Customer Retention Analytics of SHDR Annual Pass

Capstone Project Final Presentation to Disney

NYU Shanghai - NYU Stern
Data Analytics and Business Computing

Yulin Ding, Rongtian Luo, Dakota Schey, Xinran Xie, Zhaopeng Wan, Xingying Wang, Jiayi Xiang

May 2022

Contents

PART 01

PROJECT OBJECTIVES & SCOPE

PART 01

Objective

Project Goal

Use machine learning algorithm in analyzing annual pass holders' data to predict who are unlikely to renew, to identify the driving factors of renewal and to enable marketing team to act accordingly in affecting customers decisions.

Scope

Integrate yearly customers behavioral data and surveys responses for the input of machine learning model; use the input data to output renewal prediction

PART 02

SUMMARY OF DATASETS

Basic Data Description: Survey Data

Time Period of Survey Data: 2020.11.01-2021.10.31

Welcome Survey

Observations: 1,197

Attributes: 224

Quarterly Survey

Observations: 2,944

Attributes: 158

Renewal Survey

Observations: 272

Attributes: 238

Exit Lapsed Survey

Observations: 167

Attributes: 141

Basic Data Description: Non-Survey Data

Time Period of non-Survey Data: 2020.11.01-2021.10.31

Visitation

Observations: 56,813

Attributes: 8

Record the visit time/counts of each pass

Merchandise

Observations: 134,526

Attributes: 6

Record the detailed product purchased and the amount and count

Food and Beverage

Observations: 162,597

Attributes: 6

Record the food/beverage purchased and price

PART 03

DATA CLEANING

Dataset Processing

- **FnB and Merch**
  - Chose order amount/item count for food/merchandise purchased during visit
  - Merge on ‘source_id’ and ‘Date’ to get total order amount and item count for each pass holder during the past year
- **Visit**
  - Added ‘Season’ variable: classified each visit to a season based on date of visit
  - Merge on ‘source_id’ to get total visit times for each pass holder in the past year
- **Quarterly survey**
  - Survey contains missing value: drop column with substantial amount of missing value; fill the rest of missing value with average from that column

Feature Engineering

- **Data Integration**
  - Introduce ‘SpendPerVisit’ variable - total purchase amount/# of “Times Visited”
  - Combine ‘FnB’, ‘Merch’, ‘Visit’ and ‘QG_survey’ to final ‘df’ training dataset
- **Data Preparation**
  - Identify Y variable: ‘is_renewal’ -- ‘0’ non-renewal (15,438) and ‘1’ otherwise (20,285)
- **Data Split**
  - Divide the dataset into training and test subsets so that we are able to measure the performance of our model on new, previously unseen examples.
  - Test_size = 0.3

Feature Engineering

- **Outlier Detection**

  Code was applied to test different thresholds of outlier detection. Different IQR ranges exist to cut off data points that were outside of a percentile range of values:

  - Lower range: 15 - 40% quantile
  - Upper range: 60 - 85% quantile

Best model performance exists at:

Lower threshold: <40% Quantile ***Upper: >60% Quantile***

Strengthened predictive accuracy, although it cut the dataset size significantly.

Feature Engineering

- **Feature Selection and Interaction**

Analysis done with the correlation function and Random Forest Importance function:

- Dropped 75 additional variables from our dataset of 103 different ones
- Created additional interactive features:
    - Added “FnB” and “Merch” amounts to get “Totals”
    - Multiplied “Season” counts by # of “Times Visited”

Both methods weakened predictive accuracy

- **Feature Scaling (Standard Scaler)**
    - Lessens distance between varying data points to improve generalization

PART 04

MODELING

Models Evaluation

Performance Evaluation Metrics:

- **ROC & AUC:** Derived from confusion matrix
- **MAPE Score:** Mean absolute percentage error

Improving the model – hyperparameter tuning

- **Method:** GridSearchCV and Cross Validation
  - Select the combination of parameters that gives the best accuracy score

Confusion Matrix

- A Table used for **describing the performance of a classification model**
- True Positive (TPR): People who Churned, True Negative (TNR): People who didn't Churn
- In terms of evaluation accuracy, want as little overlap as possible between **TPR** and **TNR** - "Separability"

Actual Class		Predicted Class: Positive	Predicted Class: Negative
	Positive	True Positive	False Negative **Type II Error**
	Negative	False Positive **Type I Error**	True Negative

Confusion Matrix

- When overlap between TPR and TNR does occur, we get **Type I/Type II** errors
- **Type I** error (False Positive Rate): predict a churned AP holder will renew
- **Type II** error (False Negative Rate) : predict a renewed AP holder will not renew
- Focus on **minimizing Type I** error:
    - Retention rate will be harmed if we predict a certain class to renew but they actually end up churning

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive	False Negative **Type II Error**
	Negative	False Positive **Type I Error**	True Negative

ROC & AUC

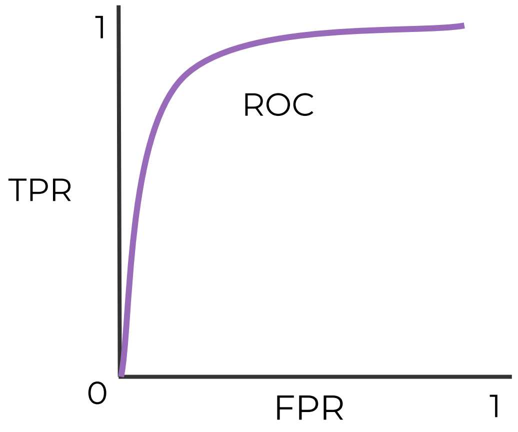
ROC Curve:

- Receiver Operating Characteristics Curve - “Probability Curve”
- Plots TPR vs. FPR at different probability thresholds
- Normally used for evaluating binary classification models

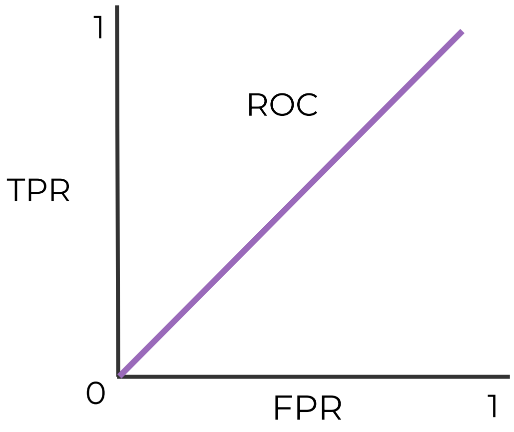
AUC

- Area Under the (ROC) Curve - “Measure of Separability”
- Higher AUC means that the model is better at distinguishing between classes, ultimately maximizing the **TPR** at the expense of the **FPR (Type 1 error)**
- Based on this logic, our models have ~70-75% accuracy of distinguishing a “Churner” from a “Non-Churner”

ROC & AUC



Strong performing AUC/ROC curve



Weak performing AUC/ROC curve

PART 04

Models Applied

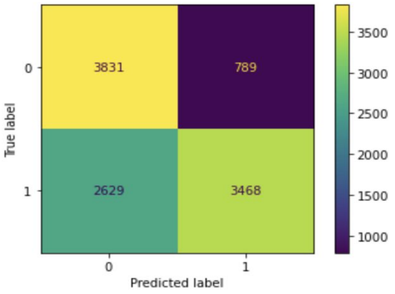
Classification Models

- CatBoost
- MLP (Multilayer Perceptron)
- Decision Tree
- Random Forest
- Gradient Boosting
- KNN (K-nearest Neighbor)
- Stacking

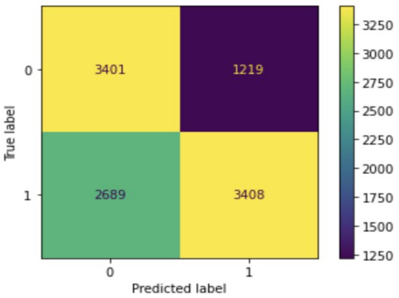
PART 04

Confusion Matrices

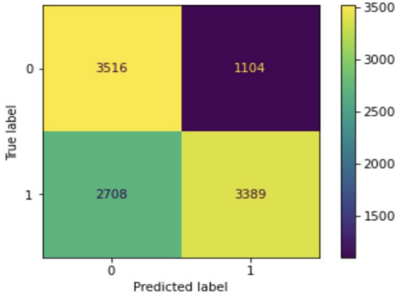
Catboost



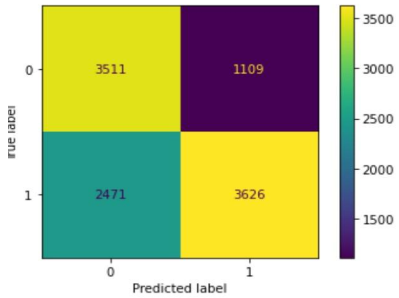
MLP



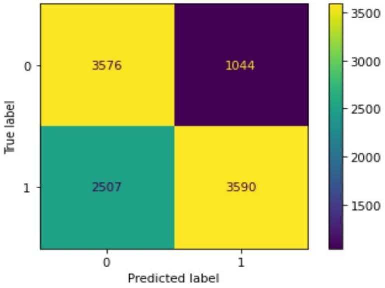
Decision Tree



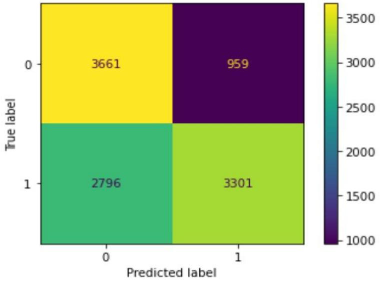
RF



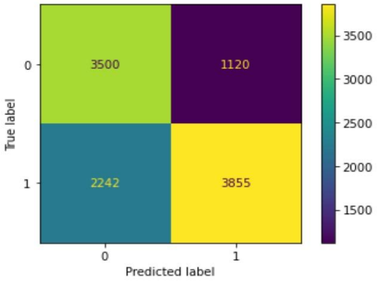
Gradient Boosting



KNN

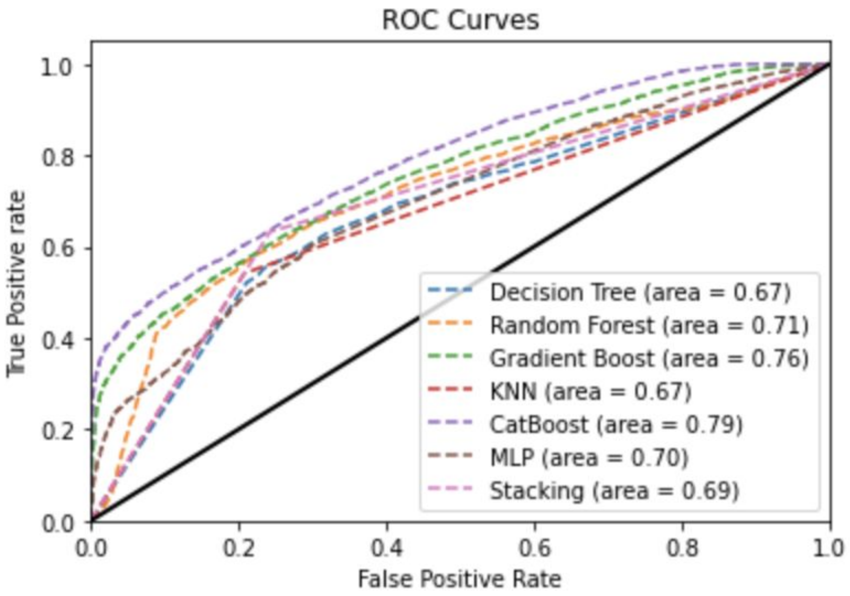


Stacking



PART 04

Model Performance



Models	MAPEs
Decision Tree	33.6
Random Forest	33.4
Gradient Boost	33.1
KNN	35.0
CatBoost	31.9
MLP	36.5
Stacking	31.4

NYU

PART 05

INSIGHTS

What We Have Learned

Model-wise

- ❖ **CatBoost gives the best performance with an AUC value of 0.79 and Type I error percentage of 7.36%.**
- ❖ **AP holders' spending levels, pass types, ages, and if they visited the resort in the last 3 months are among the most important features in our models.**
- ❖ **Overall model performances can be considered fair but rooms for improvement remains.**
  - ➢ Averaging AUC value is approximately **0.71**.
  - ➢ Type I error percentages are **≤ 11.5%**.
- ❖ **However, in terms of MAPE, performances are not ideal (average MAPE is 33.8%).**
  - ➢ MAPE is mainly used for regression models instead of classification ones, the results may not be accurate.

What We Have Learned

Business-wise:

- **Optimize Marketing Campaigns**
  - Increase Customer Retention
    - Features can be interpreted as **flags of actions**.
    - If a given AP holder is predicted to churn, we can look at the dominant features, combined with the AP holders' appeal from the survey dataset, to develop customized campaigns to retain the AP holder.
  - Increase Profit Margin
    - We can precisely locate loyal AP holders and provide them with (customized) promotions (e.g. cross-sell and up-sell) to increase the efficiency of promotions and increase revenue.

PART 06

CHALLENGES & UNSOLVED ISSUES

Low AUC/MAPE value for our models?

Even with extensive dataset preprocessing, feature engineering and hyperparameter tuning, our models performed only averagely...Why?

- **Feature Engineering**
  - Could develop and combine more features together
- **Hyperparameter Tuning**
  - Tried only 600 combinations of parameters for each model on average (500 for CatBoost), could try more

External Challenges

- Require more confidential personal information of annual pass holders, like living geolocational data (cities) and individual renewal date times
- Integrate Covid-19 related data with existing database:
  - Date of reported community transmission in Shanghai
  - Declarations of medium and high risk areas in cities where pass holders reside
- Concise survey to limited important features for ease of use and higher response rates
- Examine more combinations of parameters
  - The 3-month dataset performed significantly better than this year-long dataset because it contained almost 10 times as many data points

PART 07

FUTURE DIRECTIONS

Next Steps

- **Involve the impact of Covid-19**
  - Potential dominant factors in the future
- **Perform text and sentiment analytics of survey results**
  - Direct response from customer is more reliable than numbers alone
- **Identify who will respond positively to marketing campaigns in pass holders with low renewal intentions**
  - Conduct precise marketing to realize cost-saving or efficiency maximization

Next Steps

- **Determine Discount Rates needed to prevent Churn**
  - Because our models didn't perform as well as we had hoped, it would be difficult to accurately estimate discount rates needed.
  - However, with 4 different passtype tier memberships, you can:
    - Build logistic/linear regression models to ACCURATELY predict churn rates of each customer
    - Develop lift curves for each customer in each tier based on a set retention threshold to uncover optimal discount rates
    - Use these discount rates to offer various incentives to customers, **lessening the churn rate**

THANK YOU!