Predicting Customer Churn

Microsoft Azure

Agenda

- The Problem

- Methodology

- Conclusions & Suggestions

The Problem

- Companies lose $1.6 trillion per year due to customer churn!

- It costs 5 times more to acquire new customers than retaining an existing one

- The more customers a business retains, the more revenue it makes!

- Customers at high risk of churning, represent a huge additional potential revenue source

Analyze and predict customer churn

CUSTOMER CHURN
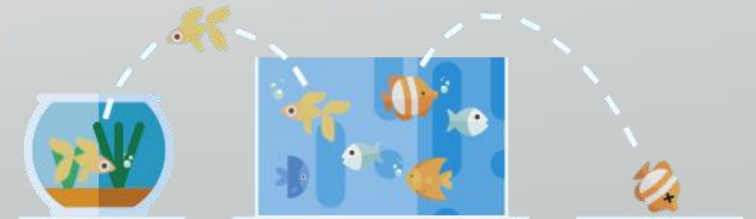
"There is a big difference between a satisfied customer and a loyal customer."

Shep Hyken

"If your customer retention is poor then nothing else matters."

Brian Balfour

Methodology

Data Processing
- Data Acquisition
- Data Wrangling
- Feature Engineering

Model Building
- Data Splitting
- Model Training
- Feature Importance
- Model Scoring

Model Evaluation
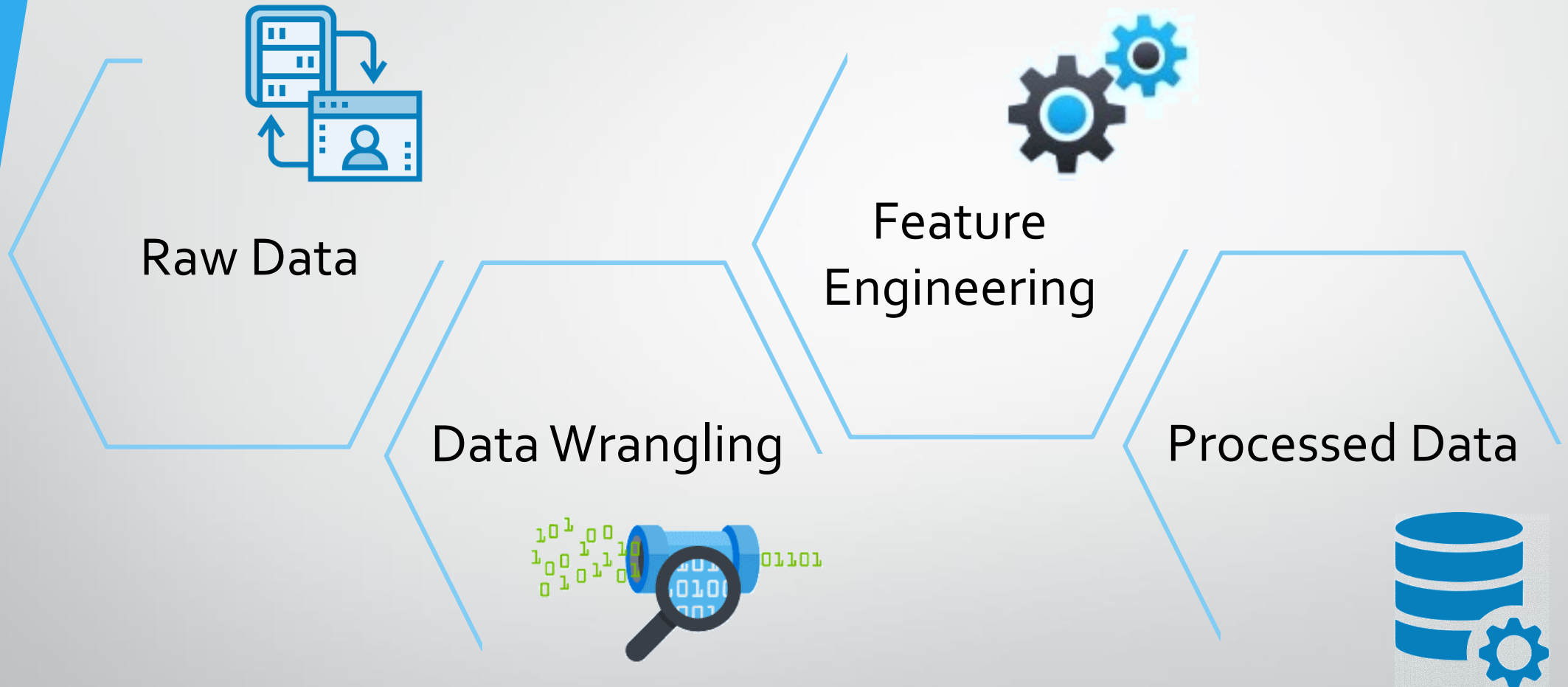- Confusion Matrix
- Performance Measurements

Model Deployment
- Batch Inference

Conclusions
- Results
- Suggestions

Features Used

- Number of times a customer collected (Sum_collect_points)

- Total collected points (Sum_collect_points)

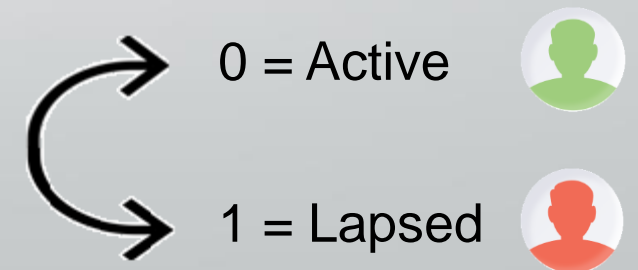- Number of times a customer redeemed (Sum_redeem)

- Years passed since customer's registration to the program (Years_in_the_program)

- Months passed since customer's last action (Months_since_last_transaction)

- ✓ Target Feature: Customer's lapsed status (State)
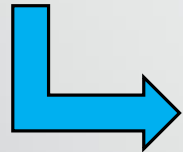
0 = Active

1 = Lapsed

Model Building

Data Splitting:

- 80% training, 20% testing

- Stratified sampling based on target variable (i.e. State)

Model Training:

- Binary classification problem (Active, Lapsed)

    Two-Class Boosted Decision Tree

DECISION TREE

Model Building

Two-Class Boosted Decision Tree:

✓ Non-parametrical method

• Classify customers to active & lapsed

• Hyper-parameter tuning

• Feature Importance

• Score Model

• Model Evaluation

• Model Deployment

Hyper-parameters	Values
Maximum number of leaves per tree	20
Minimum number of samples per leaf node	10
Learning rate	0.2
Number of trees constructed	300

Permutation Feature Importance

❖ More customers lapsed identified, more unnecessary retention costs saved
  ↳ Main performance metric used: **Recall**

➢ Month's passed since customer's last action
  ✓ Most important feature for classification
  ✓ Most customers collected or redeemed points less than 2 months ago

➢ All other features seem to have insignificant effect
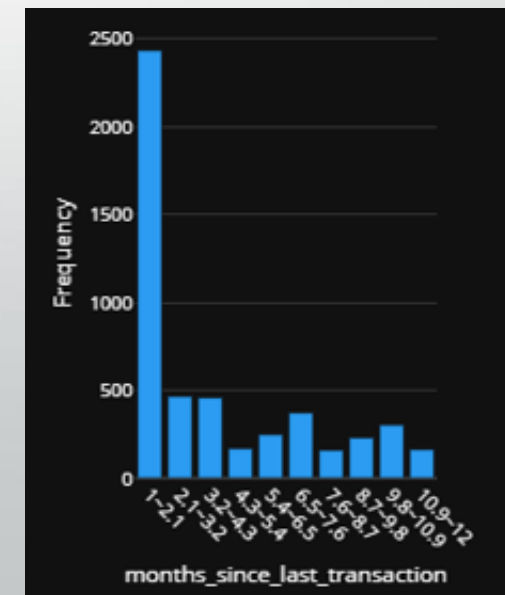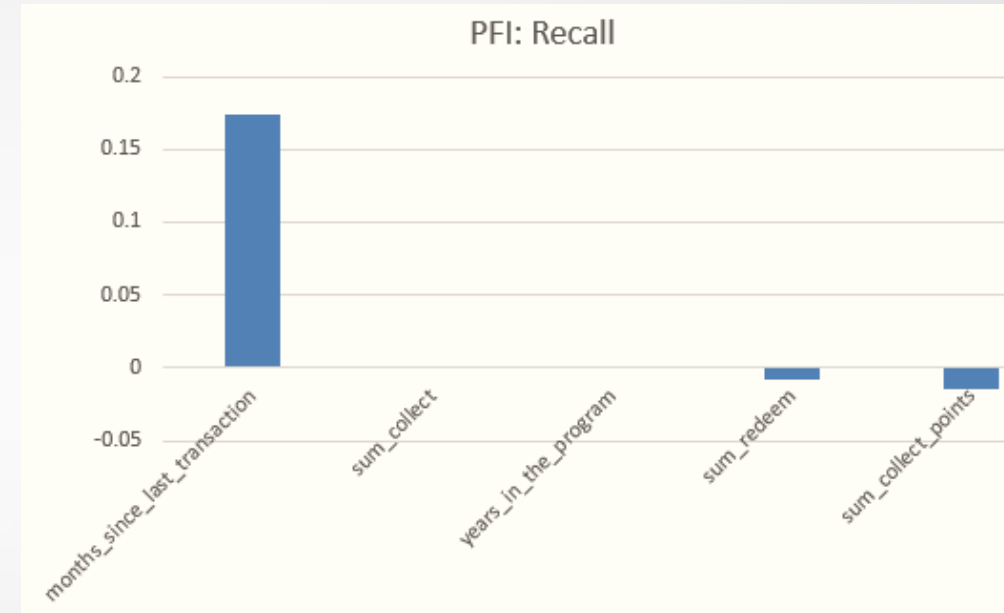
Model Evaluation

From confusion matrix:

✓ Good classifier (values accumulated on the diagonal)

From other typical performance metrics:

✓ High accuracy, precision, recall, AUC

Accuracy	81.7% (of customers correctly classified)
Precision	81.5% (of predicted lapsed customers correctly classified)
Recall	82% (of actual lapsed customers correctly classified)
AUC	.895 (higher the AUC, better the classifier)

✓ Good classification model (will generalize well to out-of-sample data)

Actual

		Lapsed	Active
Predicted	Lapsed	409	93
	Active	90	408

Model Deployment

- Deploy model to a given batch of observations

- Use model to generate customer predictions (active, lapsed in the next 12 months) based on their previous behavior

sum_collect	sum_redeem	sum_collect_ points	years_in_the_ program	months_since_last_t ransaction	Scored Labels	Scored Probabilities
13	0	1,269	19	1	0	.0071
6	0	2,890	10	10	1	.9998
2	0	422	7	9	1	.9999
7	1	2,690	16	1	0	.0144
7	0	211	2	3	0	.4053

Customers whose last action (collection or redemption) was:

- more than 9 months ago, have a huge probability to lapse (more than 99%).

- one month ago, have a very small probability to lapse (about 1%).

Conclusions

- About 1 out of 2 customers were lapsed customers

- 82% of the lapsed customers correctly identified via the model

- Months passed since customer's last action was the only feature that revealed an evident pattern for customer churn probabilities (also suggested from PFI technique)

- The fewer the months passed without a transaction the more loyal the customer

- The best customers are those who are active on a monthly basis

- Customers who are active quarterly are at a high risk of churning

Suggestions

- Keep the most loyal customers as satisfied as possible

- Reward the more active customers

- Provide incentives to customers who are active once per quarter or per trimester to prevent them from lapsing (e.g. discounts, special offers)

- Save retention costs from customers who are very likely to lapse

- ✓ <u>Note</u>: For a successful retention campaign, the decision threshold for identifying a lapsed customer should be decided by the management team or calculated from the profit matrix

Thank you for your attention!