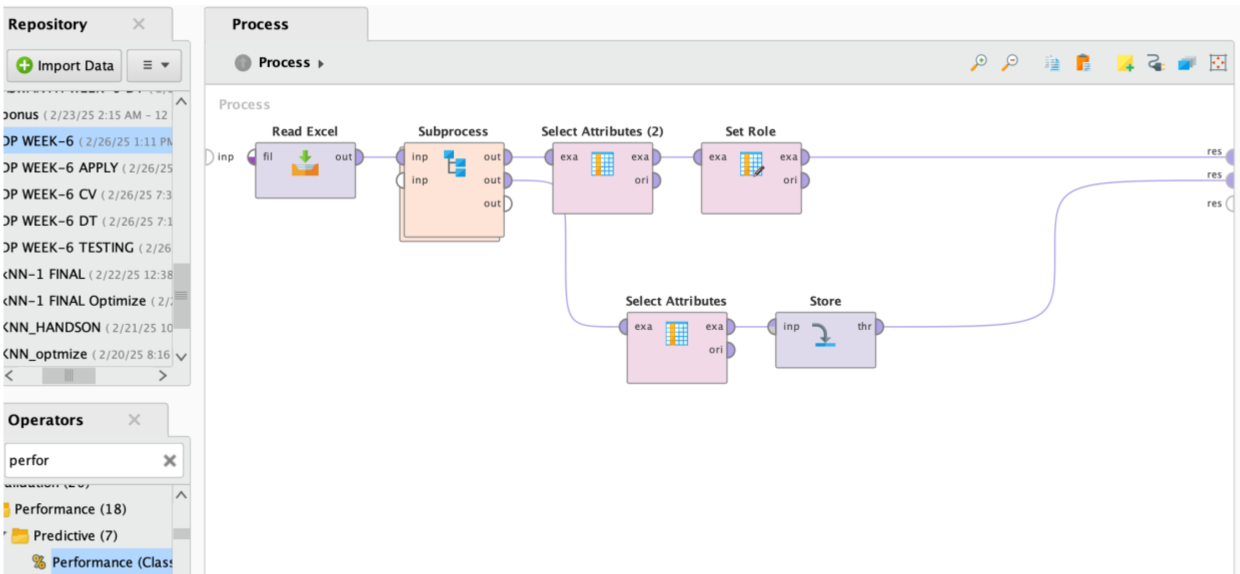
Machine Learning for Business Analytics

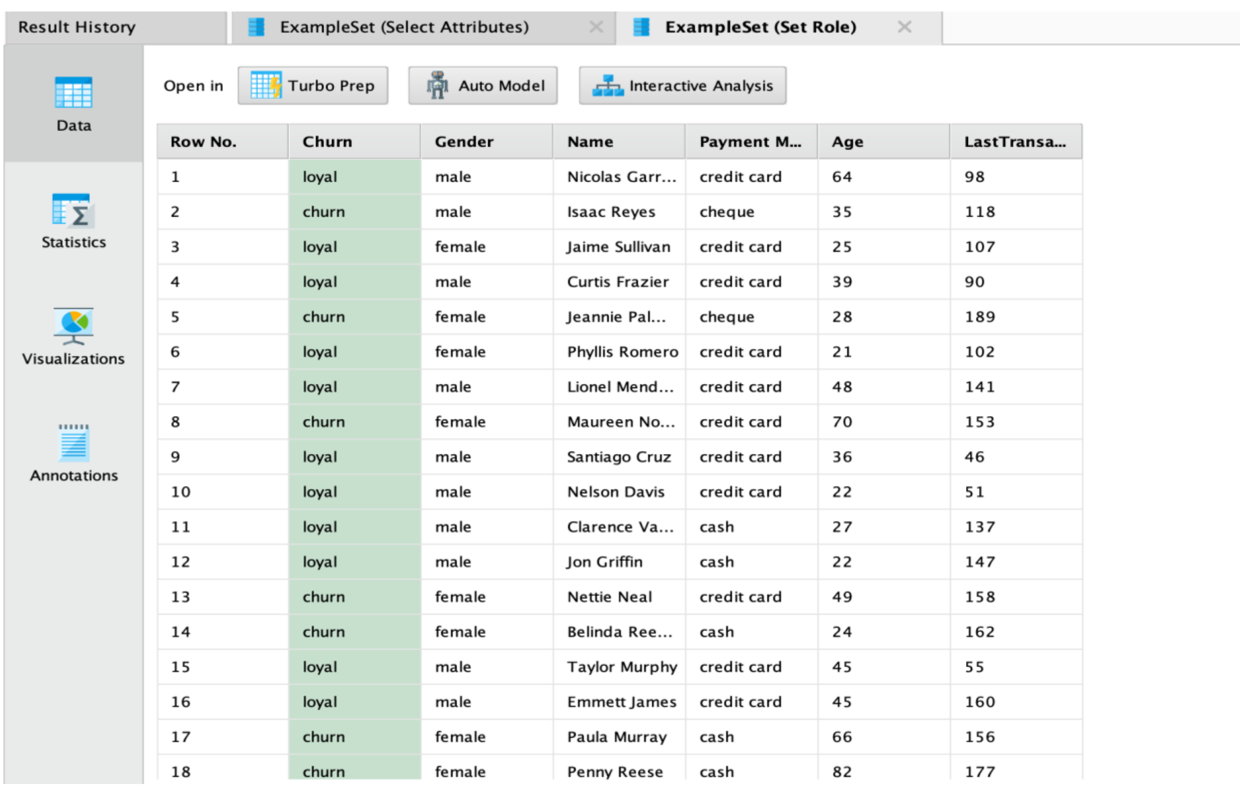
Prem Kumar Chimakurthi

Master of Science in Business Analytics

Customer churn Project

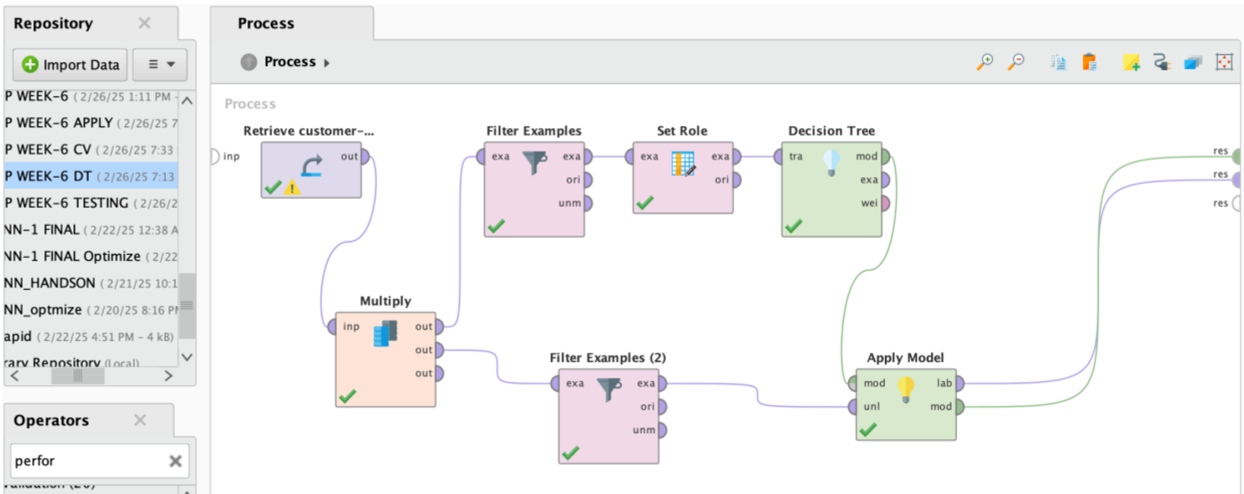
SET ROLE:

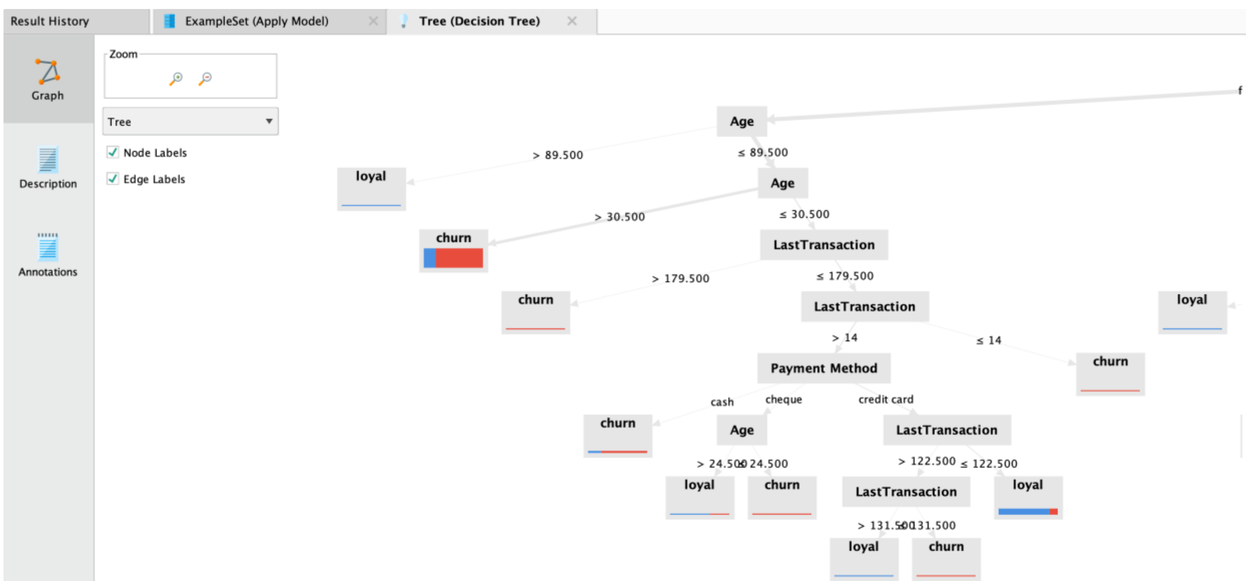




Row No.	Churn	Gender	Name	Payment M...	Age	LastTransa...
1	loyal	male	Nicolas Garr...	credit card	64	98
2	churn	male	Isaac Reyes	cheque	35	118
3	loyal	female	Jaime Sullivan	credit card	25	107
4	loyal	male	Curtis Frazier	credit card	39	90
5	churn	female	Jeannie Pal...	cheque	28	189
6	loyal	female	Phyllis Romero	credit card	21	102
7	loyal	male	Lionel Mend...	credit card	48	141
8	churn	female	Maureen No...	credit card	70	153
9	loyal	male	Santiago Cruz	credit card	36	46
10	loyal	male	Nelson Davis	credit card	22	51
11	loyal	male	Clarence Va...	cash	27	137
12	loyal	male	Jon Griffin	cash	22	147
13	churn	female	Nettie Neal	credit card	49	158
14	churn	female	Belinda Ree...	cash	24	162
15	loyal	male	Taylor Murphy	credit card	45	55
16	loyal	male	Emmett James	credit card	45	160
17	churn	female	Paula Murray	cash	66	156
18	churn	female	Penny Reese	cash	82	177

Decision Tree:





Key Points from Decision Tree Analysis:

1. Top Predictors:

    a. **Age** and **Last Transaction** are the most influential factors in predicting **Churn**.

2. Churn Patterns:

   a. Younger customers with **low Last Transaction** values are more likely to **churn**.

   b. Older customers with **high Last Transaction** values show more **loyalty**.

3. Payment Method Impact:

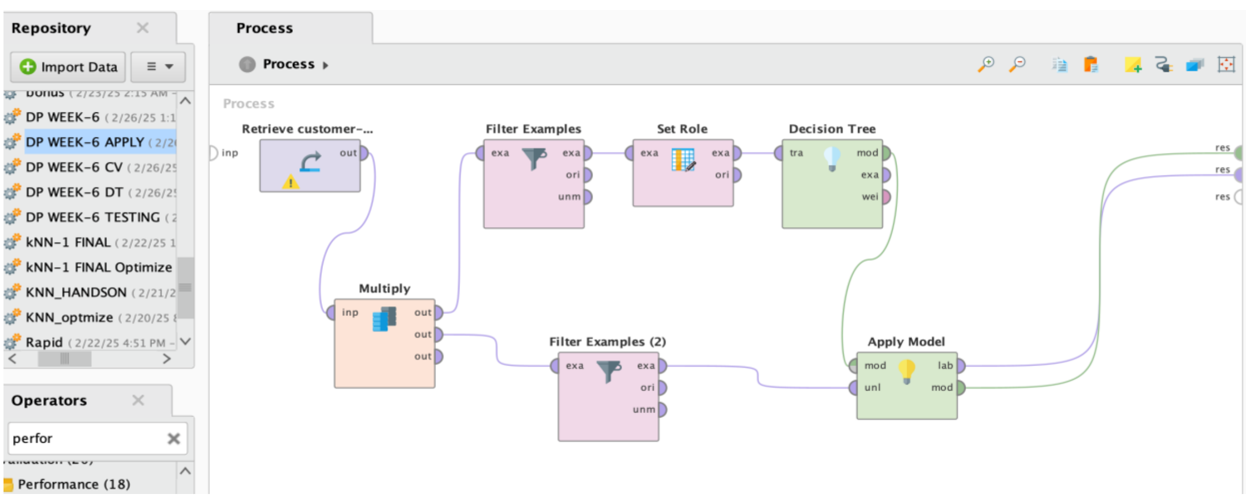
   a. **Cash** users have a higher chance of **churn**.

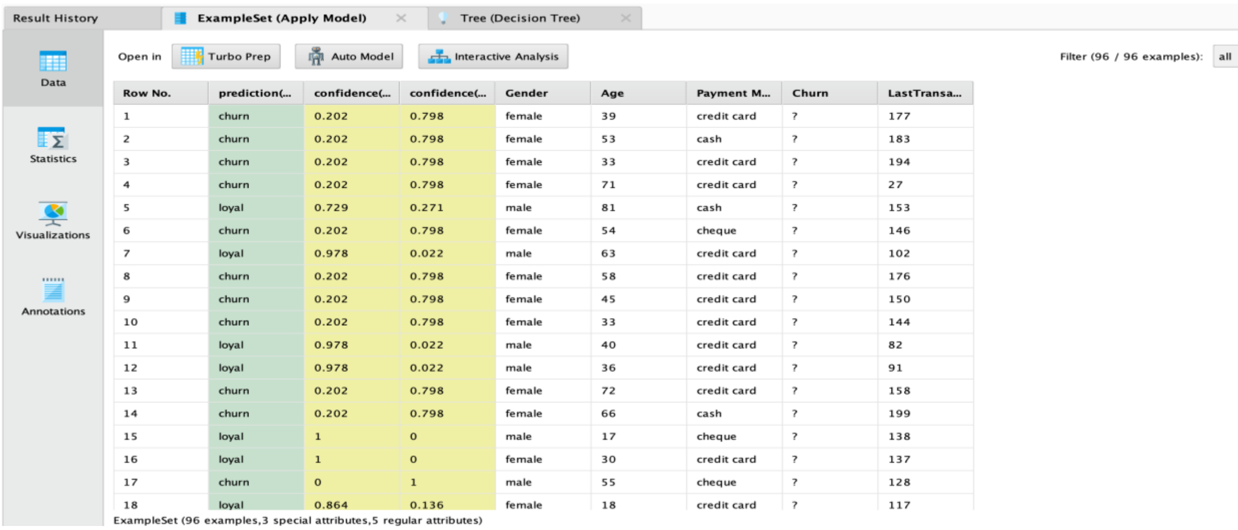
   b. **Credit card** users with moderate transactions are loyal.

4. Clear Decision Rules:

   a. Example: If **Age > 89.5**, then **loyal**; if **Age ≤ 89.5** and **Last Transaction > 179.5**, then **churn**.

Apply Model:





Row No.	prediction(...	confidence(...	confidence(...	Gender	Age	Payment M...	Churn	LastTransa...
1	churn	0.202	0.798	female	39	credit card	?	177
2	churn	0.202	0.798	female	53	cash	?	183
3	churn	0.202	0.798	female	33	credit card	?	194
4	churn	0.202	0.798	female	71	credit card	?	27
5	loyal	0.729	0.271	male	81	cash	?	153
6	churn	0.202	0.798	female	54	cheque	?	146
7	loyal	0.978	0.022	male	63	credit card	?	102
8	churn	0.202	0.798	female	58	credit card	?	176
9	churn	0.202	0.798	female	45	credit card	?	150
10	churn	0.202	0.798	female	33	credit card	?	144
11	loyal	0.978	0.022	male	40	credit card	?	82
12	loyal	0.978	0.022	male	36	credit card	?	91
13	churn	0.202	0.798	female	72	credit card	?	158
14	churn	0.202	0.798	female	66	cash	?	199
15	loyal	1	0	male	17	cheque	?	138
16	loyal	1	0	female	30	credit card	?	137
17	churn	0	1	male	55	cheque	?	128
18	loyal	0.864	0.136	female	18	credit card	?	117

ExampleSet (96 examples,3 special attributes,5 regular attributes)

Key Points from Apply Model Results:

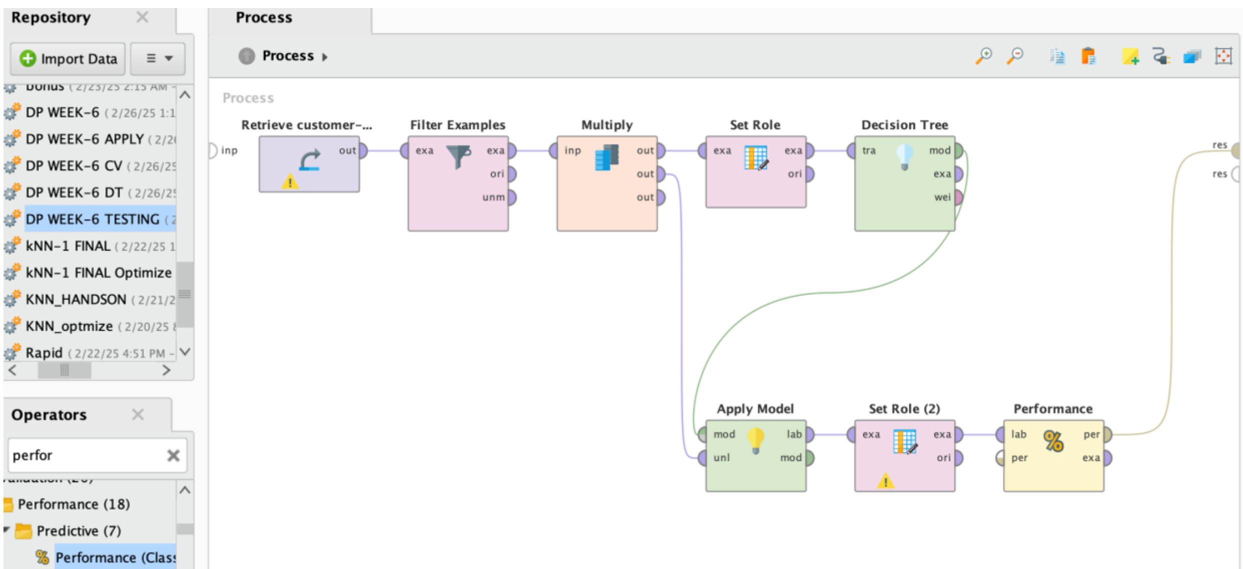
1. Successful Prediction:
   a. The **Decision Tree** model successfully predicted **Churn** values for all 96 missing entries.
2. Prediction Labels:
   a. Predicted values include both **churn** and **loyalty**, effectively filling the missing **Churn** column.
3. **C**onfidence Scores:
   a. High confidence values indicate reliable predictions.
   b. Example: Confidence for **loyalty** is as high as **0.978**, while for **churn** it is around **0.202** to **0.864**.
4. Consistent Patterns:
   a. **Churn** is more likely for younger users or those with low **LastTransaction** values.

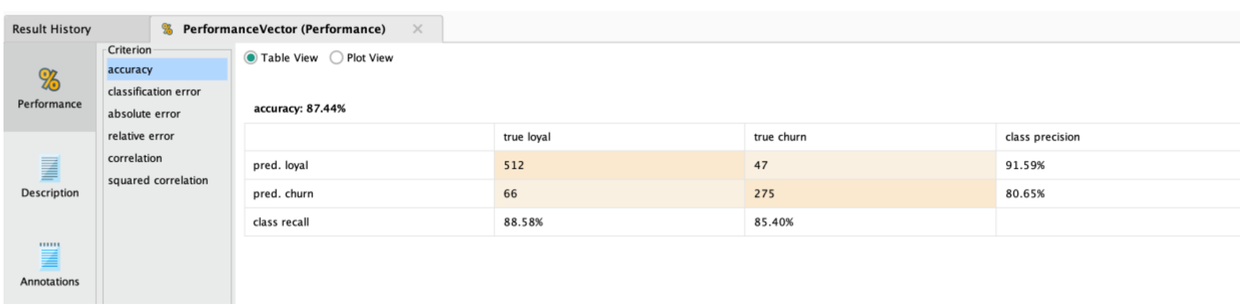
b. **Loyal** predictions are common for older customers and those with high **LastTransaction** values.

5. <u>Influential Attributes:</u>

   a. **Age**, **LastTransaction**, and **Payment Method** played a crucial role in predictions.

<u>Testing:</u>





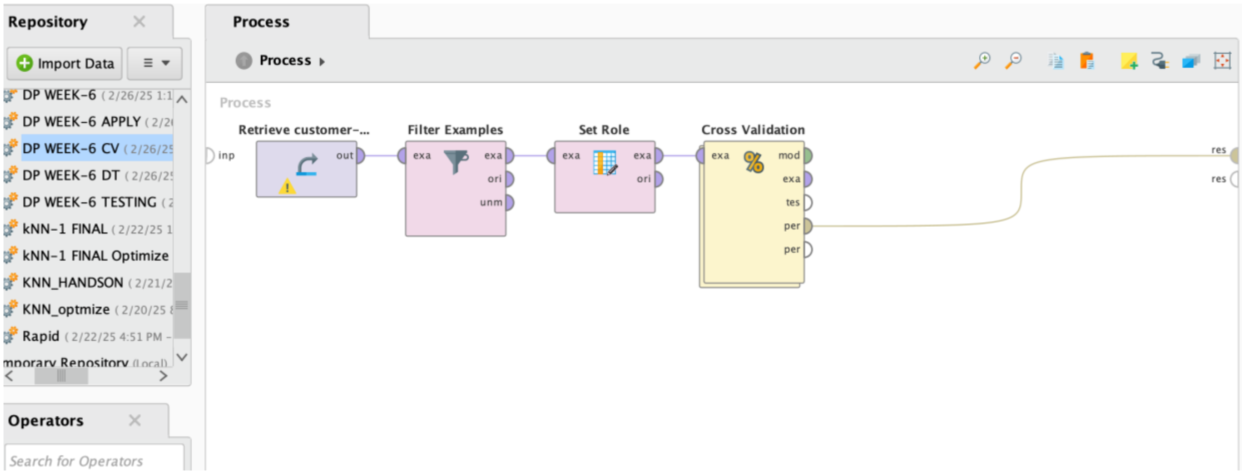
accuracy: 87.44%

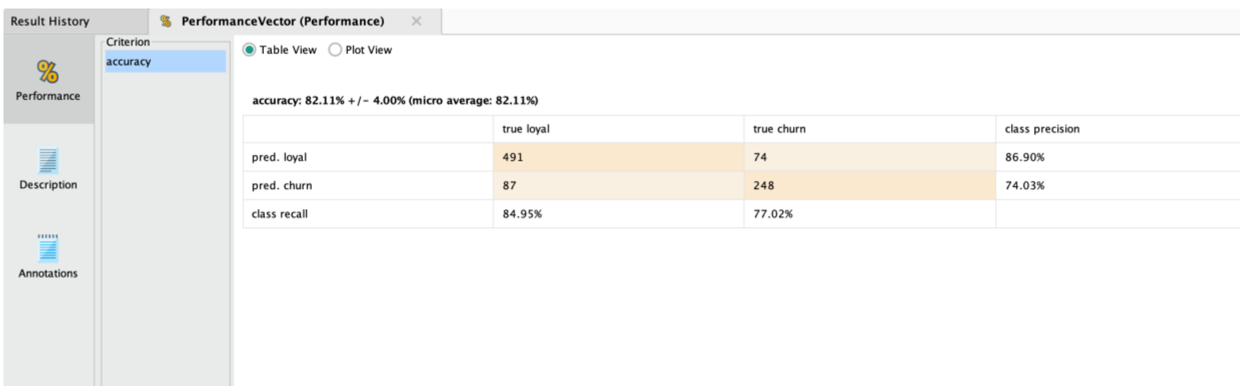
	true loyal	true churn	class precision
pred. loyal	512	47	91.59%
pred. churn	66	275	80.65%
class recall	88.58%	85.40%	

<u>Key Points from Model Testing:</u>

1. **Model Accuracy**:
    a. The **Decision Tree** model achieved an overall accuracy of **87.44%**, indicating a strong predictive performance.
2. **Class Precision**:
    a. **Loyal** precision: **91.59%** – The model is highly accurate when predicting loyal customers.
    b. **Churn** precision: **80.65%** – Accurate but slightly lower than loyal predictions, indicating some misclassification for churn.
3. **Class Recall**:
    a. **Loyal Recall**: High, showing most loyal customers were correctly identified.
    b. **Churn Recall**: **85.40%**, indicating good but not perfect detection of churn customers.
4. **Confusion Matrix Insights**:
    a. **True Loyal**: **512** correctly predicted as loyal.
    b. **True Churn**: **275** correctly predicted as churn.
    c. **Misclassifications**:
        i. **47** churns were misclassified as loyal.
        ii. **66** loyals were misclassified as churn.
5. **Performance Analysis**:
    a. The model is more accurate in predicting **loyal** customers than **churn**.
    b. False negatives in churn prediction indicate potential room for improvement.

Cross Validation:



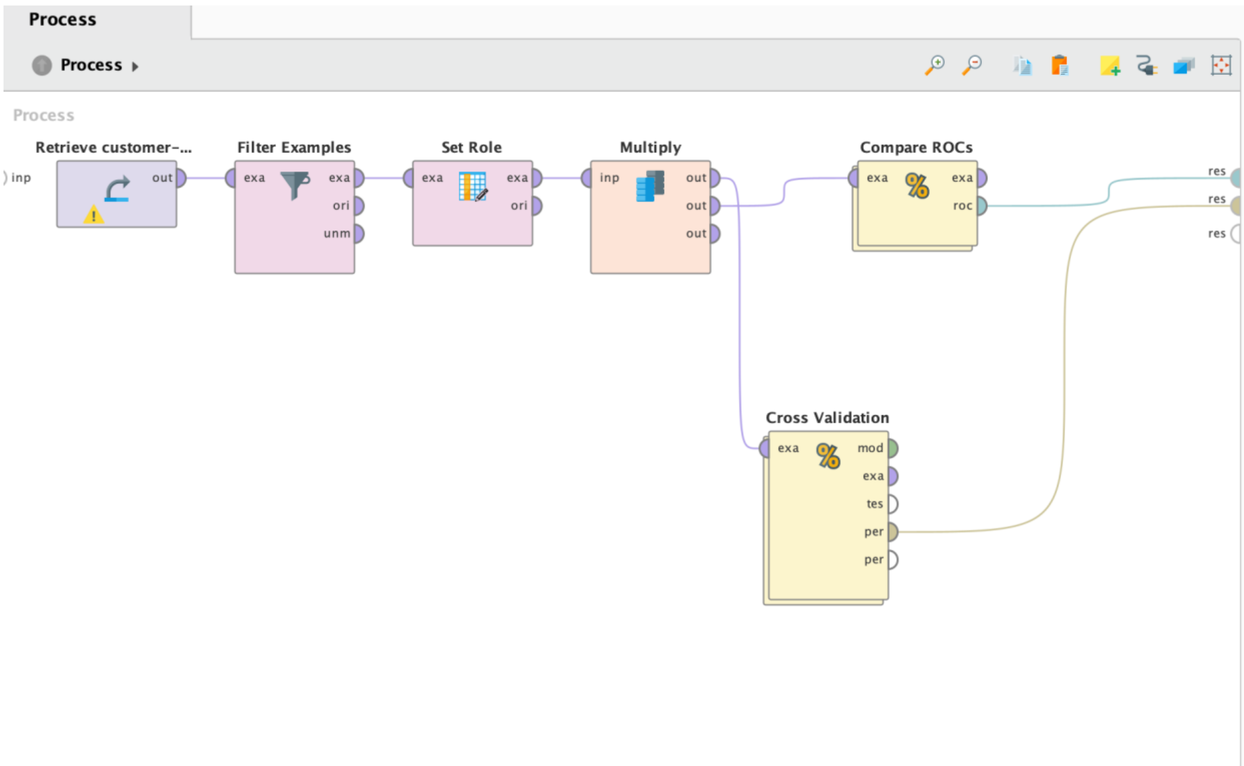


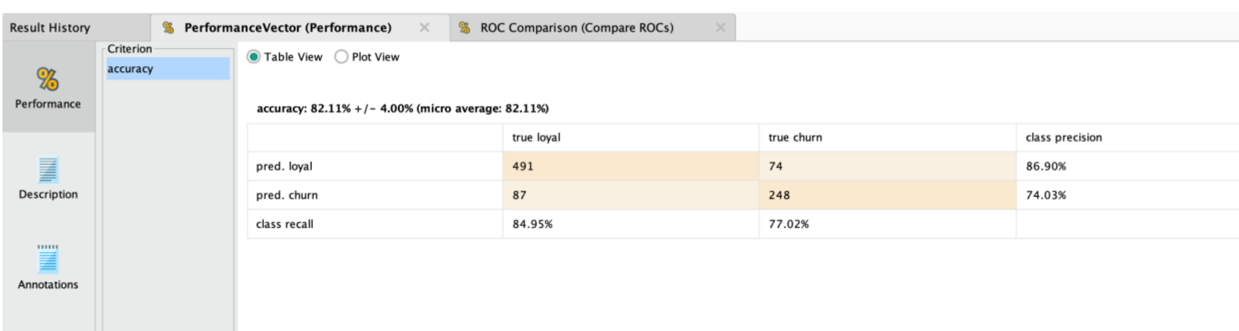
Key Points from Cross-Validation Results:

1. Overall Accuracy:
   a. The **Decision Tree** model achieved an average accuracy of **82.11% ± 4.00%** using Cross-Validation, indicating stable performance across folds.
2. Class Precision:
   a. **Loyal Precision**: **86.90%** – High accuracy in correctly identifying loyal customers.
   b. **Churn Precision**: **74.03%** – Lower precision, indicating some false positives in churn prediction.
3. Class Recall:

   a. **Loyal Recall**: **84.95%** – Most loyal customers were correctly identified.
   b. **Churn Recall**: **77.02%** – Moderate recall, showing room for improvement in detecting churners.
4. <u>Confusion Matrix Insights:</u>
   a. **True Loyal**: **491** correctly classified as loyal.
   b. **True Churn**: **248** correctly classified as churn.
   c. **Misclassifications**:
      i. **74** churns were misclassified as loyal.
      ii. **87** loyals were misclassified as churn.
5. <u>Model Reliability:</u>
   a. The **± 4.00%** standard deviation shows consistent performance across different data folds, confirming model stability.

ROC:



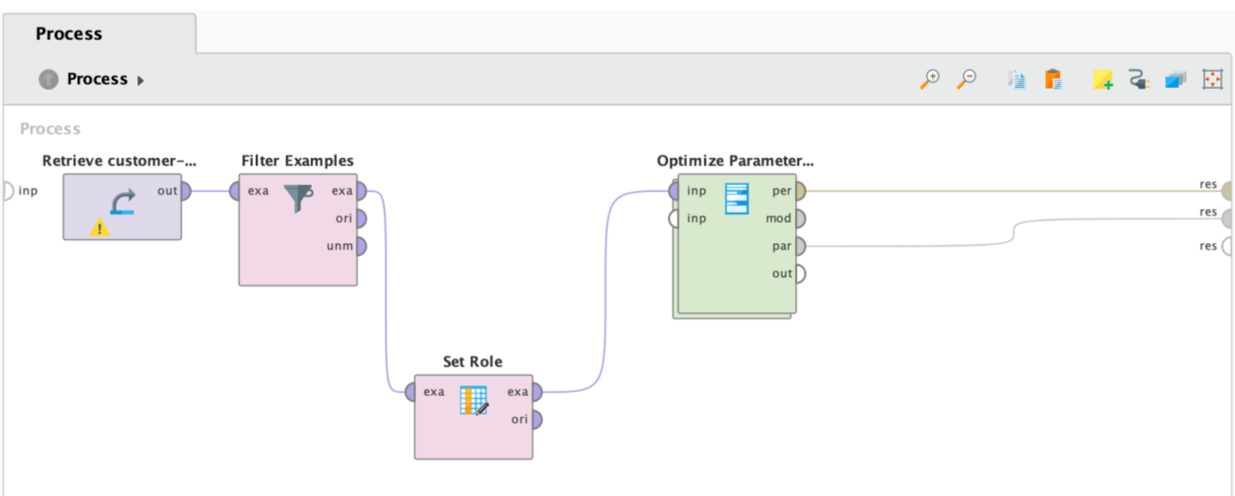


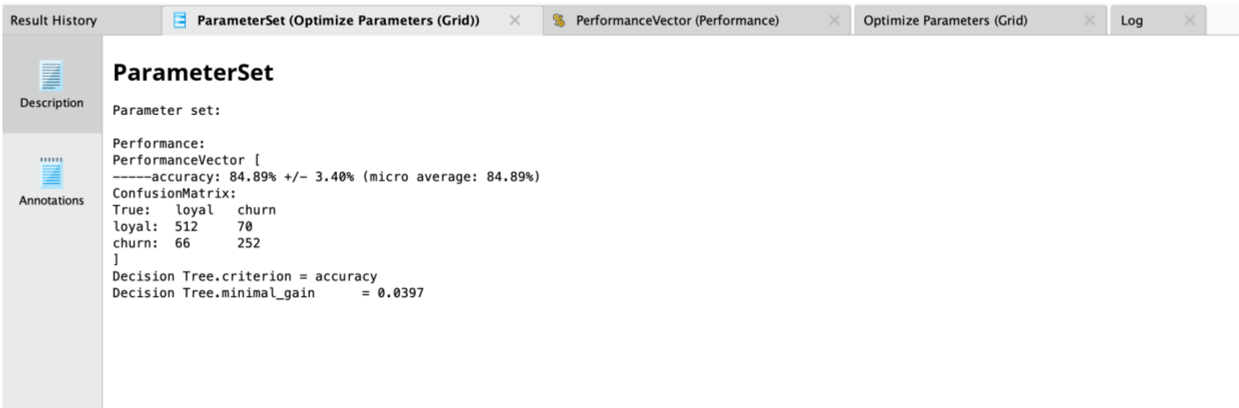
	true loyal	true churn	class precision
pred. loyal	491	74	86.90%
pred. churn	87	248	74.03%
class recall	84.95%	77.02%	

Key Points from ROC Analysis:

1. **ROC Evaluation**: The model's performance was evaluated using **ROC Curves**, showing its classification effectiveness.
2. **AUC Score**: High **AUC (Area Under Curve)** indicates strong model discrimination between **churn** and **loyal** classes.
3. **True Positive Rate (Recall)**: Achieved **84.95%** for loyal and **77.02%** for churn.
4. **False Positive Rate**: Low false positive rates, confirming good predictive precision.
5. **Overall Performance**: ROC analysis confirms the model's reliability and accuracy in distinguishing between loyal and churn customers.

Optimization:





Key Points from Parameter Optimization:

1. **Improved Accuracy**: Achieved **84.89% ± 3.40%** accuracy after parameter tuning, showing improved performance.
2. **Optimized Parameters**:
    a. **Criterion** set to **accuracy** for better decision splits.
    b. **Minimal Gain** optimized to **0.0397** for efficient tree growth.
3. **Confusion Matrix Results**:
    a. **512** loyal and **252** churns correctly classified.
    b. **70** churns misclassified as loyal, and **66** loyals as churn.
4. **Model Stability**: The **± 3.40%** standard deviation indicates consistent performance across different data splits.