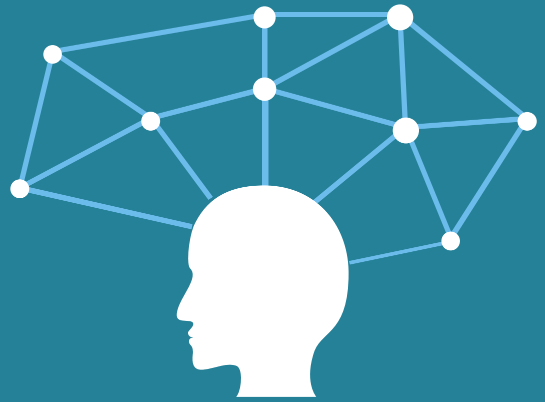
RETAINING THE ROCKS

What makes White Rock's Customers stay and/or go?

BC2406 Seminar Group 3 Team 7
Bairi Sahitya, Cai Xinrui, Ernest Ang Cheng Han, Malcolm Tan Yen Da

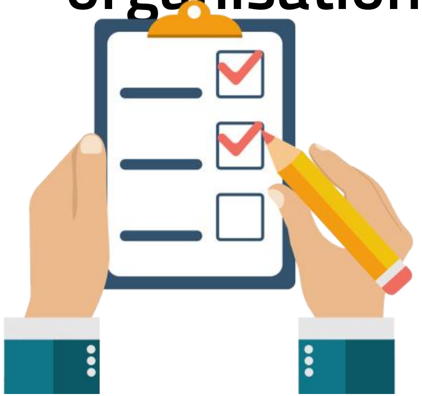
THE

BUSINESS PROBLEM



CUSTOMER RETENTION

One of the **key success metrics** of any organisation



Traditional Collection Method: Customer Satisfaction Survey

- Time Lag
- No objective way of verifying data
- Frequent surveys lead to customer burnout

Malcolm

THE BENEFITS OF RETENTION

High customer retention brings about **GREATER** profits through:

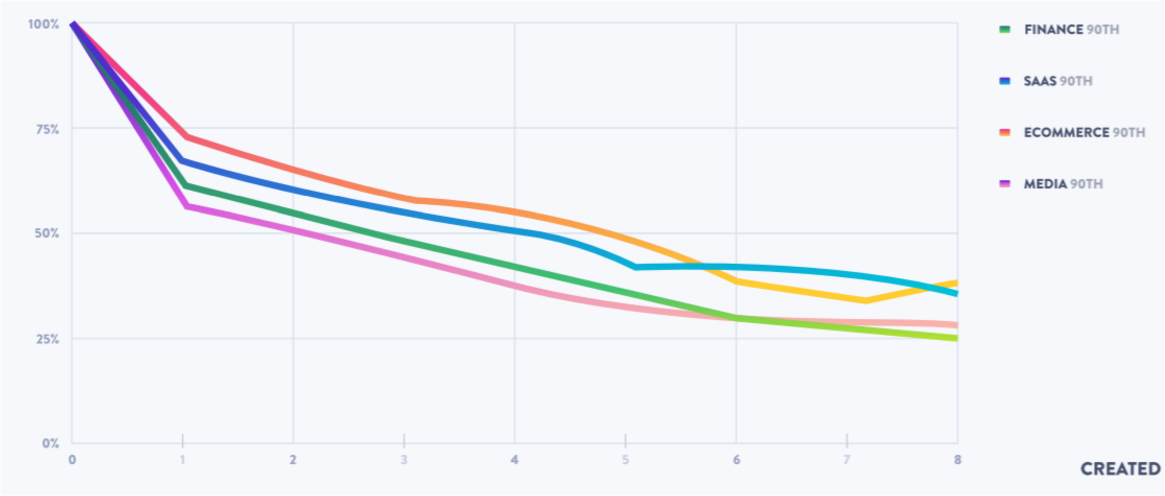
Higher Revenue: A retained customer is **60-70%** more likely to purchase

Lower Cost: **FREE** marketing through the word-of-mouth

Malcolm

WHY IS THIS RELEVANT?

Retention is particularly poor in the **Finance (Asset Management) Industry**; likely caused by **huge supply of investment firms.**



Malcolm

WHAT AFFECTS IT

INTERNAL VARIABLES

- **Have** control over
- Eg: NumberOfComplaints

EXTERNAL VARIABLES

- **No** control over
- Eg: Gender

Malcolm

THE BUSINESS PROBLEM

The focus of our analysis and proposed solution will be on...

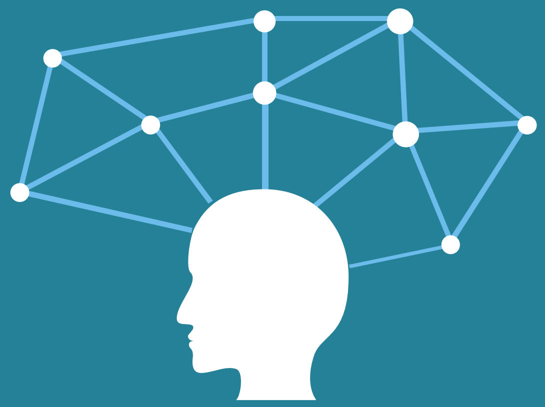
CUSTOMER RETENTION

and their **IMPORTANT FACTORS**

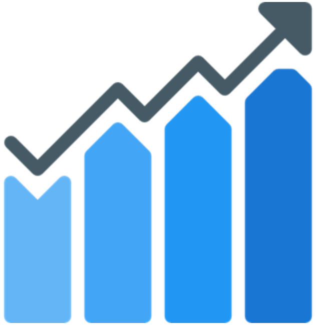
for **WHITE ROCK**

Malcolm

THE ANALYTICS SOLUTION



AUTOMATED ANALYTICS PROCESS



Increasing trend of companies utilizing data-driven processes to draw meaningful insights

Opportunity: To provide faster and more accurate retention predictions

Malcolm

RETENTION ANALYSIS MODEL

1. Identifies **customers** of White Rock who are at high risk of leaving

1. Identify **factors** that have the greatest impact on customer retention

Malcolm

PROJECT FEASIBILITY

01

Predictive Need:

Retention and factors for timely intervention

02

Imperfect Knowledge:

Lack of full awareness of customer's intentions

03

Available Training Data:

Credible and Relevant

Malcolm

DESIRED BUSINESS OUTCOMES

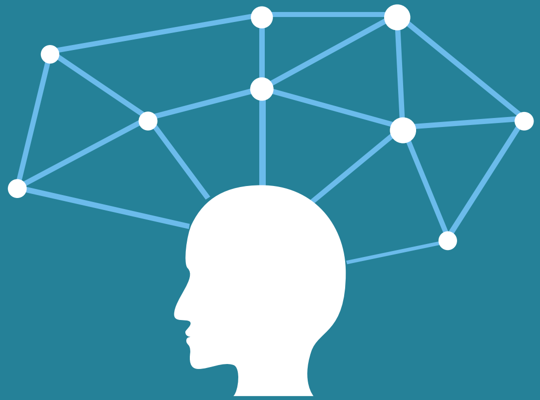
Identifying Customers for Targeted Retention Efforts

Improving Resource Allocation and Reducing Cost

Malcolm

DATA EXPLORATION

INSIGHTS

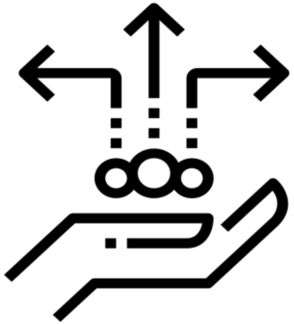


Data Preparation

Original Dataset

10000 rows and 14 columns

Primary key attribute = CustomerID



Created Variables

8 variables

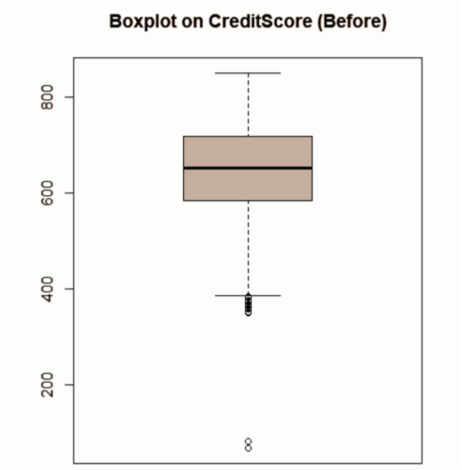
7 internal + 1 external

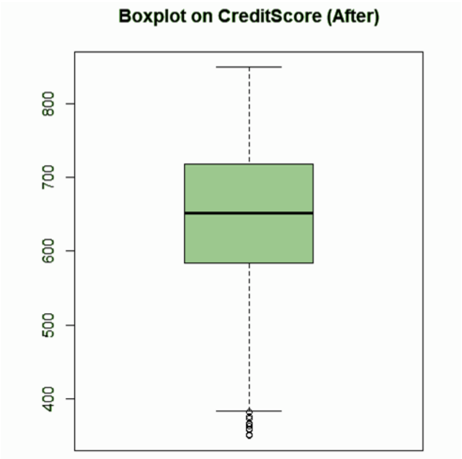
Xin Rui

Data Cleaning

1. Outliers in Credit Score

Resolved by correcting the values.



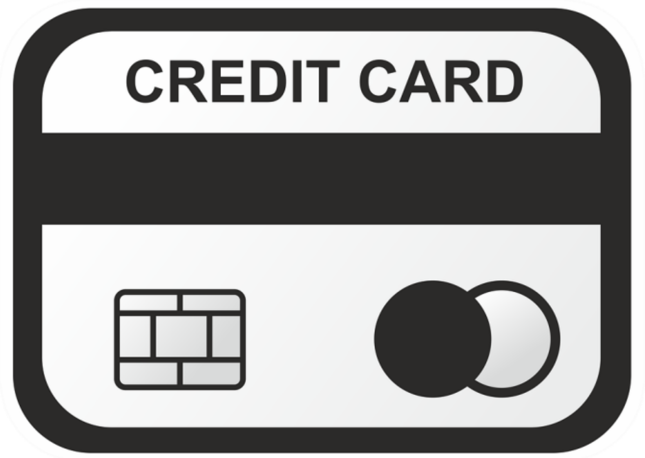


Xin Rui

Data Cleaning

2.Redundant Data (HasCreditCr)

Resolved by removing the column



Data Cleaning

3. NAs in variable Financial Literacy and Last Contact by a Banker

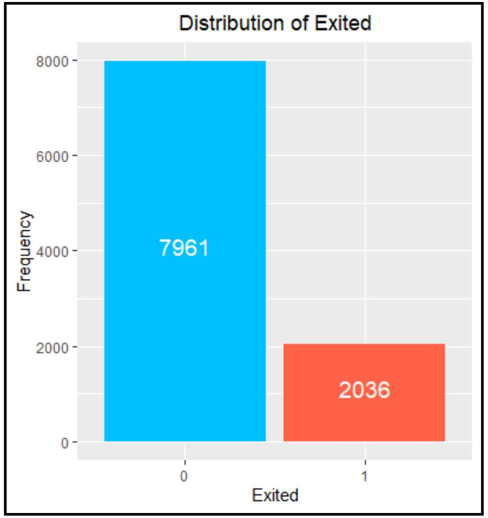
Resolved by removing the 3 rows

N.A.

Xin Rui

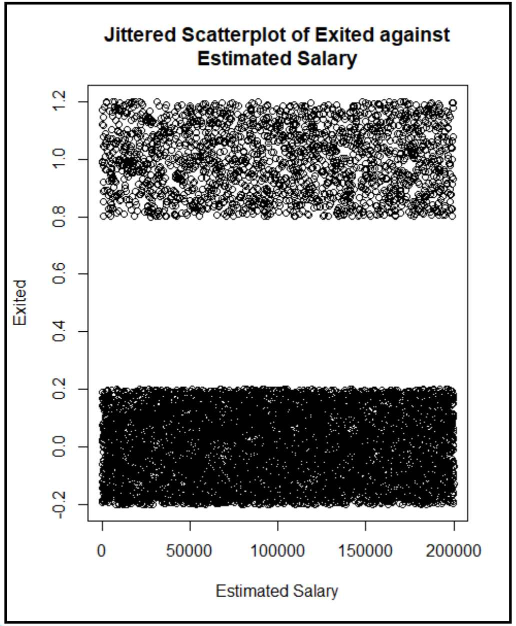
Distribution of Exited

- Binary output variable
- 20.4% versus 79.6%
- **Stratified** during train-test split in Logistic Regression model

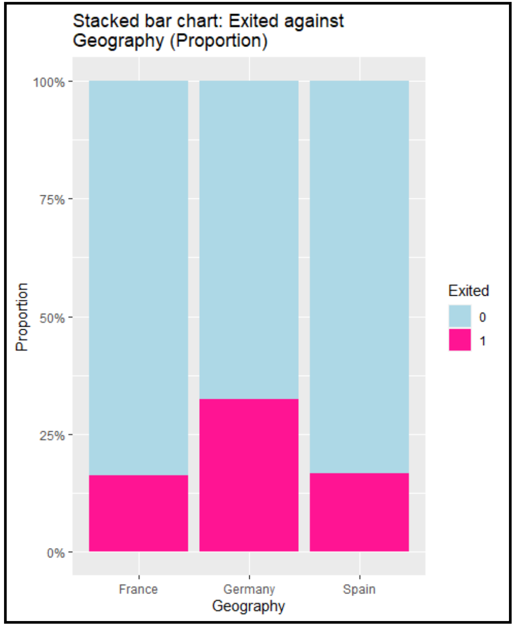


Xin Rui

Findings in the original dataset



- German branch has more customers leaving
- Estimated Salary is uniformly distributed so it should not be an influential factor

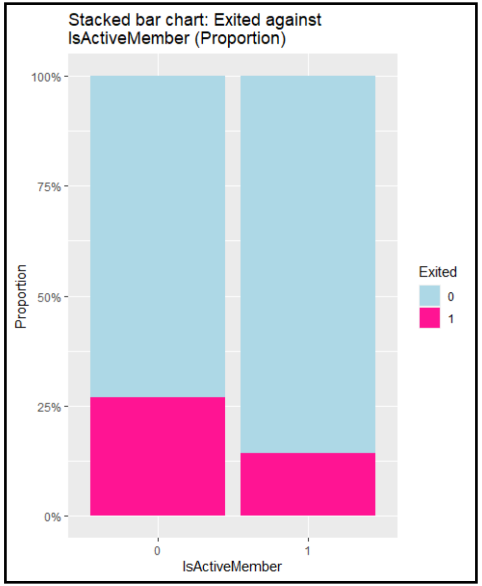


Xin Rui

Findings in the original dataset

- The only internal variable **"IsActiveMember"** showed that Active members have a **lower chance of leaving**, half of that as compared to inactive members.

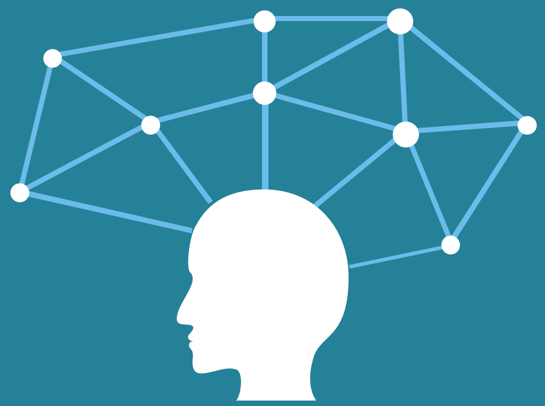
- 26.9% versus 14.3%



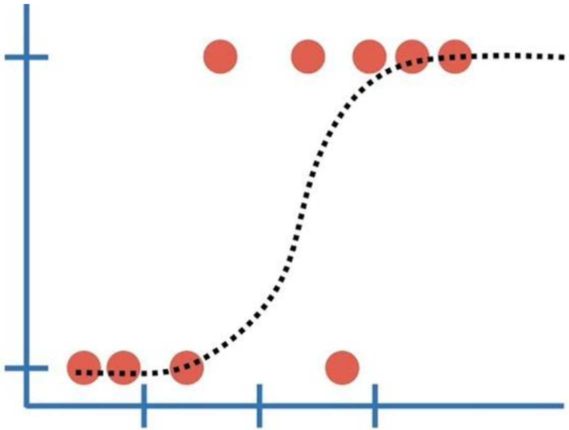
Xin Rui

THE MODEL:

LOGISTIC REGRESSION



Logistic Regression Overview

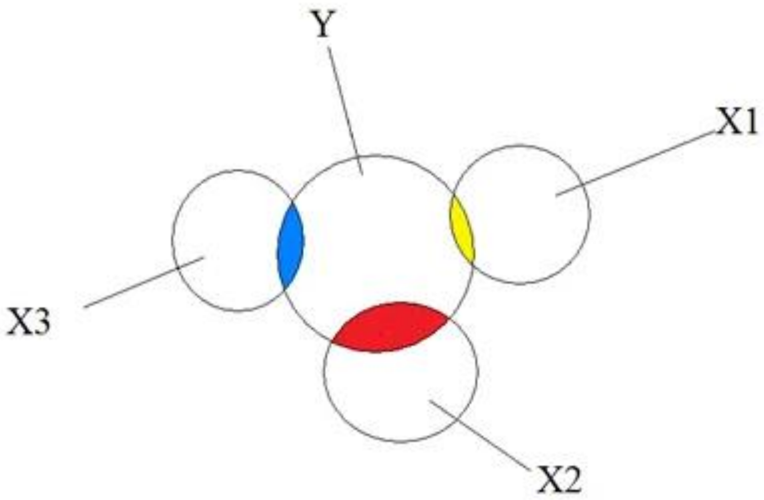


- Backward Elimination was run on all the variables
- Final model consisted of only input variables, are statistically significant at 5% α
- Clustering was performed but no clear cluster was formed

Xin Rui

Multicollinearity Issues

The **Adjusted GVIF values** of the final model are all **below 2**, suggesting that there are no multicollinearity issues between the explanatory variables.



Results

Predicted Values

Actual Values

	Retained (Negative)	Exit (Positive)
Retained (Negative)	2339 (78%)	49 (1.6%)
Exit (Positive)	84 (2.8%)	527 (17.6%)

Accuracy: 95.6% **Type I Error: 2.05%** **Type II Error: 13.8%**

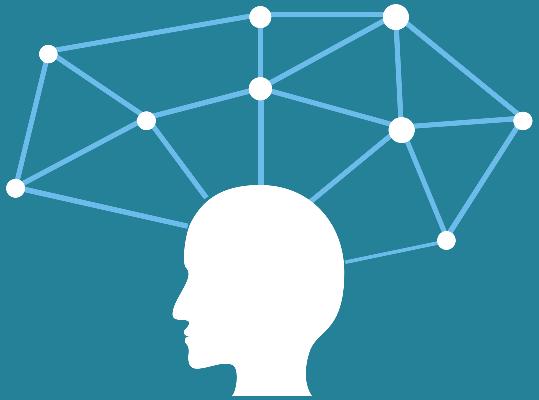
Xin Rui

Insights (Top 5 significant factors)

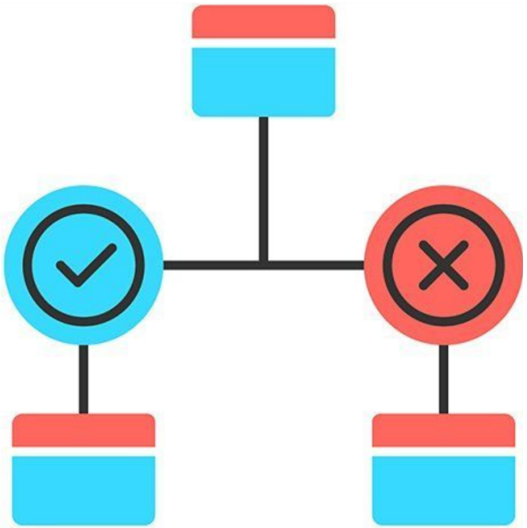
Xin Rui

THE MODEL:

CLASSIFICATION & REGRESSION TREE (CART)



CART Overview



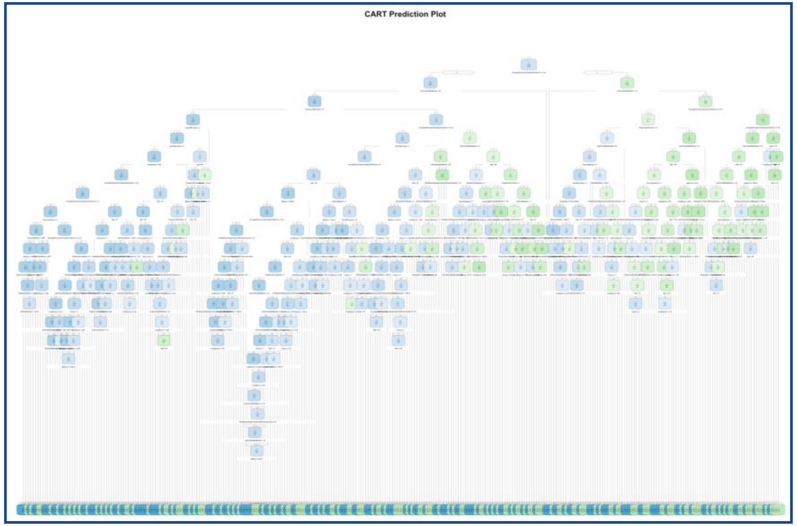
- Utilizes classification & regression to make predictions

- Trains & tests data set via 10 fold cross validation algorithm

- Generating model:
  - Phase 1: Growing to maximum
  - Phase 2: Pruning to minimum

Ernest

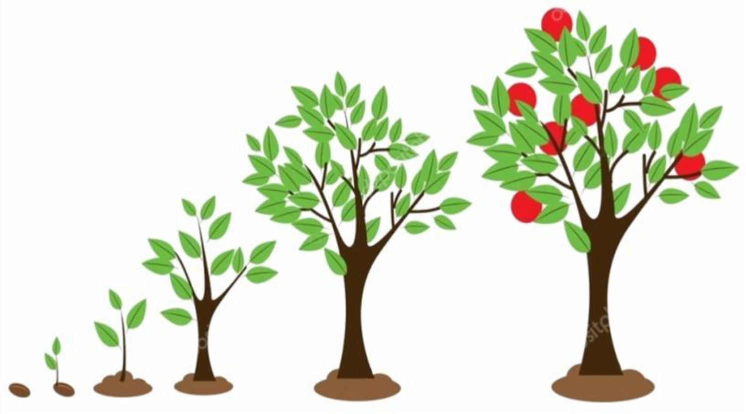
CART Phase 1: Growing



- Selecting best split point at each node of CART model
- Each node produces 2 child nodes of the highest possible purity
- Process continues till a lenient stopping condition is met
- Problem: Overfitting!

Ernest

CART Phase 1: Growing



- Due to the many variables used, problem of overfitting is observed
- CART was grown to its maximum by setting the complexity penalty to 0 units per terminal node
- Hence, CART needs to be pruned to its minimum size
- Problem: Where?

Ernest

CART Phase 2: Pruning

- Answer: Weakest Link
- Prune the tree at its weakest link to obtain the most optimal tree
- Weakest Link definition: minimum value of complexity penalty that would trigger pruning
- Problem: How?

CART Phase 2: Pruning

01

Step 1:
Identify tree with lowest cross validation error

02

Step 2:
Calculate the cross validation error cap

03

Step 3:
Identify simplest tree (nearest node) whose cross validation error is below the error cap

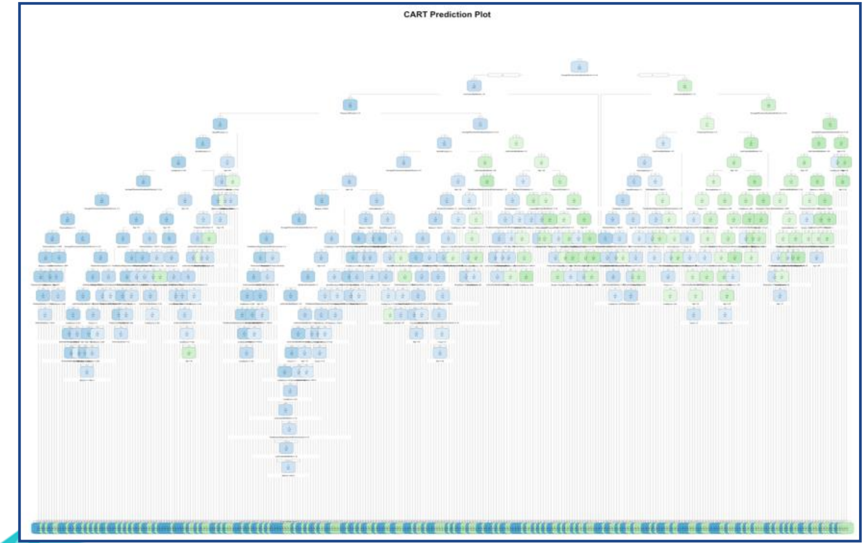
04

Step 4:
Minimum complexity penalty derivation: geometric complexity penalty mean of simplest tree

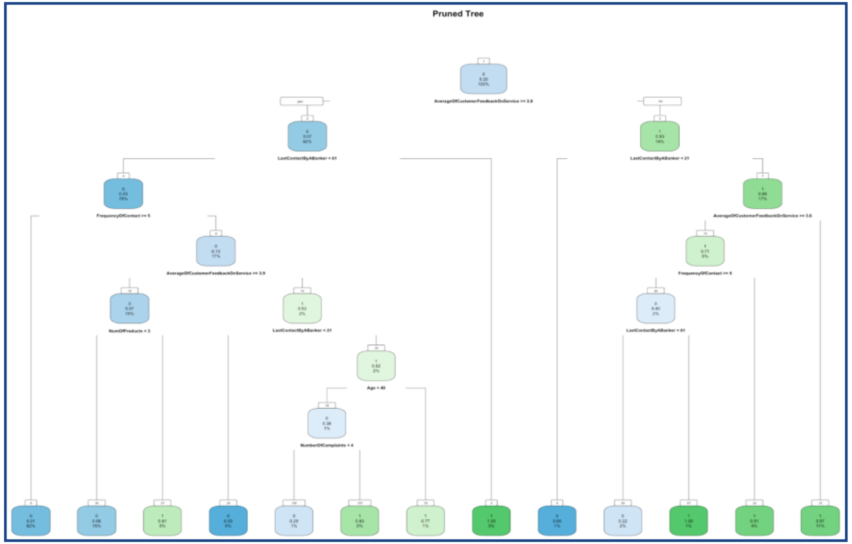
Ernest

CART Phase 2: Pruning

Before:



After:



Ernest

CART Prediction Results

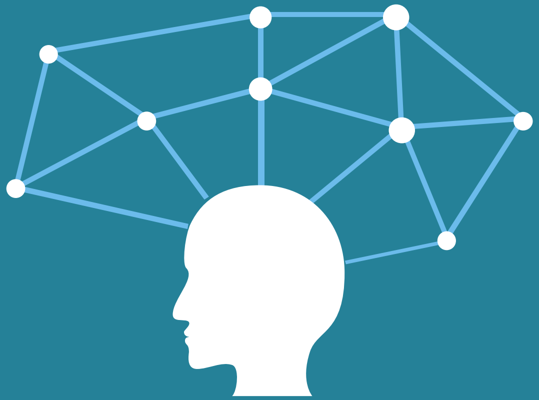
Predicted Values

Actual Values

	Retained (Negative)	Exit (Positive)
Retained (Negative)	7865 (78.67%)	96 (0.96%)
Exit (Positive)	203 (2.03%)	1833 (18.34)

Accuracy: 97.0% **Type I Error: 1.21%** **Type II Error: 9.97%**

Ernest



THE MODEL

EVALUATION

Model Selection & Evaluation

CART

Logistic Regression

Ernest

Model Selection & Evaluation

CART

Accuracy Rate: **97.0%**

Error Rate: 3.00%

Type I Error: 1.21%

Type II Error: **9.97%**

Logistic Regression

Accuracy Rate: 95.6%

Error Rate: 4.40%

Type I Error: 2.05%

Type II Error: 13.8%

Ernest

Model Selection & Evaluation

1. Having a personal advisor
1. Male gender
1. Located in Germany, Spain
1. Active member
1. High average customer feedback

Customer Retention Factors

1. Female gender
1. Older age
1. Lower estimated salary
1. High financial literacy
1. Low credit score

Customer Churn Factors

Ernest

Literature Review & Expert Opinion

According to a study, the following factors influenced them to continue using their Bank's services:

1. Satisfaction with services

1. Developing personal friendships with staff

1. High financial literacy (Contradiction!)

Ernest

Overall Evaluation & Conclusion

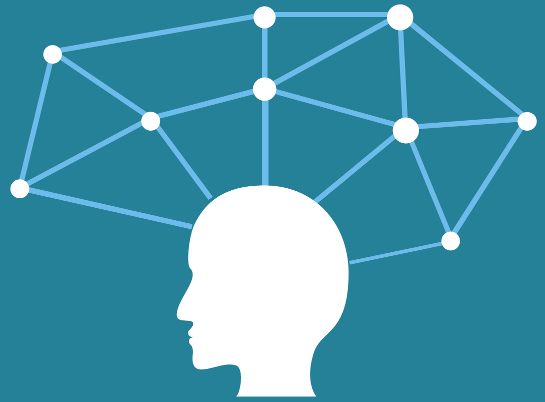
1. CART model has a higher prediction accuracy and its results are quite applicable

1. Requires further support from research and industrial knowledge when implementing recommendations from CART

Ernest

IMPLEMENTING

OUR SOLUTION



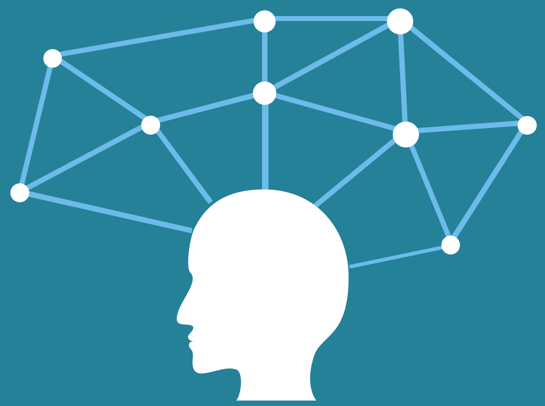
Interactive dashboard

- Displays **churn prediction** of each customer
- Displays **significant variables** affecting retention
- New factors explored frequently
- Data should be updated regularly

Sahitya

OUR

RECOMMENDATIONS



Identify important business processes

Feedback

Understand the reasons for high and low feedback from past and current customers

Address main concerns

Personal Advisor

Invest in more personalised service

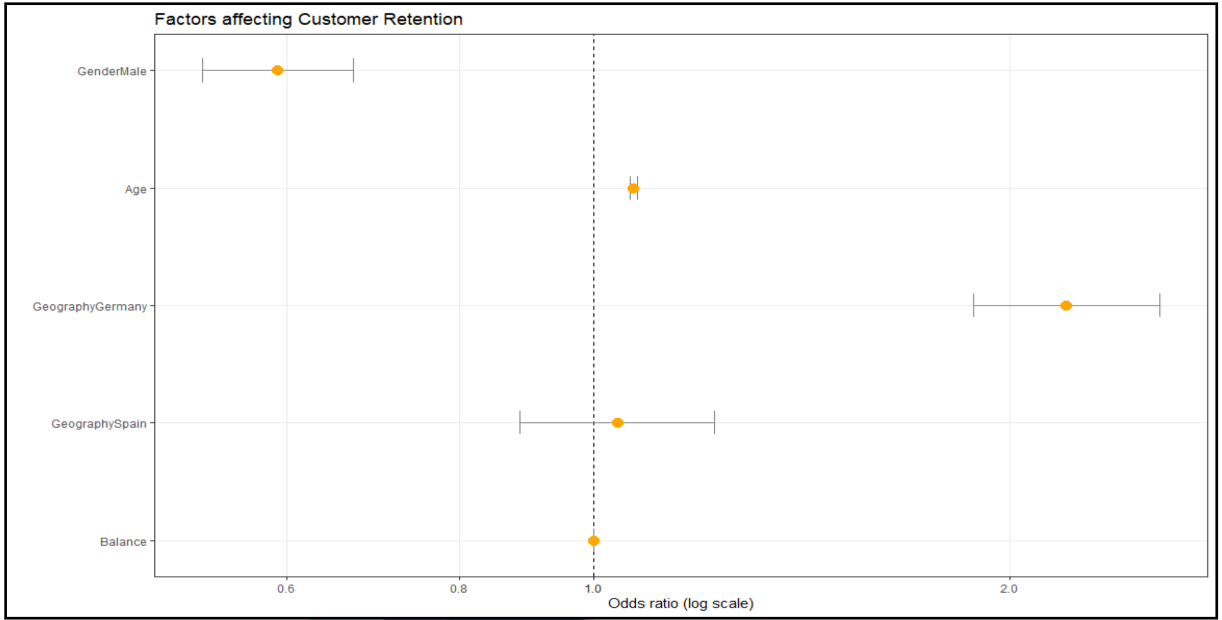
Membership usage

Market the benefits of the membership card

Sahitya

Identify favourable customers

Younger Males

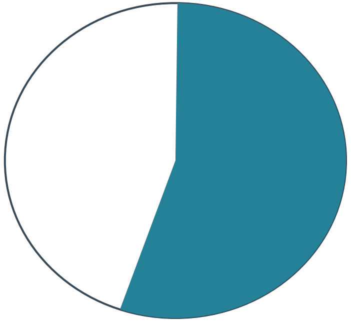


Sahitya

Identify favourable customers

Younger Males

Address concerns of less favourable groups



59% of females

“Being able to provide for their family/children and about security and comfort”

Sahitya

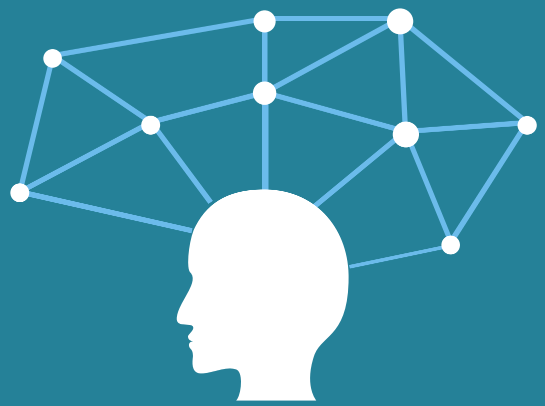
Impact of Internal vs External factors

79.3% vs 95%

Sahitya

EXTENDING

OUR SOLUTION



Creating a Personalised Solution

- Understand how the significant internal retention predictors vary with each of the external variables

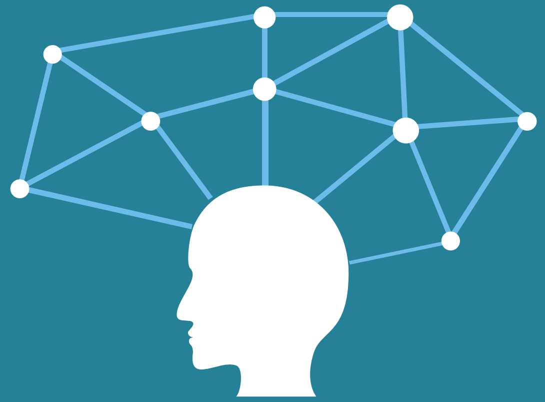
vs

- "NumberOfComplaints" is a statistically significant predictor
- **Odds Ratio** of "IsActiveMember" more significant

Sahitya

THE

LIMITATIONS



Limitations

- Inaccurate and **non-representative** data
- Cannot rely purely on statistical model

Future Research Directions

- Explore more predictors
- Include a wider spectrum of categories
- Collate other types of data
- Utilisation of more advanced artificial intelligence models

Sahitya

Q & A