

Gym Member Churn Analysis

Simran Karamchandani, Roselyn Rozario, Mathew Spencer
Linear and Non-Linear Models | ADSP 31010 | Winter 2024



Agenda

Introduction

Exploratory Data Analysis

Model Selection

Explainable AI Methods

Business Value & Final Thoughts

Organization



ABC Fitness is a high-end fitness club in Chicago. It provides thousands of members with access to a wide range of exercise equipment, group classes, a café, and other amenities.

Challenge

ABC Fitness is experiencing elevated **membership churn**, which is pressuring financial results and threatening the viability of the business.

Business Data Strategy

Our goal is to reduce churn rates by identifying members likely to churn through a mix of **linear and non-linear modeling techniques** and developing targeted retention strategies based on feature importance learnings from our **Explainable AI solutions**.

Proactively address and minimize membership churn through implementation of linear and non-linear modeling techniques.

There are plenty of commercial gyms in the U.S., but ABC Fitness's circumstances (e.g., location, amenities, membership mix) require a unique data analytics strategy

Monetary Value



Cost Savings



Pricing Optimization

- Finding new members is expensive
- Better develop pricing strategies catered to level of customer engagement

Data-Driven Decision Making



Personalized Experience

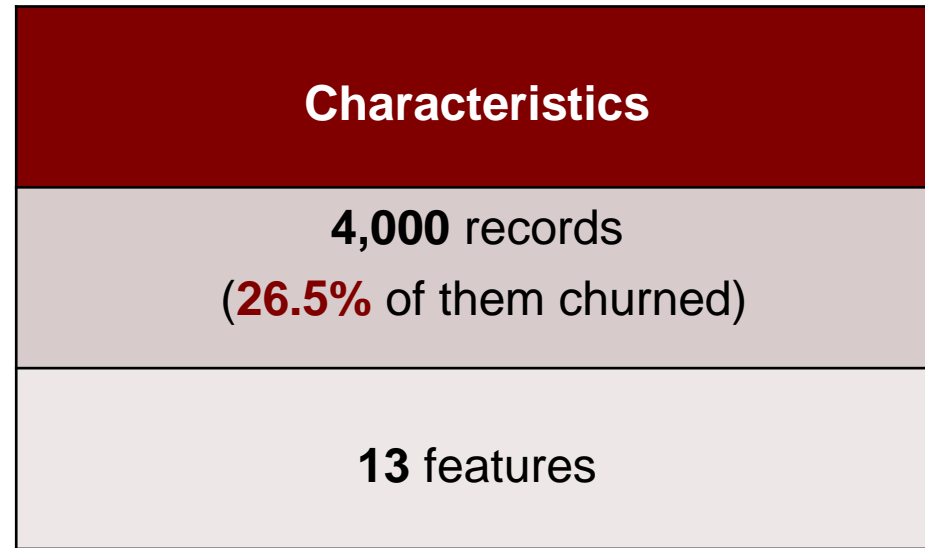


Resource Optimization

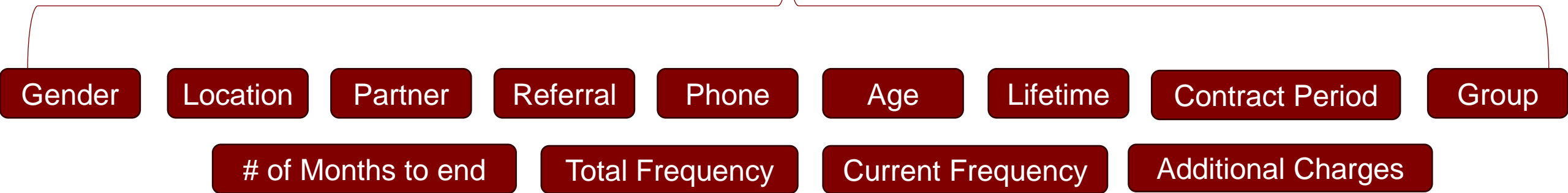
- Offer services and promotions that will peak customer interest
- Allocate resources towards favored activities

Data Overview

EDA

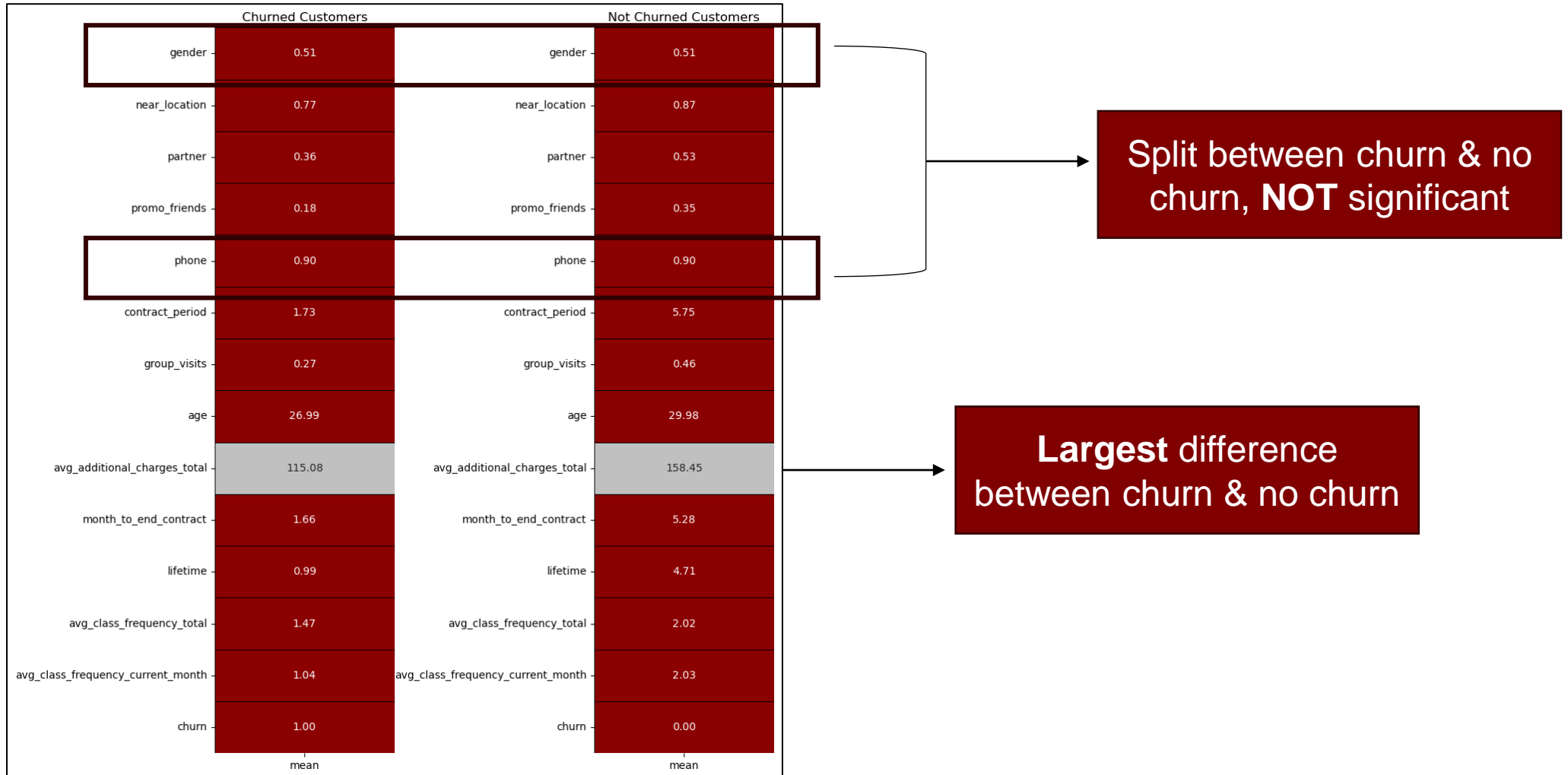


- Quality Concerns:**
- No missing values
 - Narrow age range



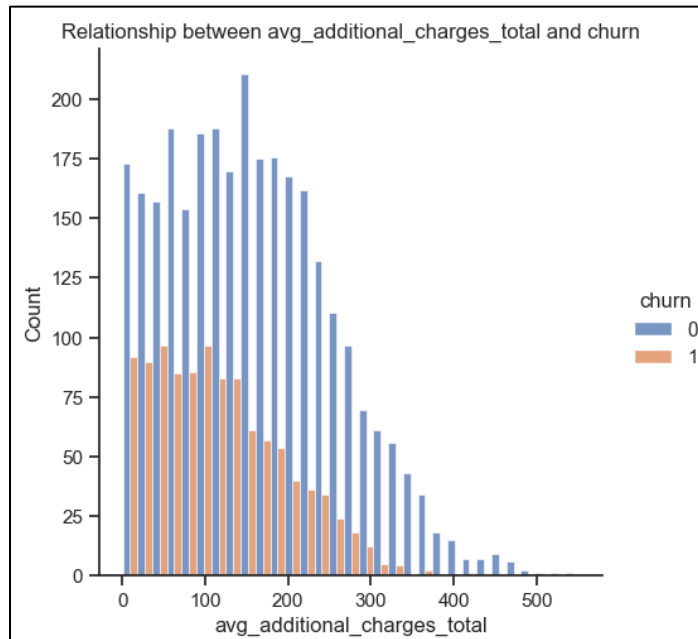
Features of Churn

EDA

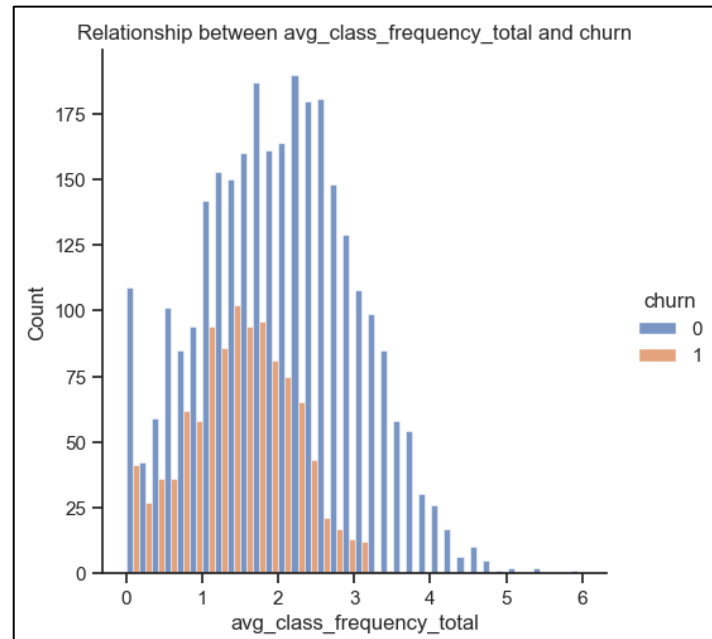


Key Characteristics of Churn

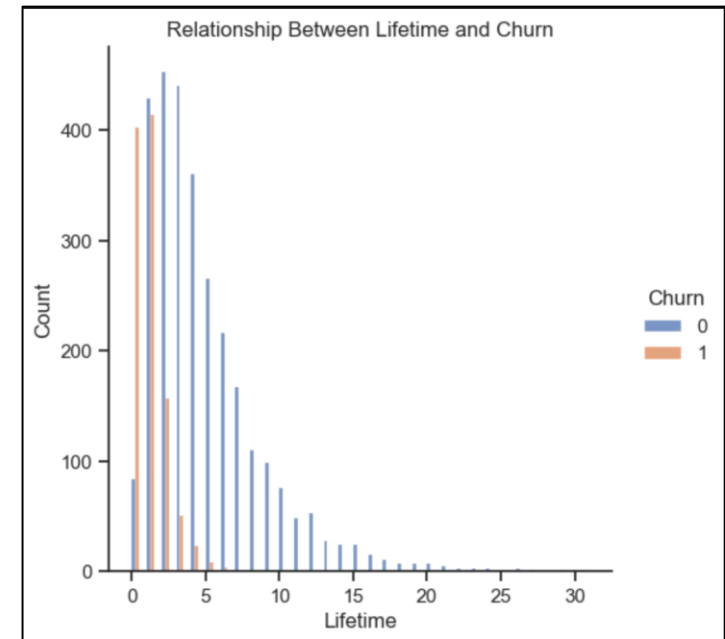
EDA



Less additional charges



Less frequent



Less committed

Linear & Non-Linear Modeling Techniques

Model Selection

LOGISTIC REGRESSION

RANDOM FOREST

PROS	<ul style="list-style-type: none">• Highly interpretable• Robust• Computationally efficient	<ul style="list-style-type: none">• Works well with heterogeneous data• Robust
CONS	<ul style="list-style-type: none">• Susceptible to overfitting• Suboptimal for complex relationships• Limited applications	<ul style="list-style-type: none">• Less efficient• Performance degradation with high dimensionality• Harder to interpret


Logistic Regression

Model Selection

Fit initial model with all 13 features

Implemented **recursive feature elimination (RFE)** for optimization of model

Reduced model to 6 defining features



	Feature	Coefficient	Odds Ratio	Standard Error	Z-value	P-value
0	Contract_period	-1.461	0.232	0.020	-73.945	0.0
1	Age	-1.053	0.349	0.020	-53.340	0.0
2	Avg_additional_charges_total	-0.505	0.604	0.019	-26.285	0.0
3	Lifetime	-3.440	0.032	0.020	-172.968	0.0
4	Avg_class_frequency_total	3.168	23.752	0.065	48.989	0.0
5	Avg_class_frequency_current_month	-4.270	0.014	0.066	-64.424	0.0

Random Forest Modeling (RFM)

Model Selection

Feature Importance

Identification of non-linear
patterns

Predictive Accuracy

Minimization of overfitting

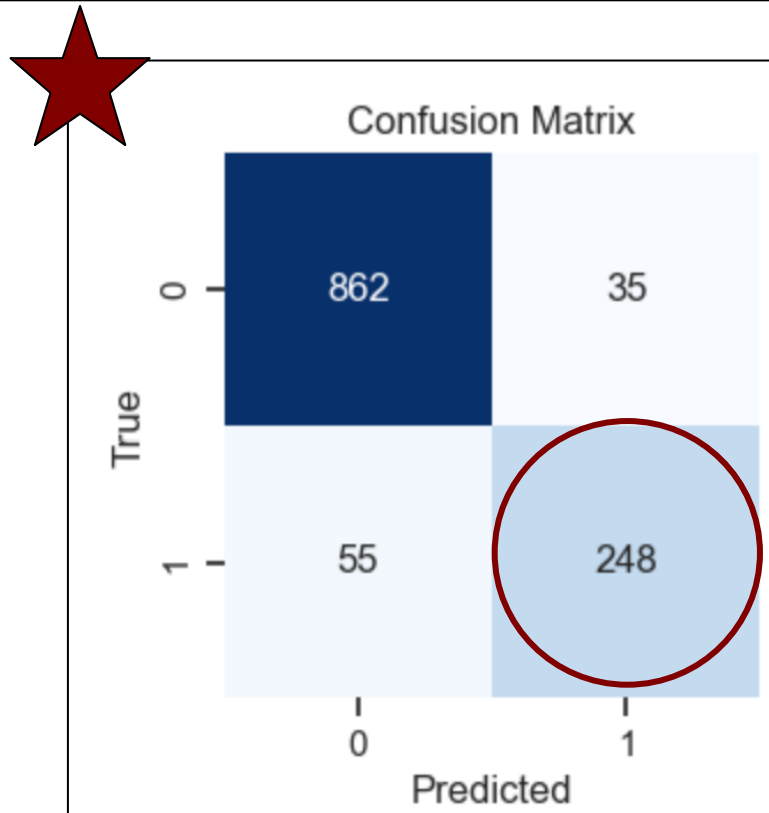
Robust

Introduction of additional
randomness

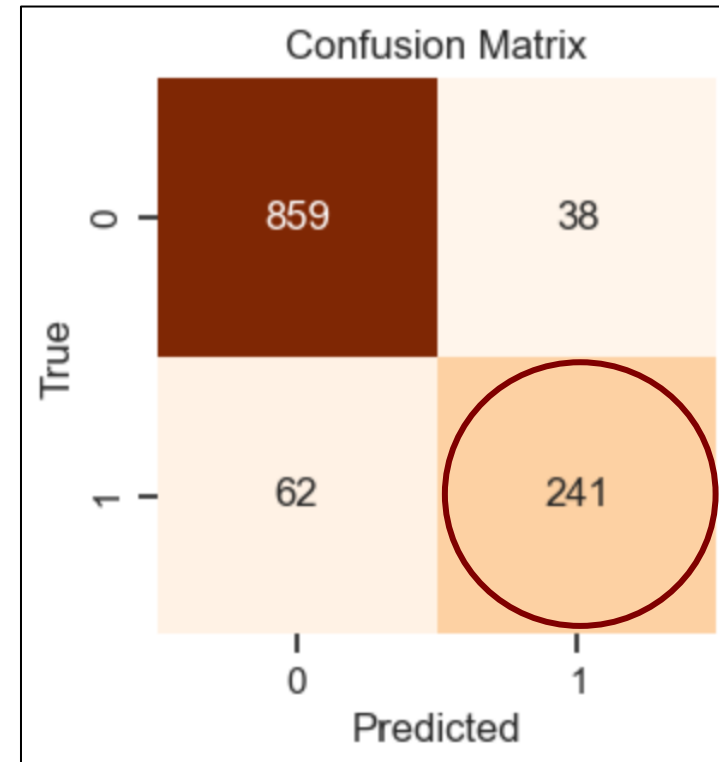
Logistic Regression vs. Random Forest

Model Selection

↑ Precision ~ ↑ # of **True Positives**



Logistic Regression



Random Forest

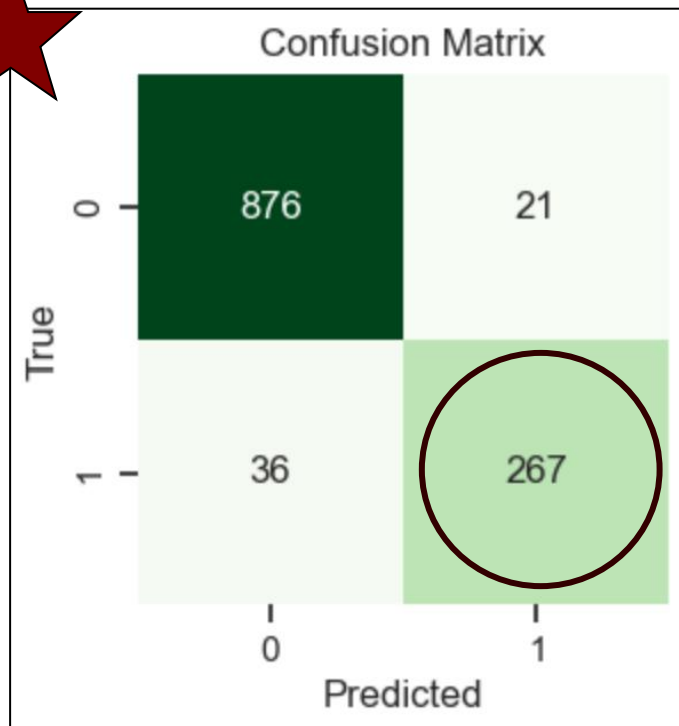
Explainable Boosting Machine (EBM) vs. General Additive Models (GAMs)

Explainable AI

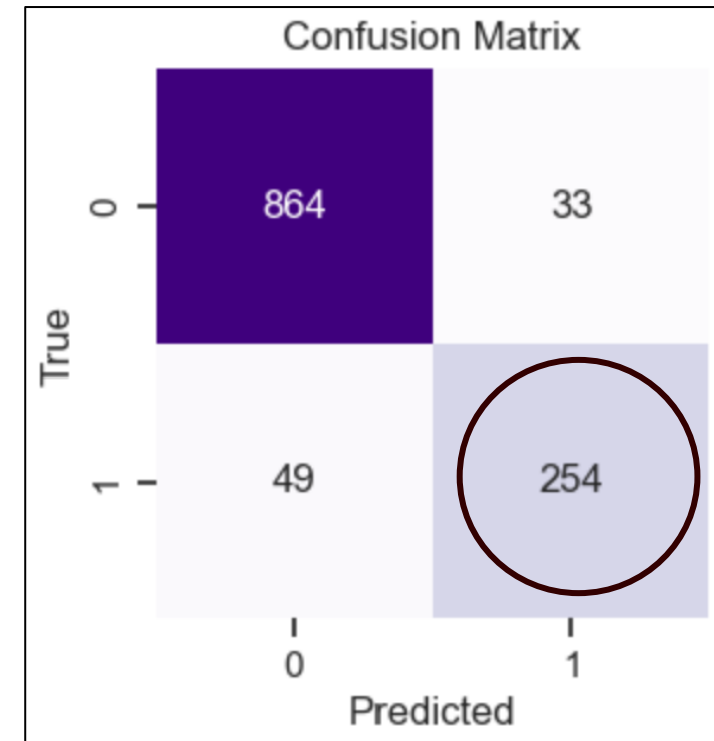
Feature Insight

Model Validation

Churn Identification



95.25% Accuracy



93.17% Accuracy

Explainable Boosting Machine (EBM)

General Additive Models (GAMs)

Local Interpretable Model-Agnostic Explanations (LIME)

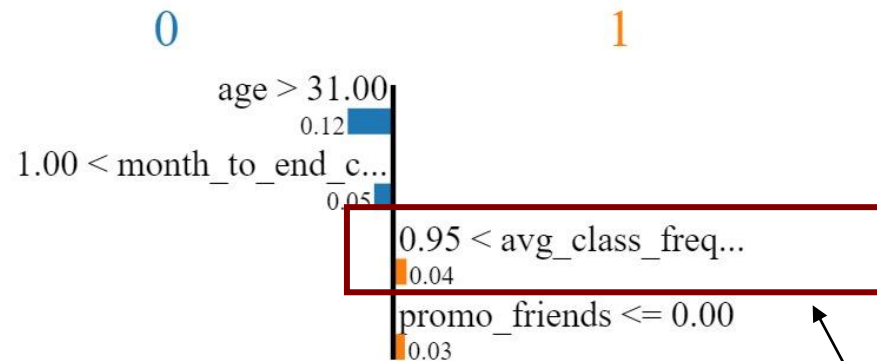
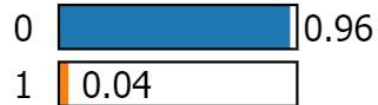
Explainable AI

Feature Insight

Model-Independent

Transparency

Prediction probabilities



Feature Value

age	37.00
month_to_end_contract	5.00
avg_class_frequency_current_month	1.61
promo_friends	0.00

1. Selection of **instance**
2. Use **RFM** to make predictions
3. Approximate **weighted** distance
4. **Fit** model
5. Analyze **influential** coefficients

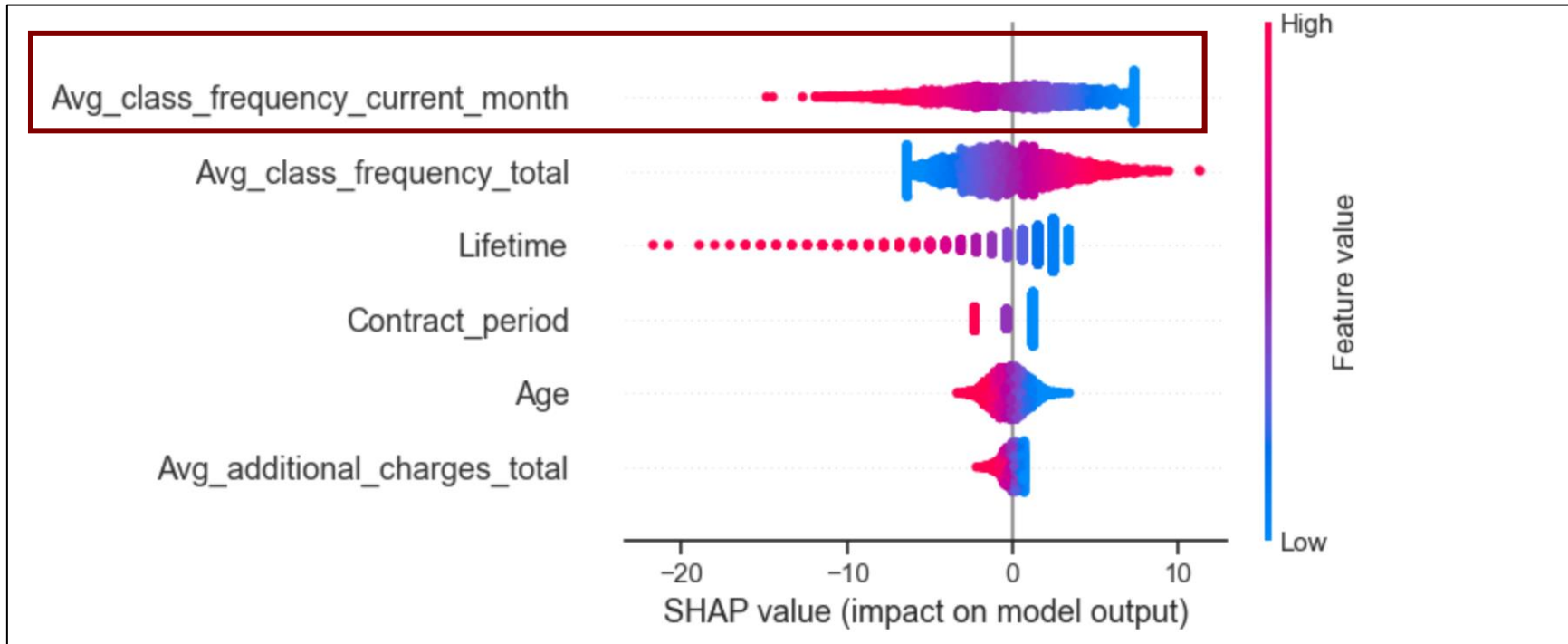
SHapley Additive exPlanations (SHAP)

Explainable AI

Feature Insight

Model-Independent

Model Validation



1. Train **logistic regression** model

2. Generate **SHAP** values from explaining model

3. Summarize **effects** of each variable based on predictions

Models Comparison

Model Selection

	Linear		Non-Linear	Explainable AI	
Metric	Logistic Regression	Logistic Regression with RFE	Random Forest	EBM	GAM
Accuracy	0.92	0.93	0.92	0.95	0.93
Correct Prediction Rate (Precision)	0.88	0.89	0.86	0.93	0.89
Missed Opportunity Rate (Recall)	0.82	0.82	0.80	0.88	0.84
Overall Performance Score (F1-Score)	0.85	0.85	0.83	0.90	0.86
ROC AUC	0.89	0.89	0.88	0.93	0.90

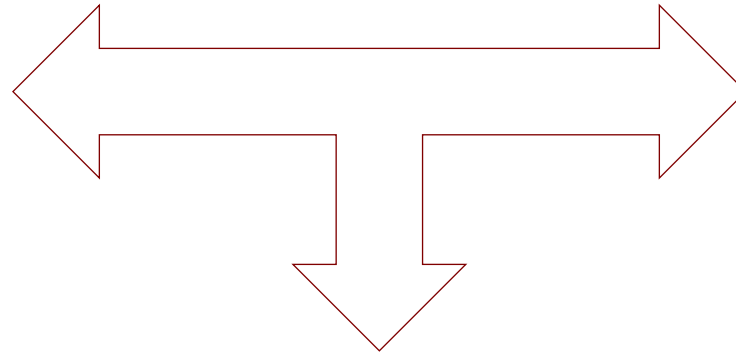
- Combination of linear and non-linear techniques with Explainable AI
 - Logistic Regression with RFE
 - Random Forest
 - Explainable Boosting Mechanism (EBM)

Metric	Model
<i>Accuracy</i>	0.92
<i>Correct Prediction Rate (Precision)</i>	0.88
<i>Missed Opportunity Rate (Recall)</i>	0.82
<i>Overall Performance Score (F1-Score)</i>	0.85
<i>ROC AUC</i>	0.89

Proposed Model & Explainable AI

Logistic Regression with RFE

- Selection of **subset** of influential features
- **Simplifies** model and computational complexity



Random Forest

- Random Forest can detect **non-linear patterns** and provide highly accurate models.

Explainable Boosting Mechanism (EBM)

- Illustrates impact of **each feature** on model's predictions through summation
- **Higher** accuracy ~ strong in making **correct** predictions

Business Considerations

Business Value & Final Thoughts

KEY INSIGHTS

Does not get a discount through their workplaces.

Does not sign up through a friend's promo/referral option.

Signs up for a 1-month contract period.

Does not take part in group-related activities.

Typically visits less than 3 times in a week.

When there isn't long left in the contract period.

NOTEWORTHY TRENDS

High churn rates amongst those within the 20-to-30-year-old age range.

High churn rates when the average additional charges total ranges between \$0 to \$300.

In order to minimize gym membership churn, ABC Fitness will need to focus on **cost for customers, frequency of visits, and sign-up/contract options.**



Expanded Offerings

- Offering Activities (i.e., Zumba Classes)
- Fitness Challenges



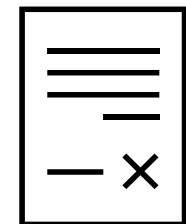
Long-Term Discounts

- Student Discounts
- Corporate Discounts



Loyalty/Rewards Programs

- Points-Based Rewards
- Referral Bonuses
- Personal Training Packages
- Exclusive Access
- Attendance Milestones



Sign-Up/Contract Options

- Special Sign-Up Offers
- Discounted Contract Promos

Completeness/Substance:

- Examined Linear & Non-Linear techniques
- Incorporated Explainable AI Methods
- Customer data trained, tested, and split
- Key insights noted from modeling techniques and Explainable AI methods

Project Design:

- **Ensemble model** is a result of **highly accurate** models that were chosen from the modelling approach process and incorporates several techniques, resulting in **reliable** predictions
- Thought-process behind ensemble model is what makes the project design sound and well-chosen

Thank you!

Simran Karamchandani, Roselyn Rozario, Mathew Spencer

Linear and Non-Linear Models | ADSP 31010 | Winter 2024

