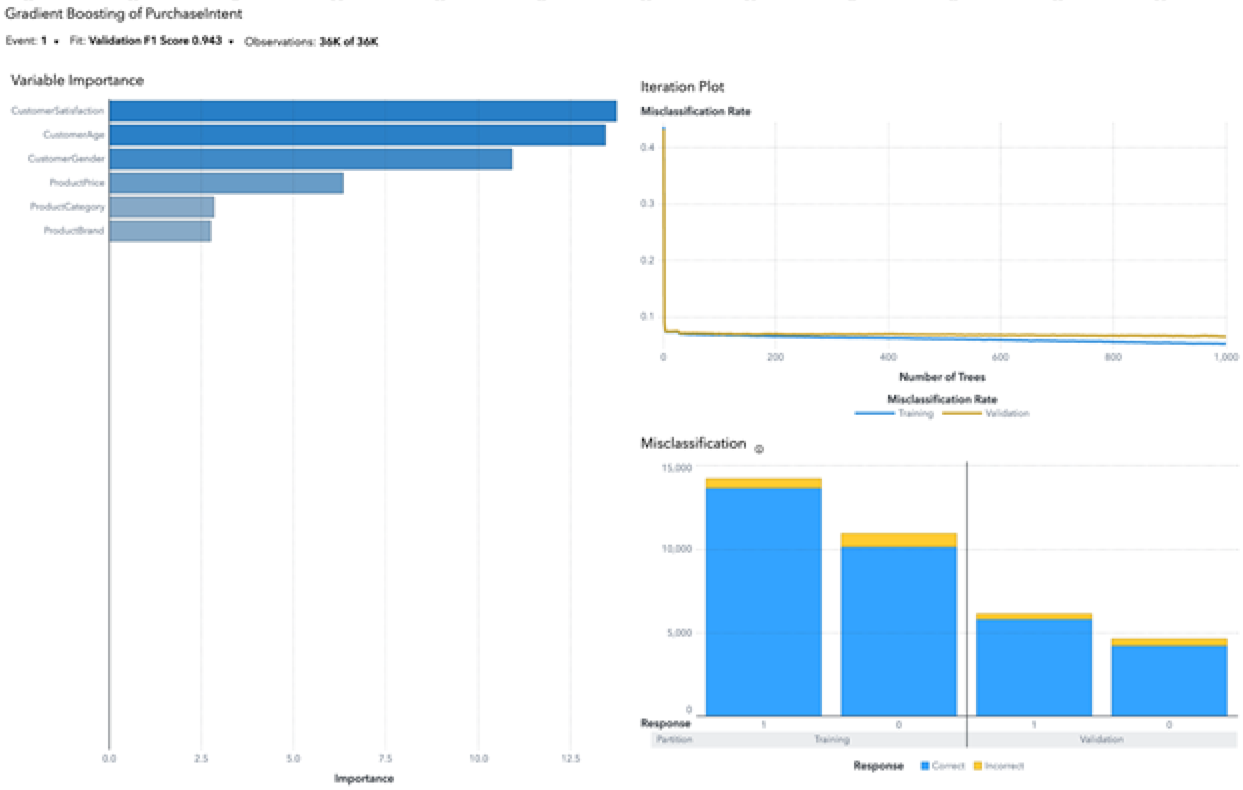
- F1 Score: 0.943 (good performance)
- KS Youden: 0.8982 (sensitive to data changes)
- Misclassification Rate: 0.0530 (less likely to misclassify the data)

04

FINAL MODEL

Gradient Boosting

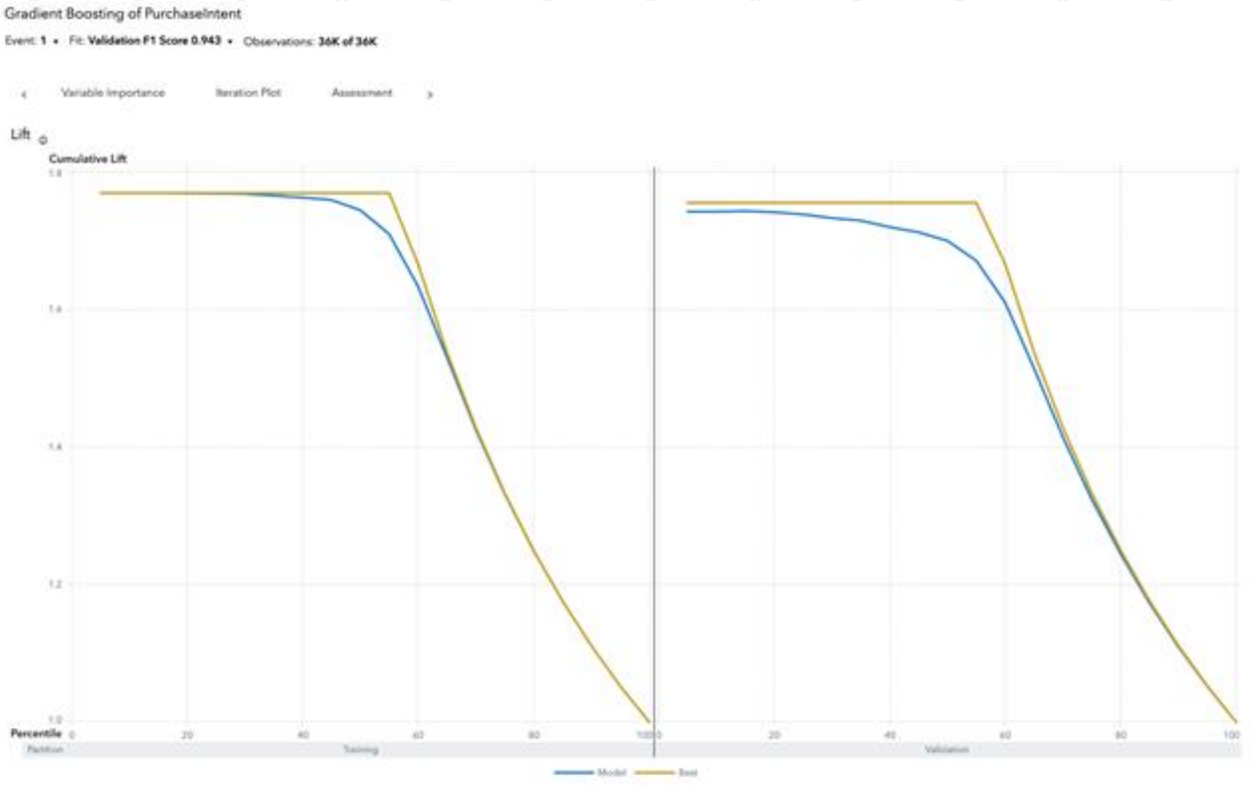
About Model



Model Statistics:

- F1 Score: 0.942
- KS-Youden: 0.8611
- Misclassification Rate: 0.0669

Lift Curve



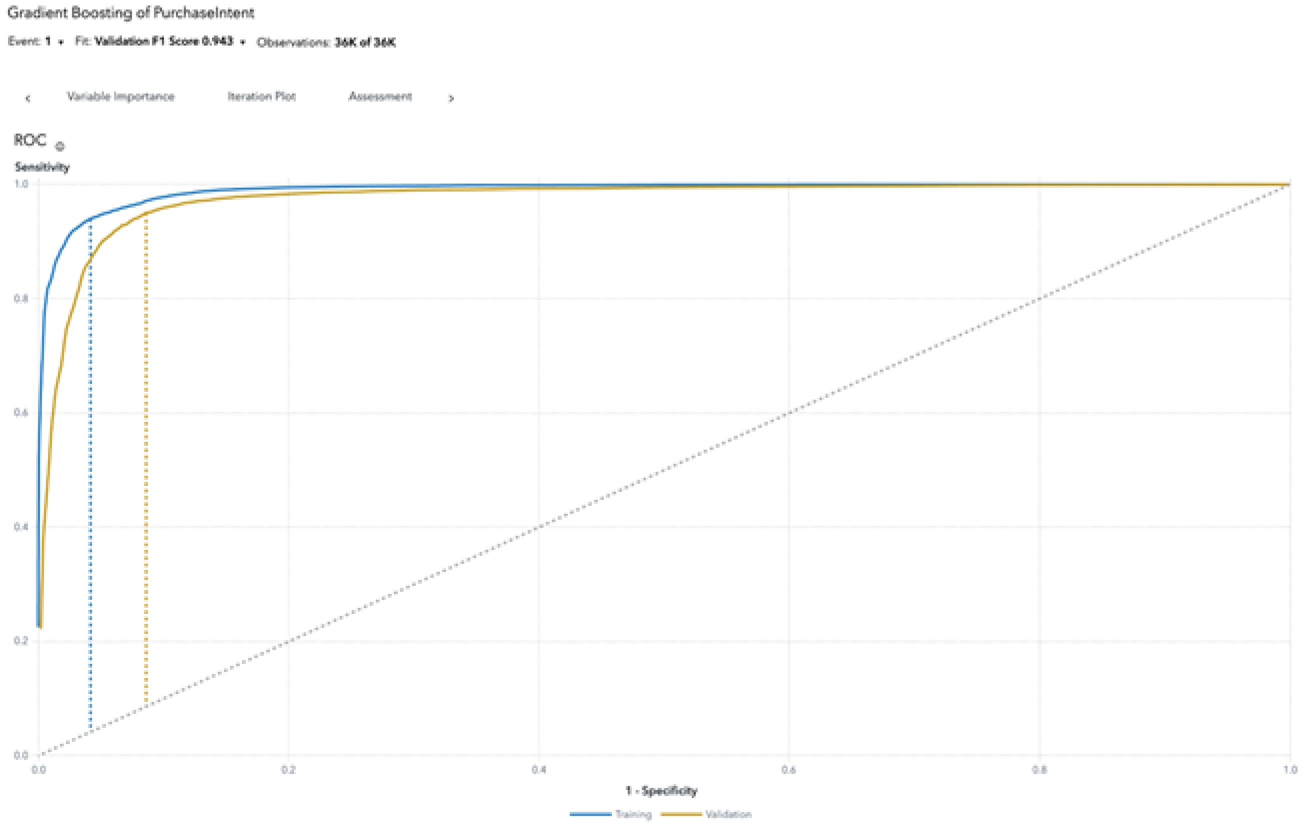
No Overfitting:

- Observation that the validation model performs relatively similar to training model

Good Model Performance:

- Model is performing relatively close to the best model

ROC Curve



Insignificant Overfitting:

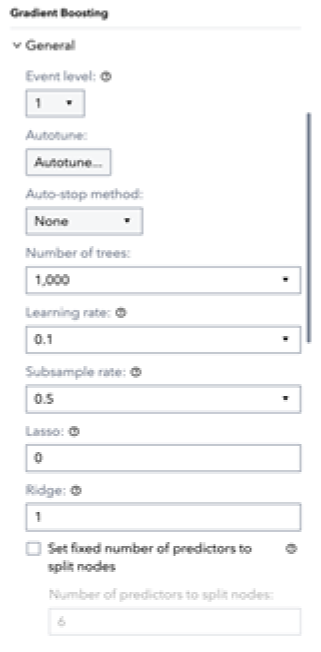
- Training curve is slightly sensitive than the validation

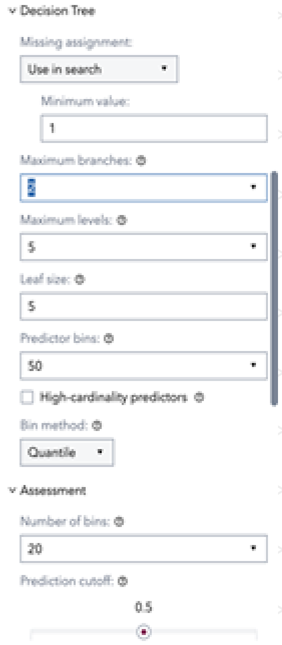
Good Model Performance:

- Curves for both models are close to the top left side which indicates a very sensitive and good model

Fine Tuning

- Higher number of trees → reducing Overfitting
- Lower number of levels → reduces the complexity of the decision tree





Model Analysis (Decision Tree)

- Demographics of customers who purchase the most: (Based on customers with the purchase intent)
    - Customer Age: 26-31/20-21
    - Gender: Female
    - Purchase Frequency: 7 to 9

06

CONCLUSION

You can enter a subtitle here if you need it

Insights

- In determining whether a customer decides to purchase the product, the below factors are important
(in descending order most → least important)
    - Customer Satisfaction
    - Customer Age
    - Customer Gender
    - Product Price
    - Product Brand
    - Product Category

Recommendations

- Ensure **maximum customer satisfaction**
- Tailor marketing strategies to **Females** aged between **26 and 31** or **20 to 21**

Conclusion

- By using the model insights, effective marketing strategies could be tailored to the specific categories of people.
- Linking back to the Theory of Planned Behaviour (TPB), we can justify that the theory does work in its own mysterious ways (with customer satisfaction as a main goal)
- With the increasing use of predictive analysis in the retail fields, companies must use such a model as a baseline for future models to observe different factors – such as Customer Satisfaction determined by the different product brands and prices
- In short, whilst this model can serve as an example, it is important to use **updated** and **customised** data **based on the needs of the specific sector**

THANKS

Do you have any questions?

youremail@freepik.com
+91 620 421 838
yourwebsite.com