Health and Vehicle Insurance: Cross-sell Insights

Team Members

1. Aarsh Muley
2. Aditya Tiwari
3. Ashutosh Jha
4. Shrey Yadav
5. Arhan Ghosh

Constituents:

1. Introduction
2. Problem Statement
3. Business Use Case
4. Data Overview and Preprocessing
5. EDA and Dashboard (Power BI)
6. Machine Learning Approach (Orange)
7. Results & Interpretation
8. Key Learnings and Challenges Faced
9. References

Introduction and Problem Statement

Our client is an Insurance company that has provided Health Insurance to its customers now they need your help in building a model to predict whether the policyholders (customers) from past year will also be interested in Vehicle Insurance provided by the company.

- Key features to put emphasis on.
- What are we aiming to achieve?
- What type of problem are we working on ?
- What are the evaluation metrics that can be helpful in such cases ?

Business Use Case

S (Strengths)	W (Weaknesses)	O (Opportunities)	T (Threats)
• Target customer identification gets easier. • Scalability gets a boost, as the health insurance company can grow by cross-selling vehicle insurance.	• Some constraints do exist, on as there are a lot of features and parameters to be taken care of. That may increase the bias of the models.	• Business Growth • To have a better customer retention as they can stay more longer connected with the company.	• Customer privacy and consent risk. • Working more with past patterns and data could lead to overfitting. • Customer Eligibility conflicts.

Getting One With Data

Data Description

1. *Train Data*

- id Unique ID for the customer
- Gender Gender of the customer
- Age Age of the customer
- Driving_License 0 : Customer does not have DL, 1 : Customer already has DL
- Region_Code Unique code for the region of the customer
- Previously_Insured 1 : Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
- Vehicle_Age Age of the Vehicle
- Vehicle_Damage 1 : Customer got his/her vehicle damaged in the past. 0 : Customer didn't get his/her vehicle damaged in the past.
- Annual_Premium The amount customer needs to pay as premium in the year
- Policy_Sales_Channel Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
- Vintage Number of Days, Customer has been associated with the company
- Response 1 : Customer is interested, 0 : Customer is not interested

27/06/25

Data Description

2. Test Data

- id : Unique ID for the customer
- Gender : Gender of the customer
- Age : Age of the customer
- Driving_License 0 : Customer does not have DL, 1 : Customer already has DL
- Region_Code : Unique code for the region of the customer
- Previously_Insured 1 : Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
- Vehicle_Age : Age of the Vehicle
- Vehicle_Damage 1 : Customer got his/her vehicle damaged in the past. 0 : Customer didn't get his/her vehicle damaged in the past.
- Annual_Premium : The amount customer needs to pay as premium in the year
- Policy_Sales_Channel : Anonymised Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
- Vintage : Number of Days, Customer has been associated with the company
- Submission
- Variable Definition
- id : Unique ID for the customer
- Response 1 : Customer is interested, 0 : Customer is not interested

Power BI Dashboard (Visualization)

CORPORATE
GURUKUL
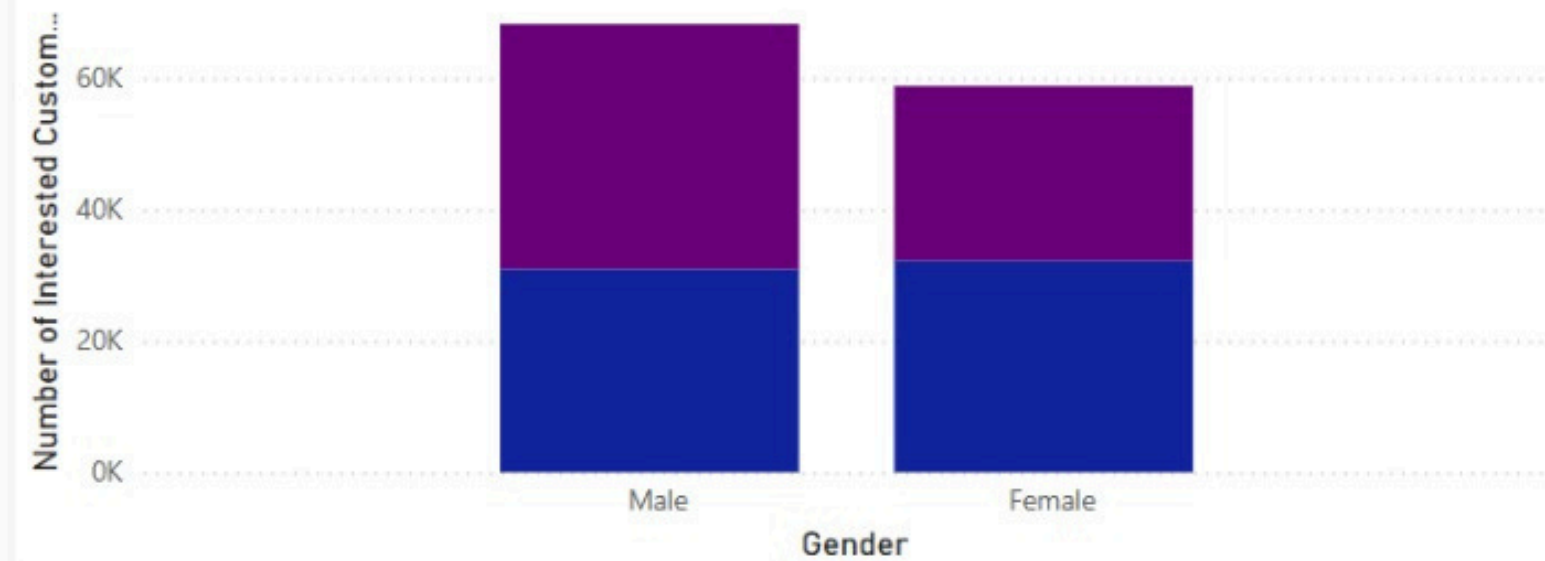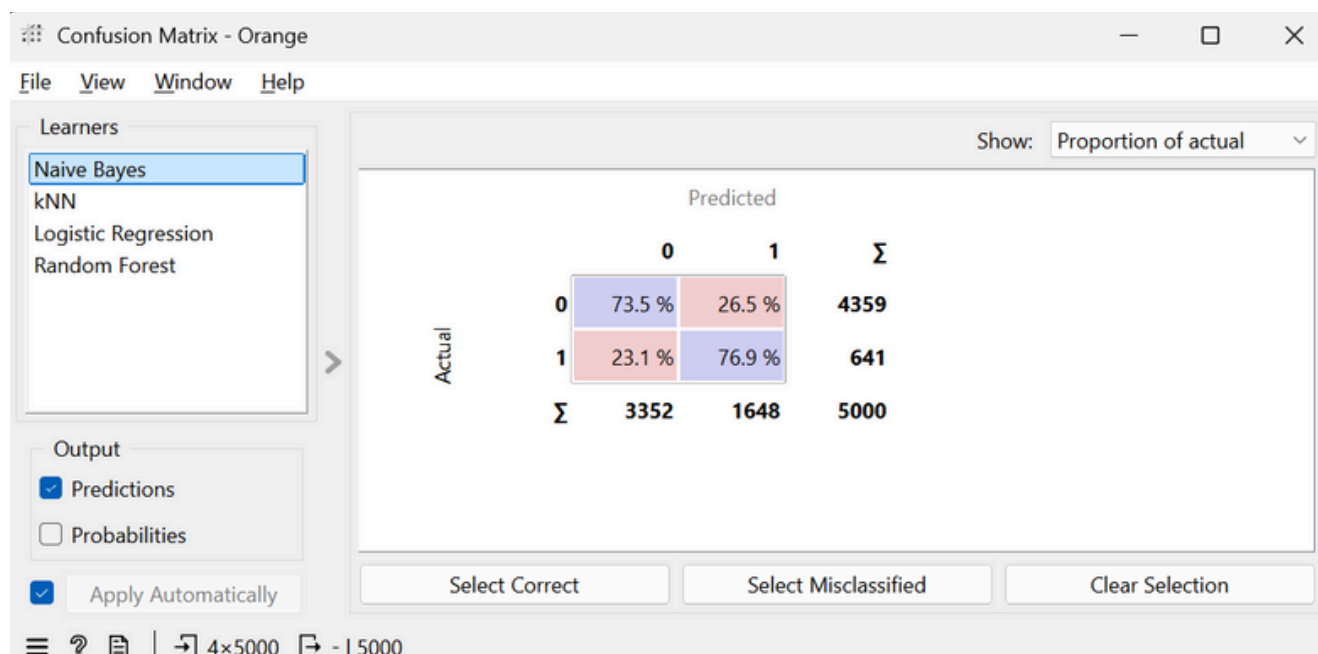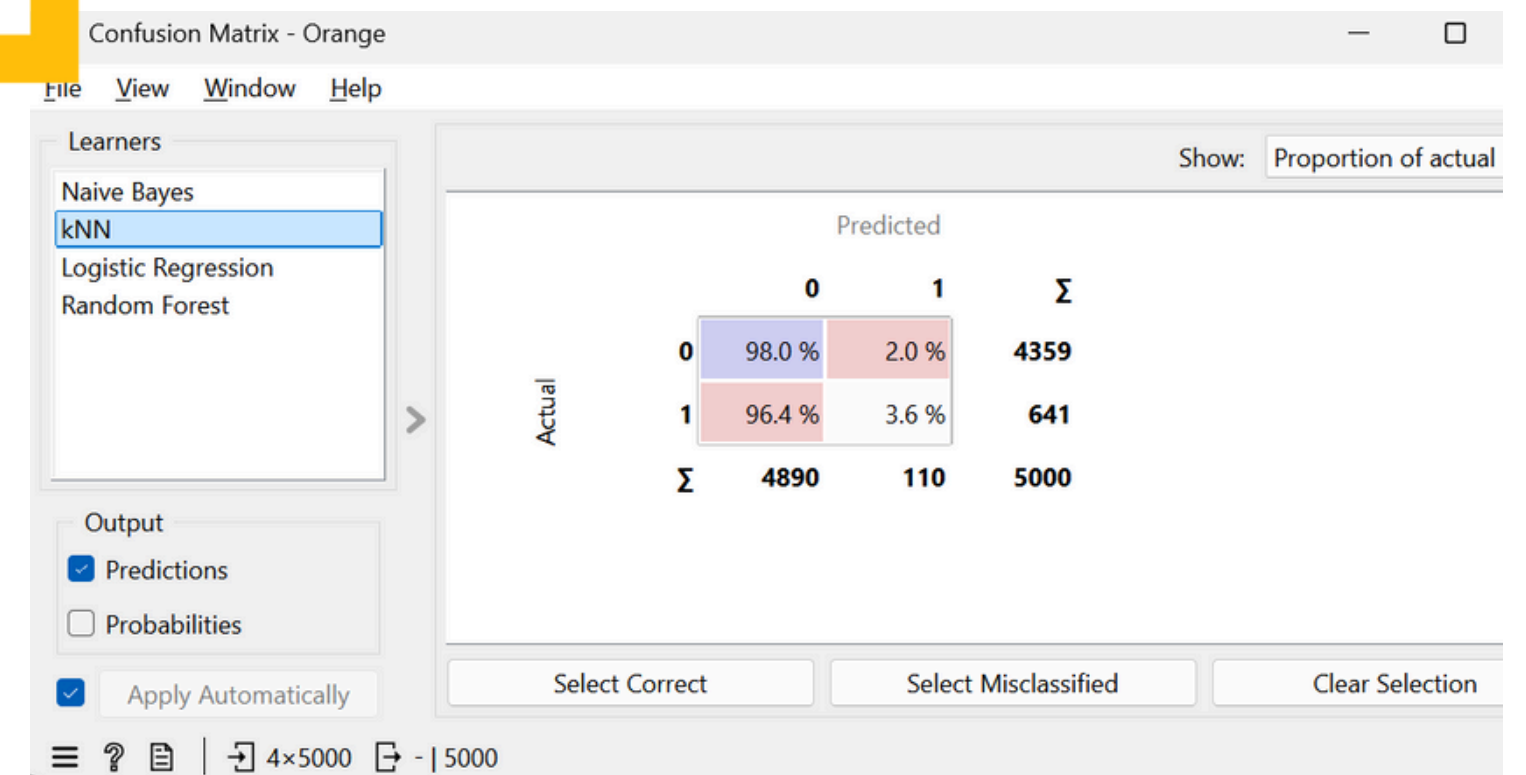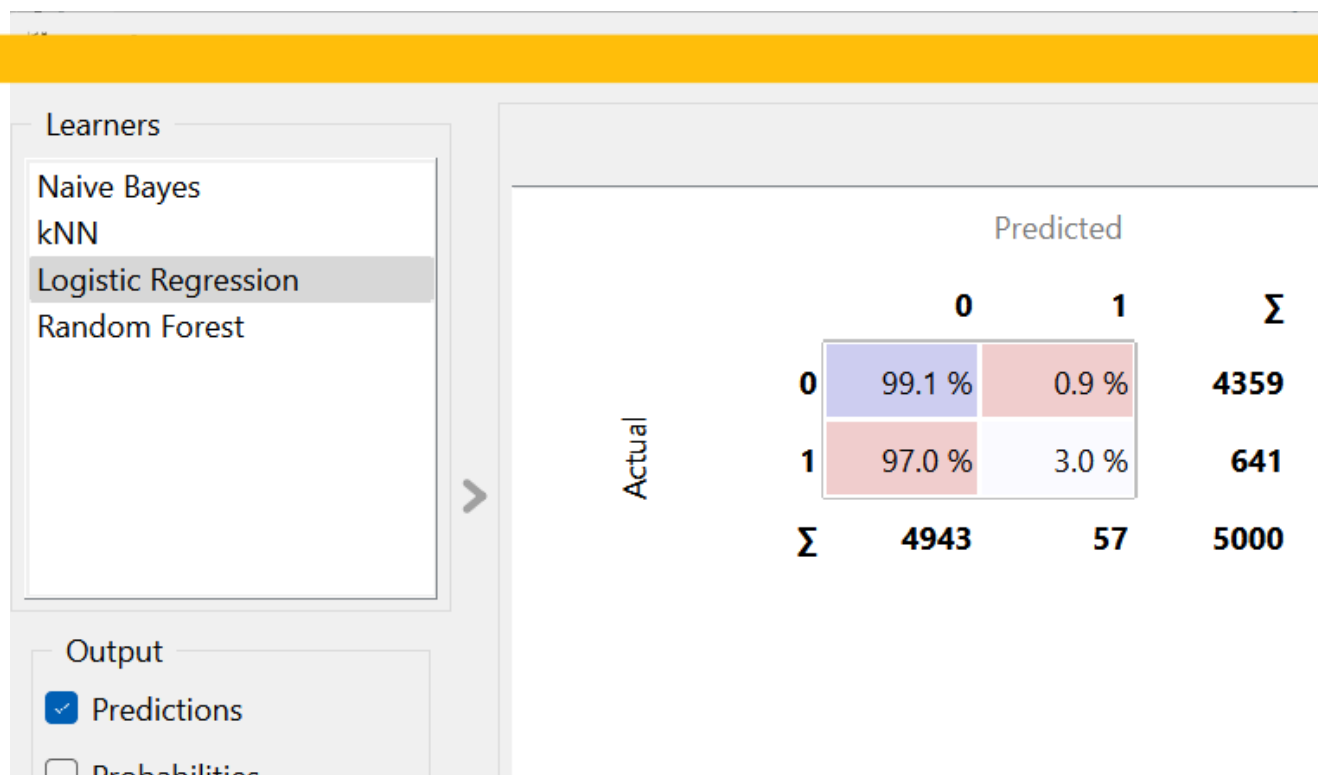27/06/25
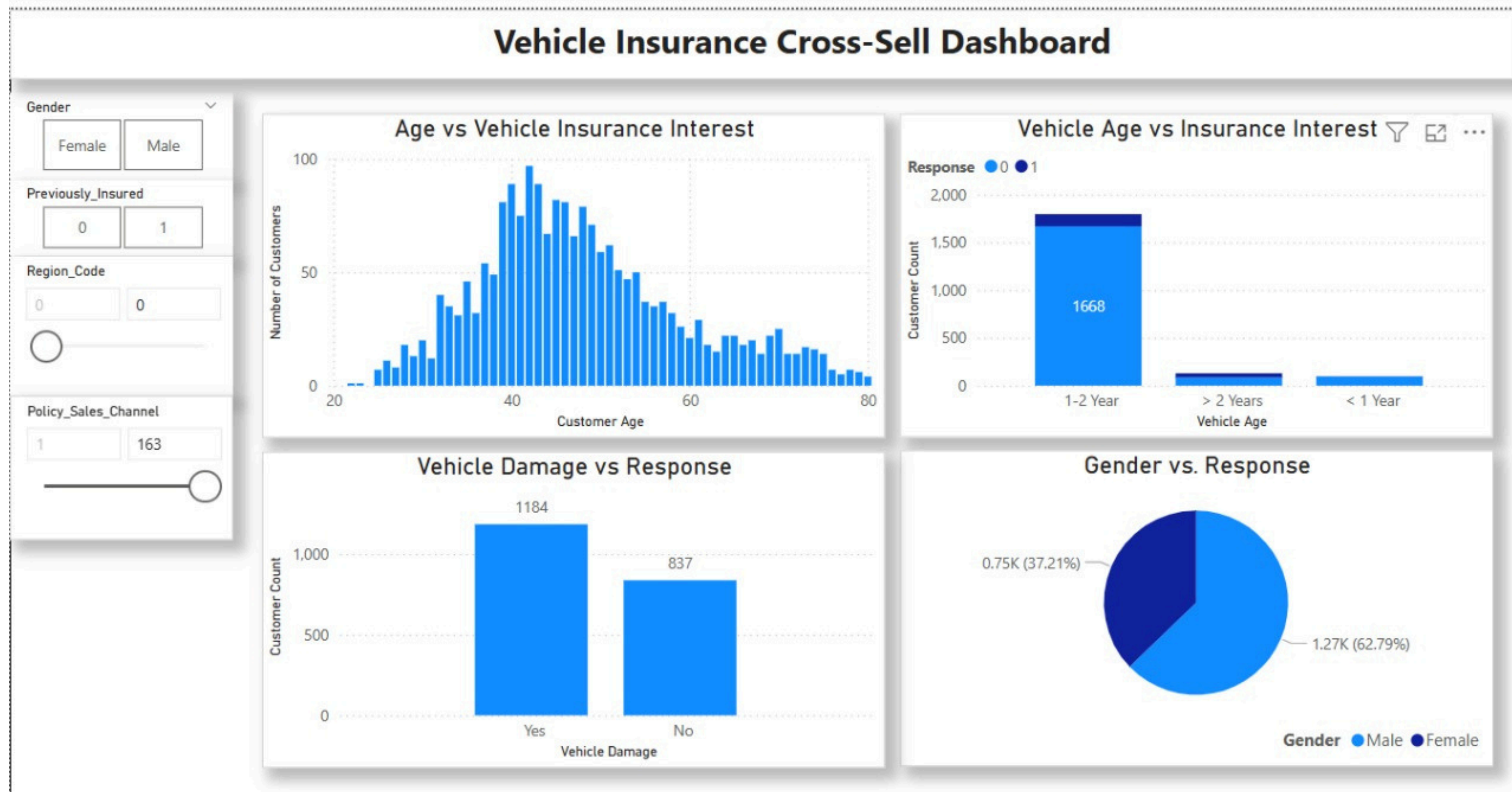
The Orange Modelling
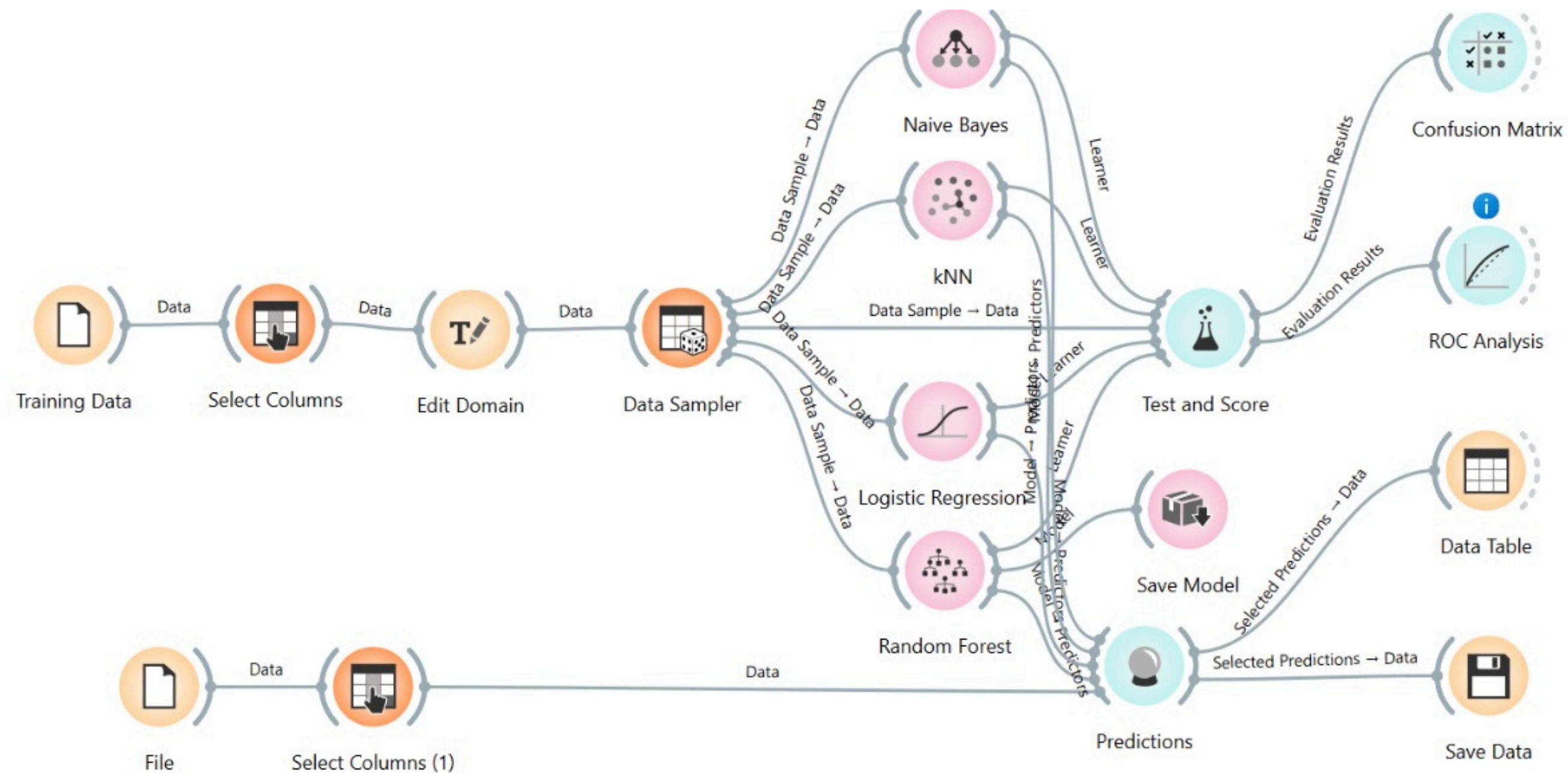
Results and Interpretation

ROC Curve
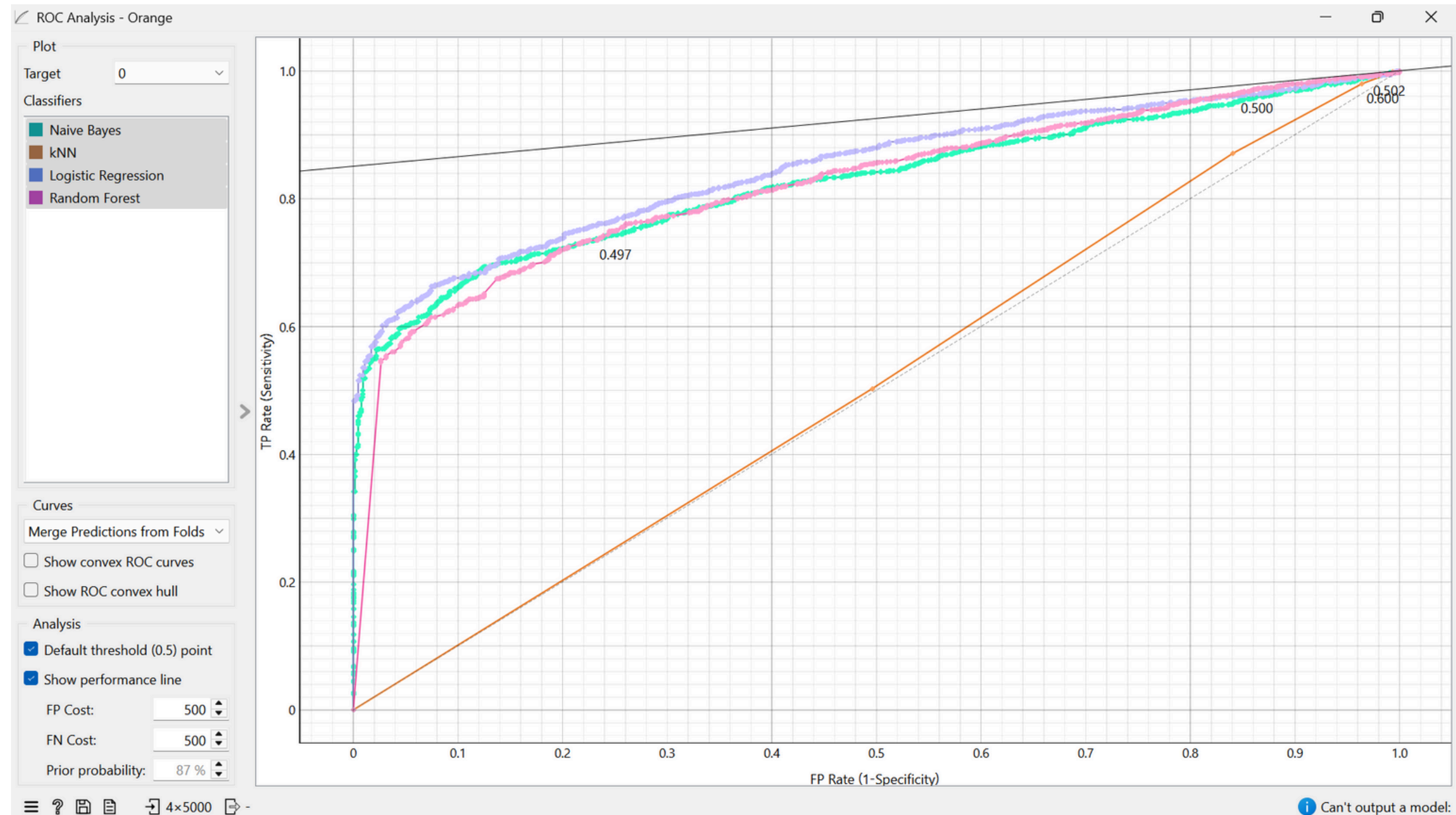
Analyzing the confusion matrix
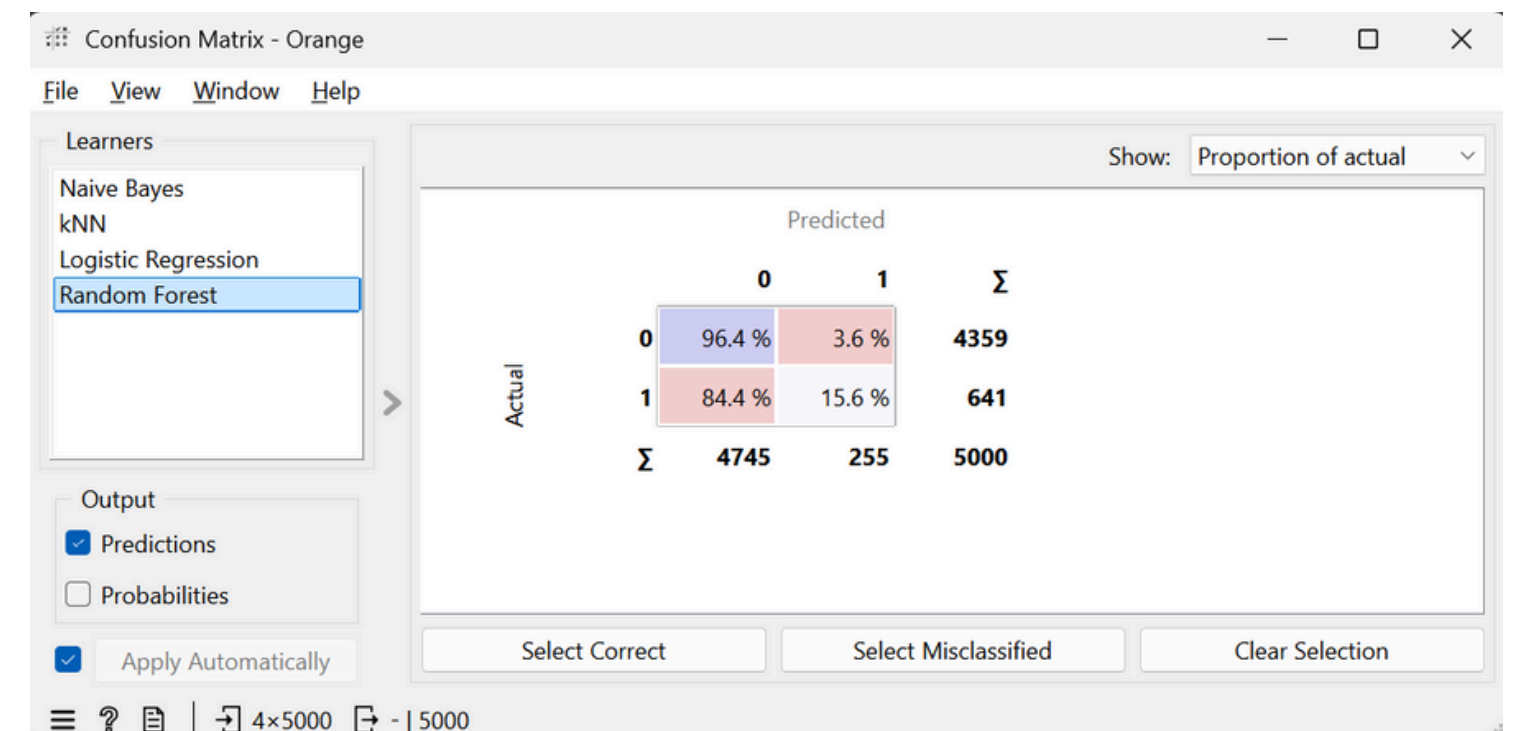
Key Learnings and Challenges Faced

Power BI
1. Drag-and-Drop Visualizations – Mastered creating charts (bar, line, maps) without writing DAX.
2. Power Query Efficiency – Automated data cleaning (removing duplicates, splitting columns) using the GUI.
3. Dashboard Design Best Practices – Learned layout principles (color contrast, hierarchy) for user-friendly reports.
4. Direct Data Sources – Connected seamlessly to Excel, SQL, and web APIs without manual coding.

Challenges Faced:
- Limited Custom Calculations
- Solution: Used built-in aggregations (sum/avg) and calculated columns in Power Query.
- Slow Refresh Times
- Fix: Reduced data volume with filters at the query level.
- Sharing/ Collaboration Issues
- Workaround: Published to Power BI Service and set up shared workspaces.

Key Learnings and Challenges Faced

Orange

Key Learnings:
1. No-Code Machine Learning – Built models (e.g., logistic regression) using pre-built widgets.
2. Data Preprocessing – Used Orange's interactive tools for normalization, missing values, and feature selection.
3. Visual Analytics – Explored patterns with heatmaps, box plots, and silhouette analysis.
4. Workflow Automation – Saved and reused workflows for similar datasets.

Challenges Faced:
- Limited Advanced Algorithms
- Solution: Combined Orange with Python scripts for extensions.
- Steep Learning Curve for Widgets
- Fix: Followed official tutorials and sample workflows.
- Handling Large Datasets
- Workaround: Sampled data or upgraded hardware.

References

1. Power BI
2. Microsoft Power BI Documentation
   a. Power Query Transformations
   b. Visualization Best Practice .
3. Community/Forums
   a. Power BI Community (for sharing/collaboration issues).

1. Orange
2. Official Orange Documentation
   a. Orange Widgets & Workflows
   b. Data Preprocessing Guide.
3. Tutorials
   a. Orange Data Mining YouTube Channel

Canva Study Materials

https://canvas.nus.edu.sg/

CORPORATE
GURUKUL

27/06/25