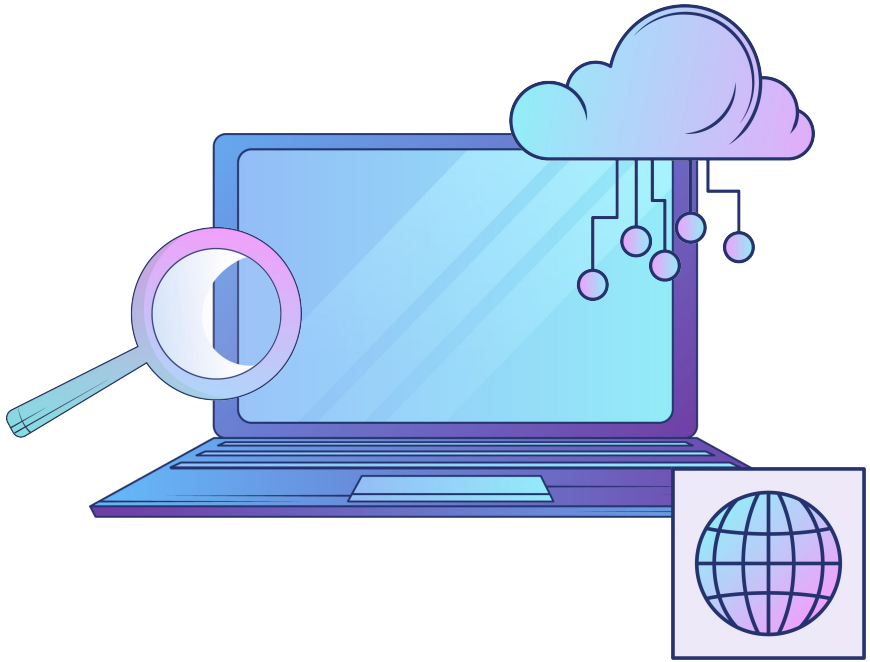
CUSTOMER CHURN

Prediction & Analysis

IS 675 – Deep Learning for Business
Group 9 – Mahaam Ahmed, Natalie Rath, Natalie Nguyen

Table of contents



01

Project Overview

Project Overview

INDUSTRY

Subscription-based entertainment company (Netflix, Disney+, Hulu, etc.)

PROBLEM

Customer churn – when customers terminate their subscription, is a crucial metric for subscription-based businesses.

Goal – Predict customer churn and analyze customer behavior

EFFECTS

Churn consequences:

- Revenue Loss
- Reduced market share
- Increased customer acquisition costs

02

About the Data

About the Data

Data Description

Anonymized information about customer subscriptions and their interaction with the service.

Data shape:
~ 240K rows x 21 columns

Independent Variables

20 features total:

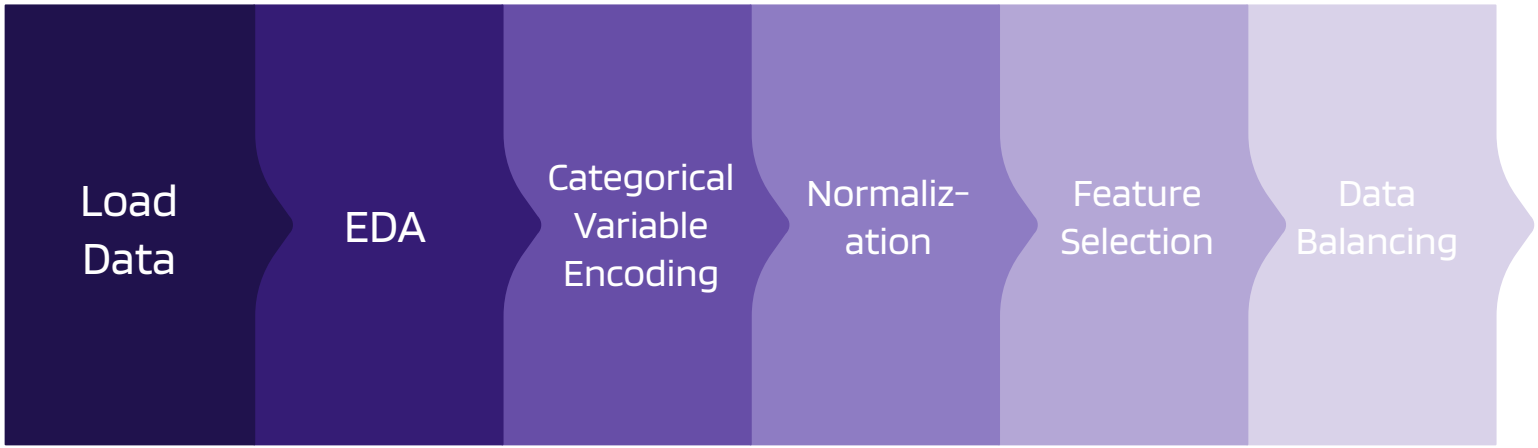
- 10 categorical
- 9 numerical
- 1 identifier

Target Variable

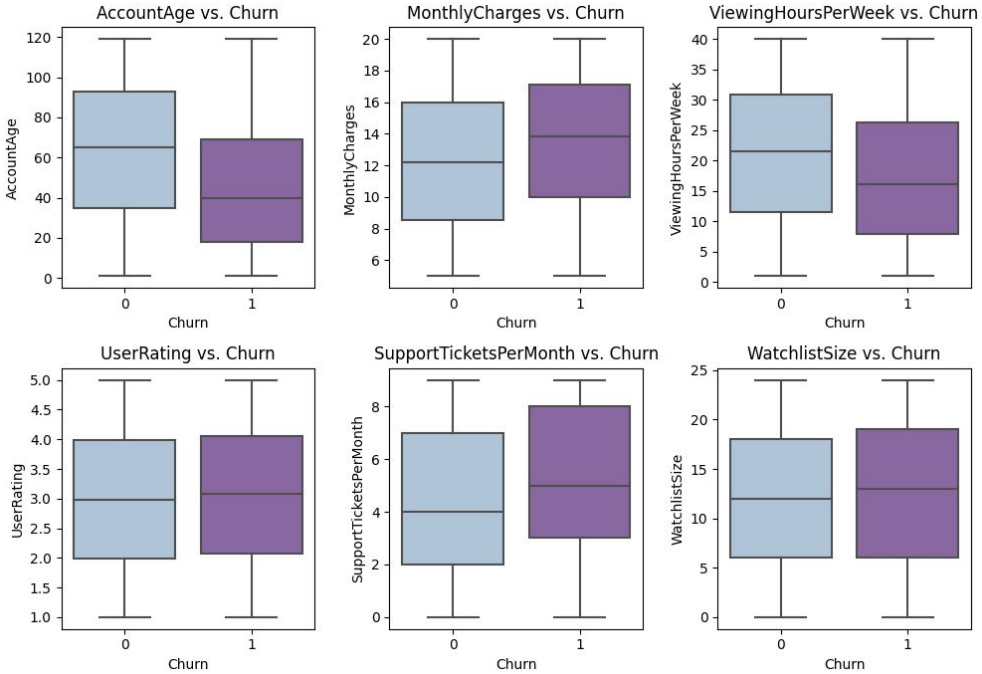
Churn:

- 0 - Not churn
- 1 - Churn

Data Preprocessing



Churn patterns in numerical attributes



Account Age vs. Churn:

- Customers with lower account ages are more prone to churn
  ⇒ Customer loyalty

Monthly Charges vs. Churn:

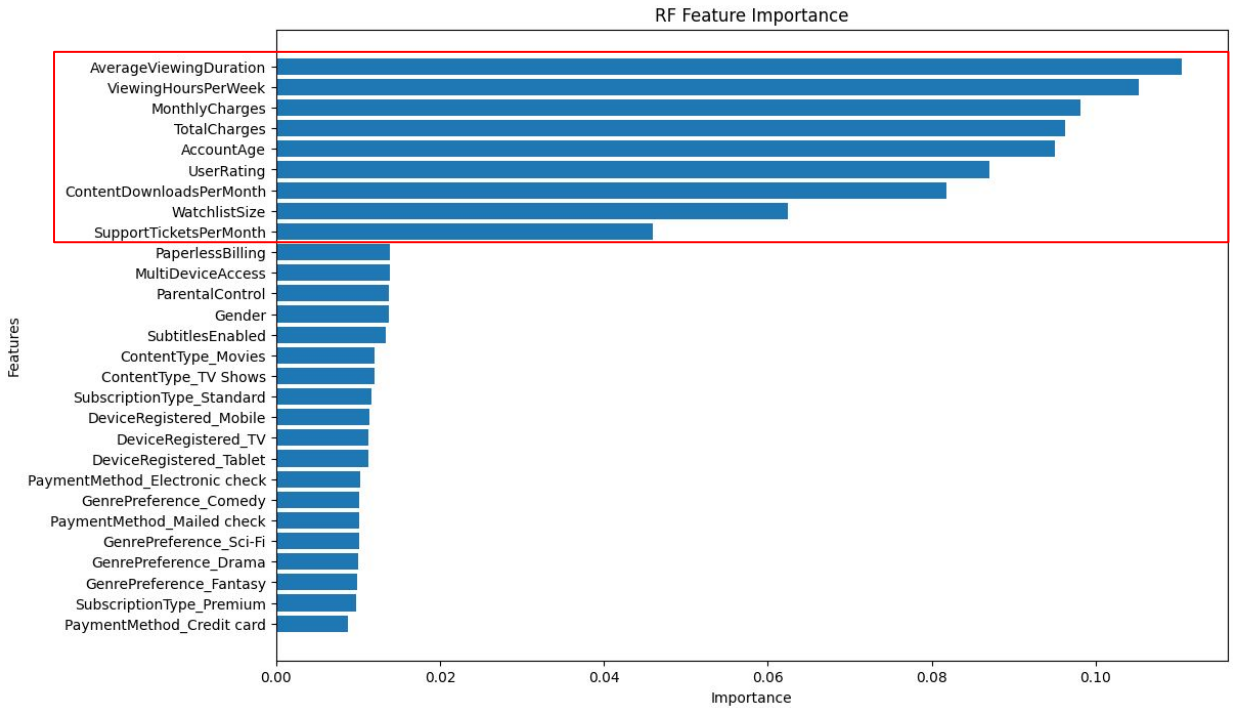
- Customers with higher monthly charges are more likely to churn
  ⇒ Cost

Support Tickets per Month vs. Churn:

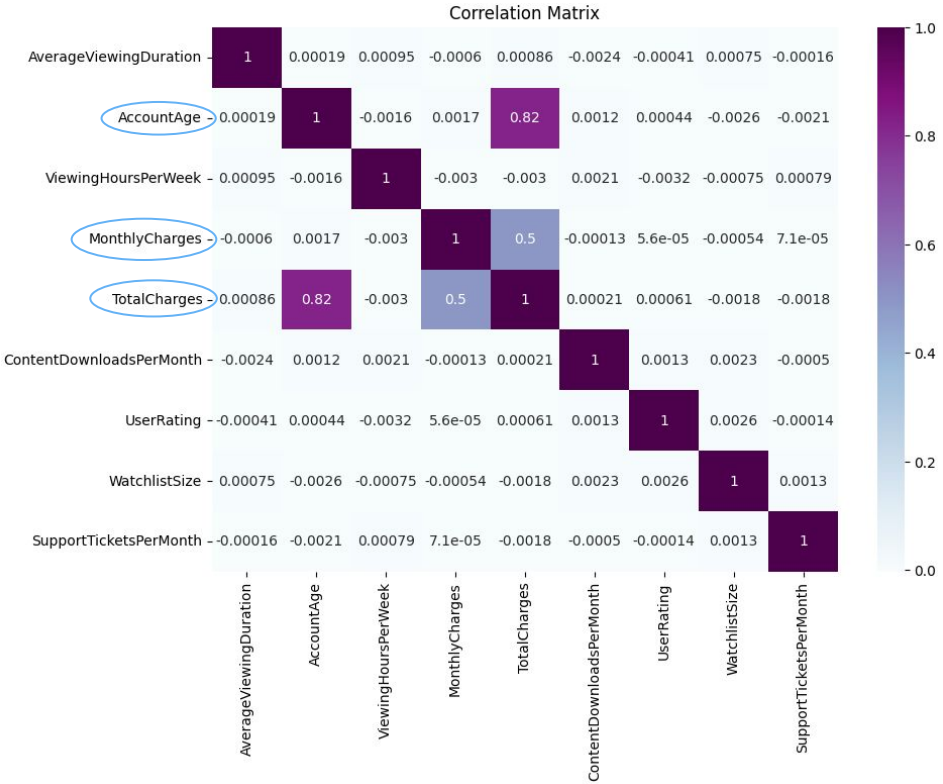
- Churned customer has higher volumes of support tickets
  ⇒ Service dissatisfaction

Feature Selection

Feature importance using Random Forest



Feature Selection



Data Balancing

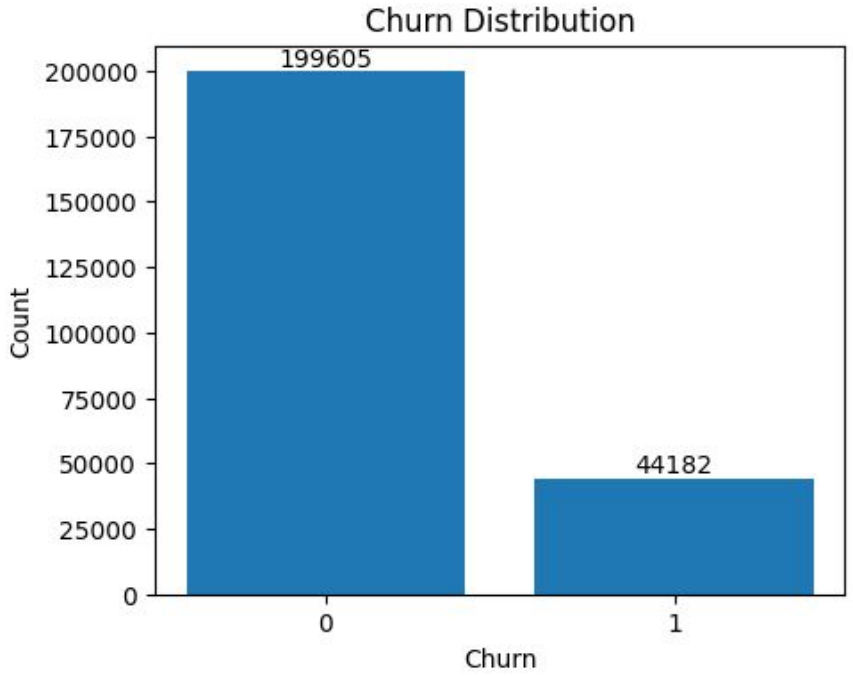
Class Imbalance

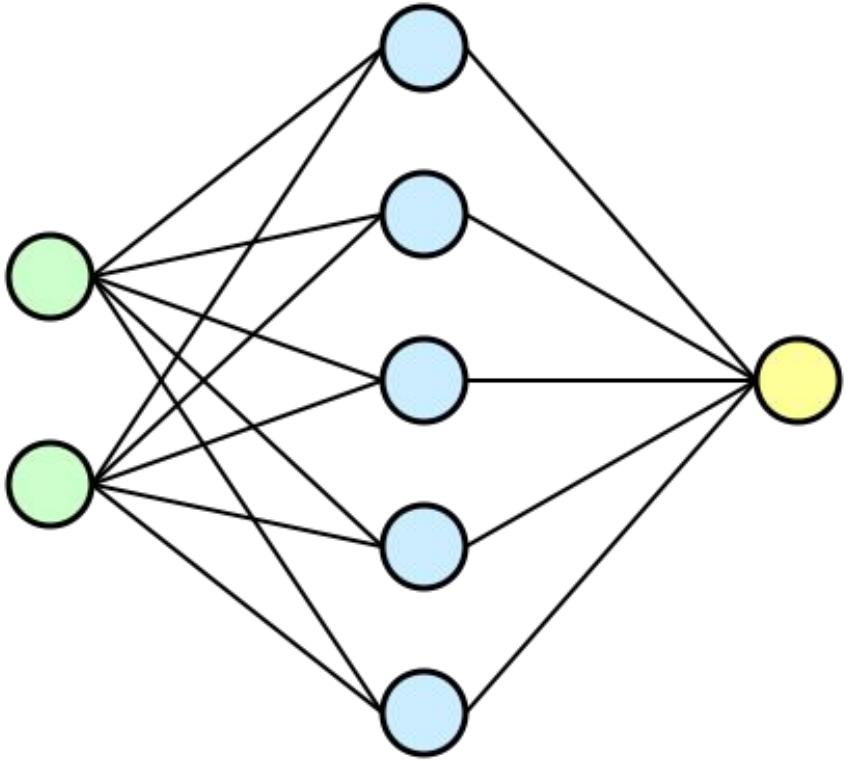
Ratio between non-churn and churn customers = **45:1**

Implemented Method

Undersampling:

- Decreasing the number of instances in the majority class to balance it with the minority class.
- Risk: Information loss





03

Neural Network Model

Model Architecture

- Data Split: 60/40
- Type: Feedforward Neural Network
- Layers: 4

  Input: Features

  Output: 0 Not Churn or 1 Churn
- Dataset size: Large

Hyperparameter Tuning : Random Search

Number of Neurons

Range from 32 to 512
(more neurons are able to capture more complex patterns to predict churn)

Learning Rate

Range from 0.0001 to 0.1
(influences stability of training process)

Activation Functions

- ReLu (most common)
- Sigmoid (probabilities)
- Tanh (zero-centered)

Optimizers

- SGD optimizer
- Adam optimizer (popular)

Hyperparameter Tuning : Best Model

Number of Neurons:

Hidden Size 1: 256

Hidden Size 2: 256

Learning Rate : 0.1

Activation Function:

Sigmoid

Optimizer:

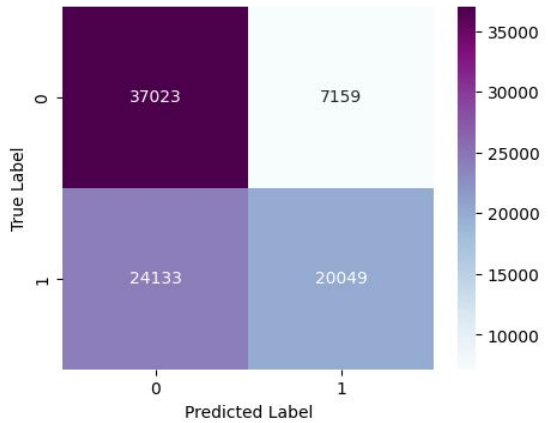
Adam

```
Hyperparameters 1: Hidden Size 1: 256, Hidden Size 2: 256, Learning Rate: 0.1000, Activation Function: Sigmoid, Optimizer: Adam
  -- Epoch 0, Training Loss: 0.7060
  -- Epoch 2, Training Loss: 0.7314
  -- Epoch 4, Training Loss: 0.7303
  -- Epoch 6, Training Loss: 0.7283
  -- Epoch 8, Training Loss: 0.7316
  -- Epoch 9, Training Loss: 0.7306
  Test Loss: 0.5252
```

Lower test loss = better performance

Performance Evaluation

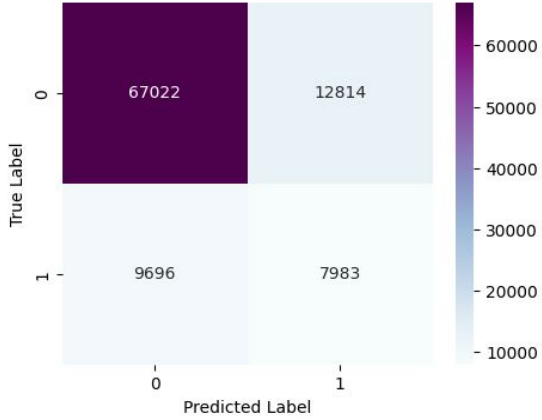
Train Data



Classification Report for Neural Network - Train Data

	precision	recall	f1-score	support
0	0.61	0.84	0.70	44182
1	0.74	0.45	0.56	44182
accuracy			0.65	88364
macro avg	0.67	0.65	0.63	88364
weighted avg	0.67	0.65	0.63	88364

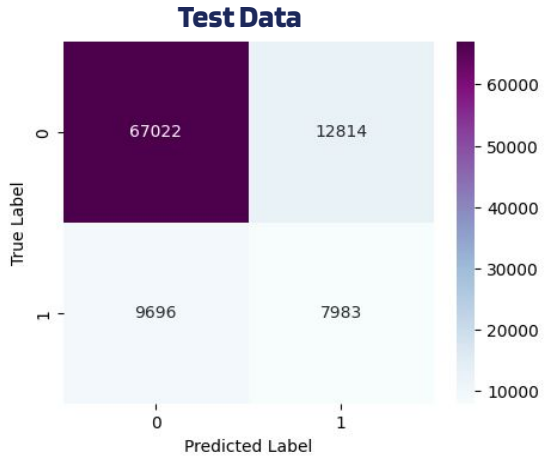
Test Data



Classification Report for Neural Network - Test Data

	precision	recall	f1-score	support
0	0.87	0.84	0.86	79836
1	0.38	0.45	0.41	17679
accuracy			0.77	97515
macro avg	0.63	0.65	0.64	97515
weighted avg	0.78	0.77	0.78	97515

Performance Evaluation



Classification Report for Neural Network - Test Data

	precision	recall	f1-score	support
0	0.87	0.84	0.86	79836
1	0.38	0.45	0.41	17679
accuracy			0.77	97515
macro avg	0.63	0.65	0.64	97515
weighted avg	0.78	0.77	0.78	97515

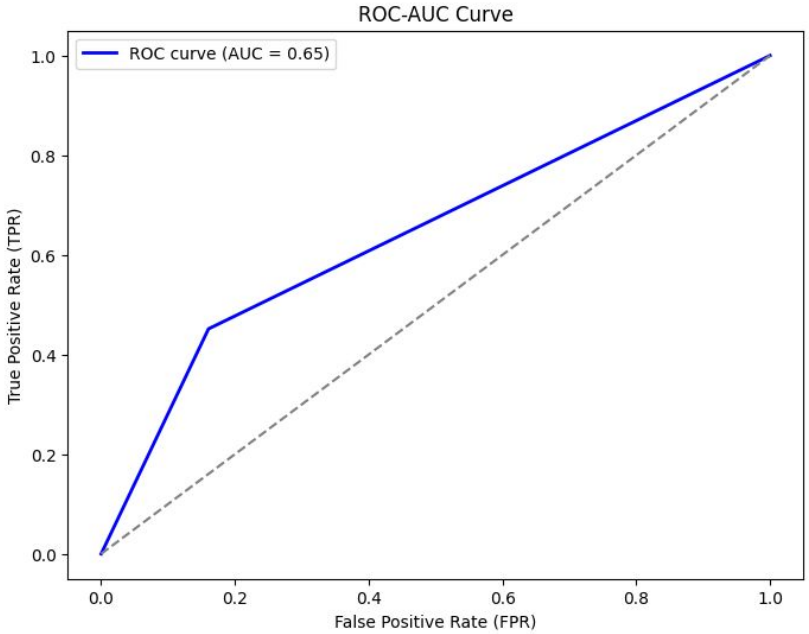
No Churn: 0

The model is predicting no churn 84% of the time, 87% of these calls are correct!

Churn: 1

The rate of predicting churn is 45% but the amount of times this is actually correct is 38%.

Performance Evaluation



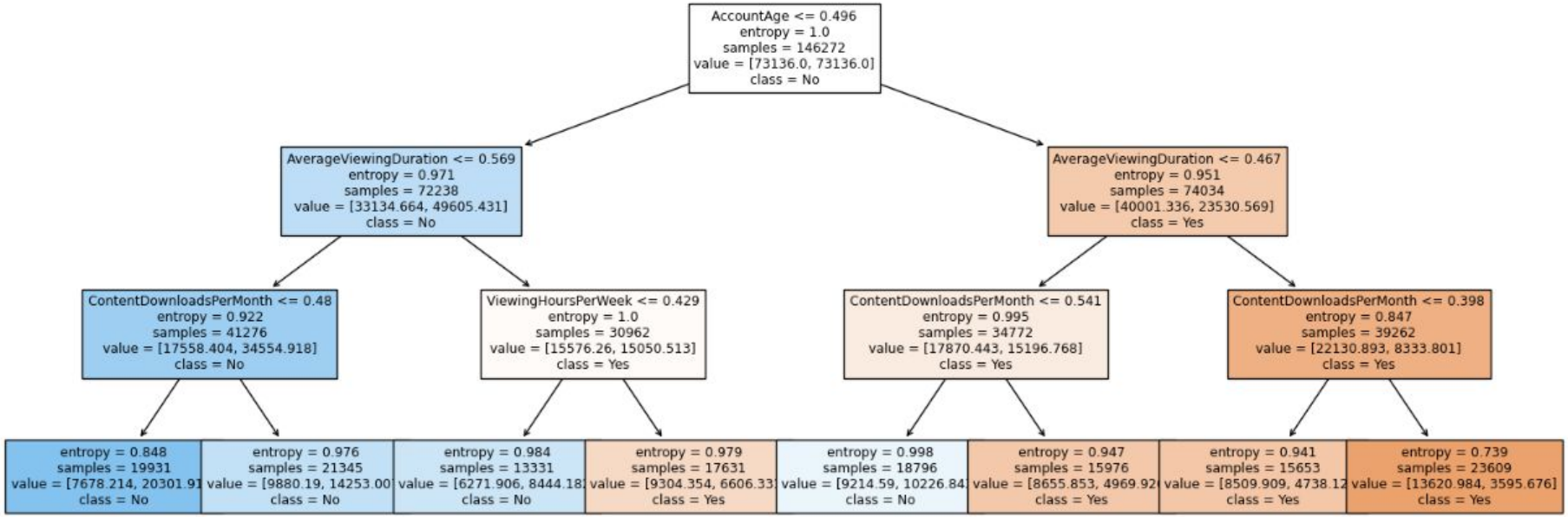
- **ROC Curve score = 0.65**
  - Demonstrates moderate ability/effectiveness to predict churn

- **Favorable Curve: Top RIght**
  - Shows good tradeoff between sensitivity (true positive rate) and specificity (false positive rate)

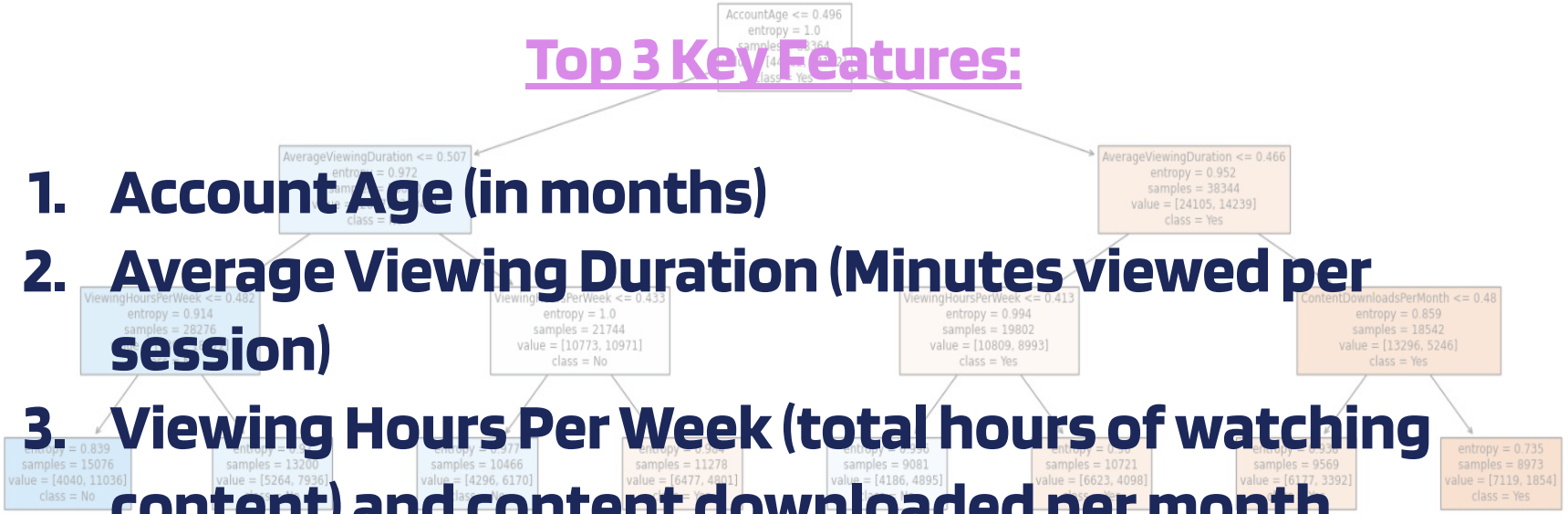
04

Machine Learning Model

Decision Tree Model



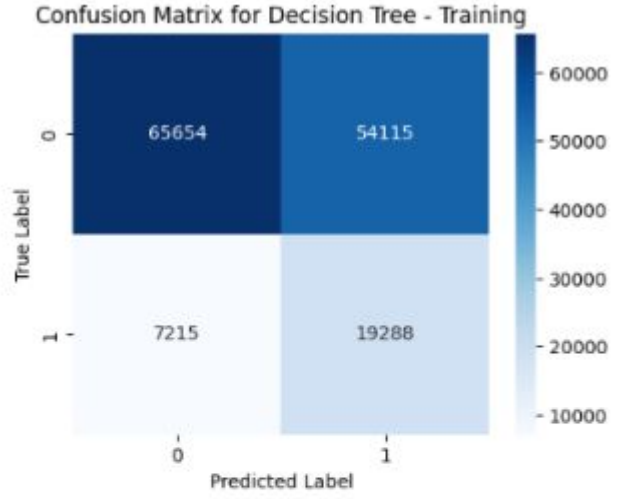
Decision Tree Model



Top 3 Key Features:

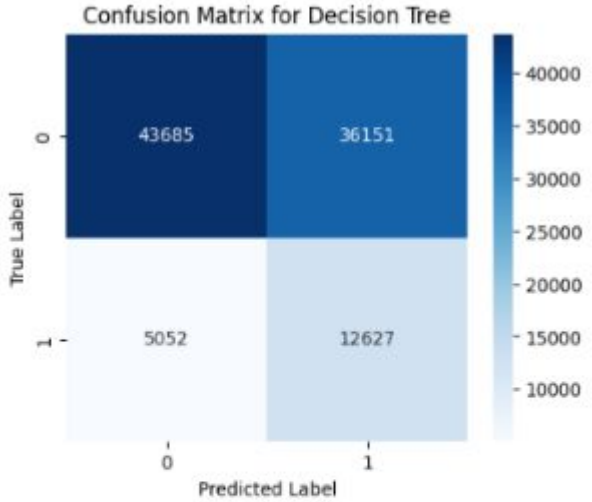
1. Account Age (in months)
2. Average Viewing Duration (Minutes viewed per session)
3. Viewing Hours Per Week (total hours of watching content) and content downloaded per month

Decision Tree Model



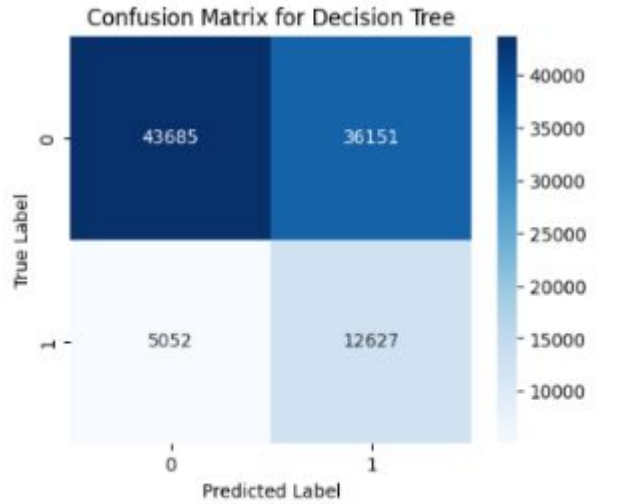
Classification Report for Decision Tree - Train Data

	precision	recall	f1-score	support
0	0.90	0.55	0.68	119769
1	0.26	0.73	0.39	26503
accuracy			0.58	146272
macro avg	0.58	0.64	0.53	146272
weighted avg	0.79	0.58	0.63	146272



	precision	recall	f1-score	support
0	0.90	0.55	0.68	79836
1	0.26	0.71	0.38	17679
accuracy			0.58	97515
macro avg	0.58	0.63	0.53	97515
weighted avg	0.78	0.58	0.63	97515

Decision Tree Model



	precision	recall	f1-score	support
0	0.90	0.55	0.68	79836
1	0.26	0.71	0.38	17679
accuracy			0.58	97515
macro avg	0.58	0.63	0.53	97515
weighted avg	0.78	0.58	0.63	97515

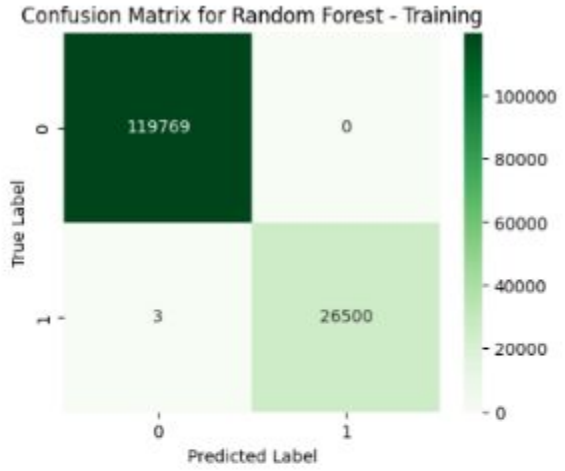
No Churn: 0

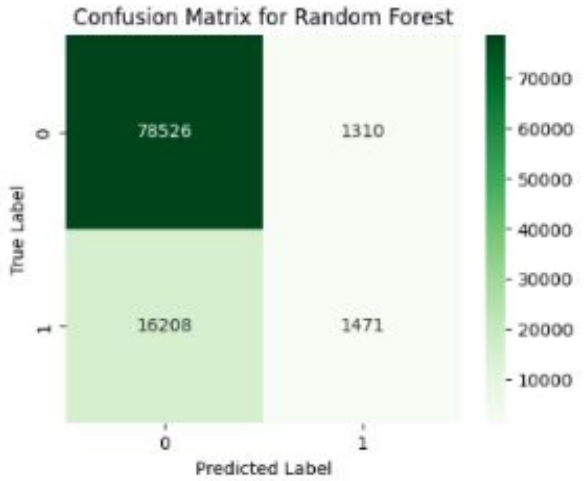
The precision is higher for predicting not churn, which as we know by now is the safe prediction. The model is predicting no churn 55% of the time, which probably contributed by the balanced data. 90% of these calls are correct!

Churn: 1

The rate of predicting churn is 71% but the amount of times this is actually correct is 26%.

Random Forest Model





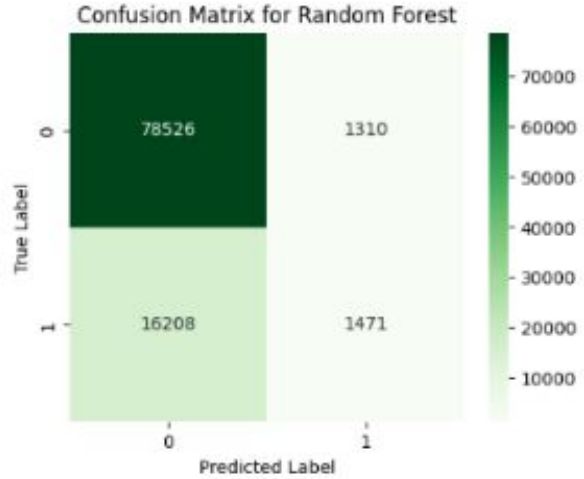
Classification Report for Random Forest - Train Data:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	119769
1	1.00	1.00	1.00	26503
accuracy			1.00	146272
macro avg	1.00	1.00	1.00	146272
weighted avg	1.00	1.00	1.00	146272

Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.83	0.98	0.90	79836
1	0.53	0.08	0.14	17679
accuracy			0.82	97515
macro avg	0.68	0.53	0.52	97515
weighted avg	0.77	0.82	0.76	97515

Random Forest Model



Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.83	0.98	0.90	79836
1	0.53	0.08	0.14	17679
accuracy			0.82	97515
macro avg	0.68	0.53	0.52	97515
weighted avg	0.77	0.82	0.76	97515

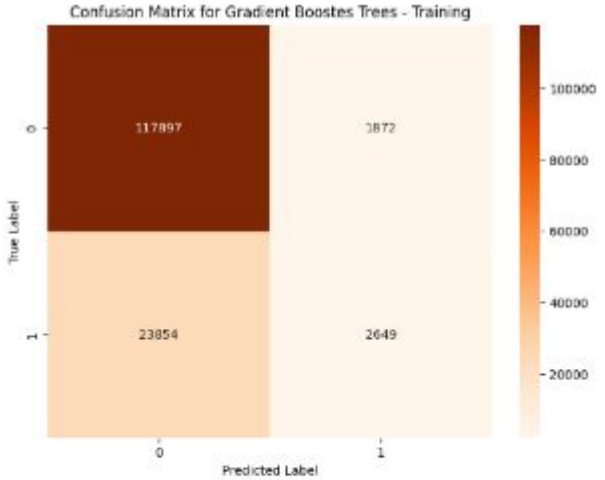
No Churn: 0

The model is predicting no churn 98% of the time, which probably contributed by the balanced data. Only 83% of these calls are correct!

Churn: 1

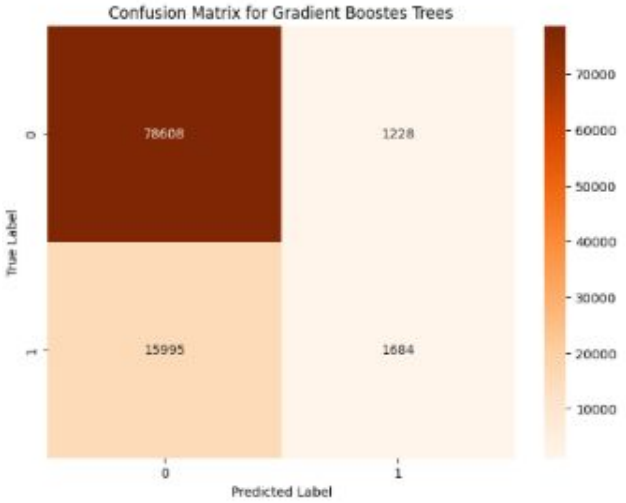
The rate of predicting churn is 8% but the amount of times this is actually correct is 53%.

Gradient Boosted Trees Model



Classification Report for Gradient Boosted Tree - Train Data

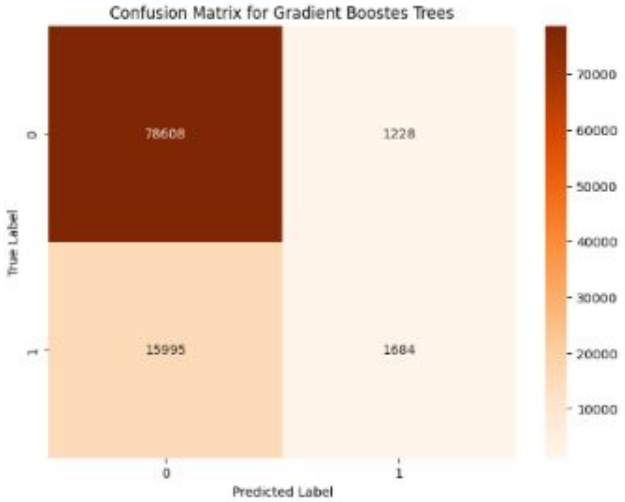
	precision	recall	f1-score	support
0	0.83	0.98	0.90	119769
1	0.59	0.10	0.17	26503
accuracy			0.82	146272
macro avg	0.71	0.54	0.54	146272
weighted avg	0.79	0.82	0.77	146272



Classification Report:

	precision	recall	f1-score	support
0	0.83	0.98	0.90	79836
1	0.58	0.10	0.16	17679
accuracy			0.82	97515
macro avg	0.70	0.54	0.53	97515
weighted avg	0.79	0.82	0.77	97515

Gradient Boosted Trees Model



Classification Report:

	precision	recall	f1-score	support
0	0.83	0.98	0.90	79836
1	0.58	0.10	0.16	17679
accuracy			0.82	97515
macro avg	0.70	0.54	0.53	97515
weighted avg	0.79	0.82	0.77	97515

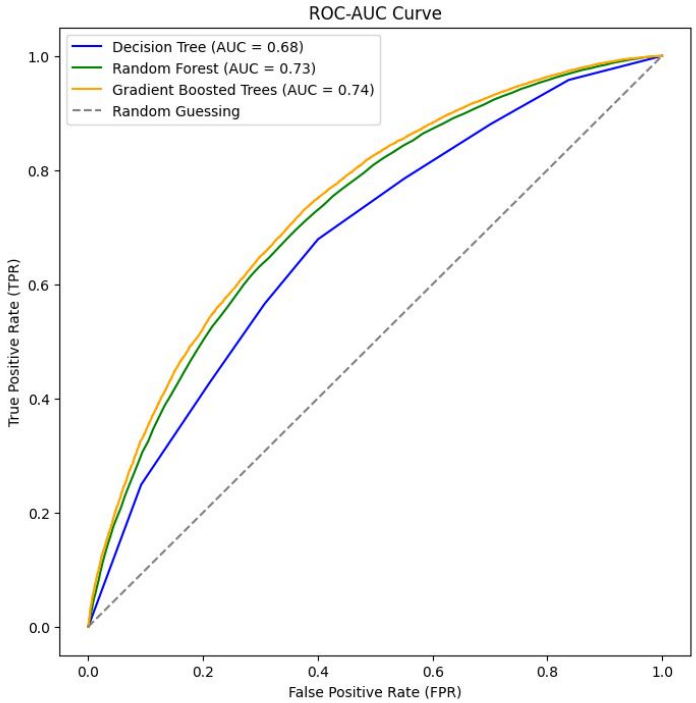
No Churn: 0

The model is predicting no churn 98% of the time, which probably contributed by the balanced data. Only 83% of these calls are correct!

Churn: 1

The rate of predicting churn is 10% but the amount of times this is actually correct is 58%

ROC-AUC Curve

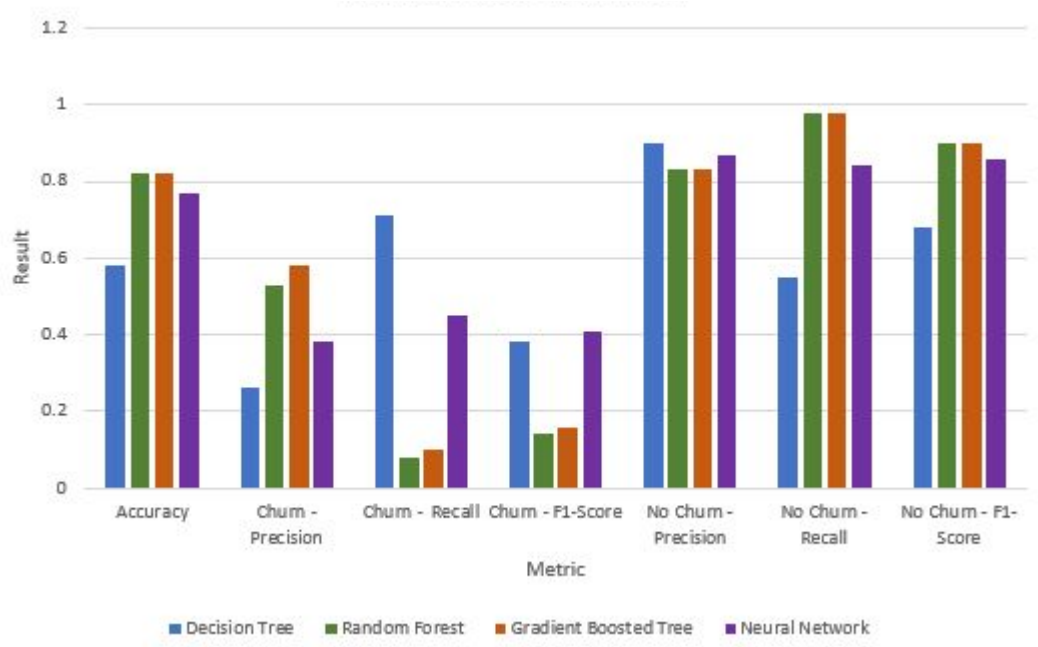


- The winning model on the ROC-AUC Model is the Gradient Boosted Trees!
- The Random Forest is a very close second
- The Decision Tree did have the best Precision for predicting no churn but that is not a measure of best overall performance.
- There is higher sensitivity and fewer false positives.
- All models are better than random guessing

05

Performance Comparison

Model Comparison & Analysis



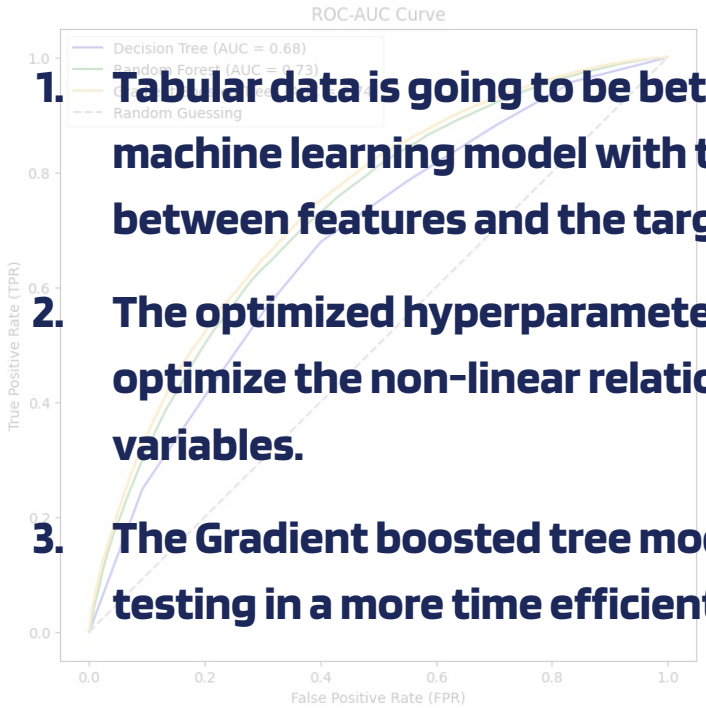
Observations

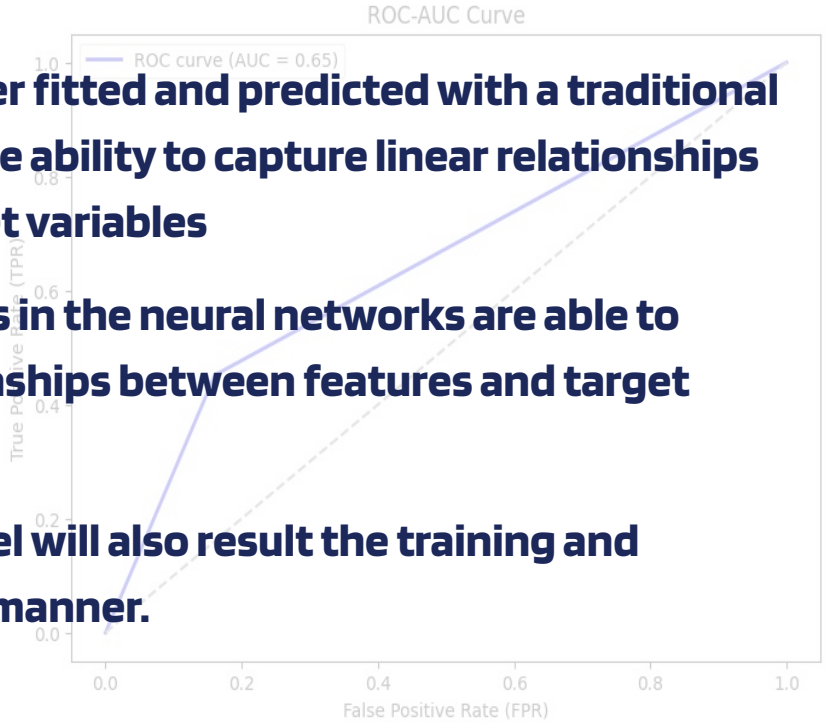
Gradient Boosted Tree ranks #1:

- Highest accuracy
- Balanced precision and recall for both classes
- Highest AUC score

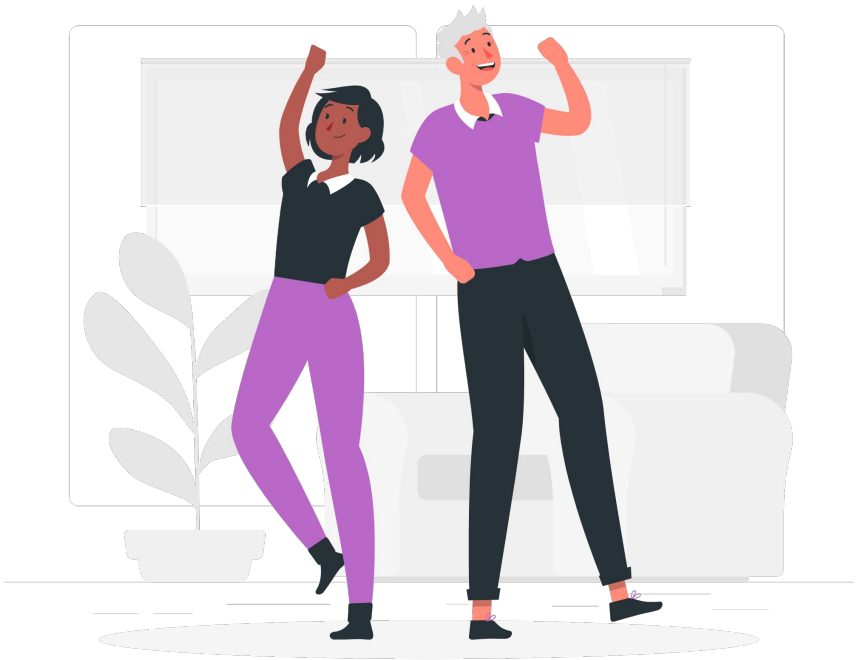
⇒ Overall better performance

Model Comparison & Analysis





1. **Tabular data is going to be better fitted and predicted with a traditional machine learning model with the ability to capture linear relationships between features and the target variables**
2. **The optimized hyperparameters in the neural networks are able to optimize the non-linear relationships between features and target variables.**
3. **The Gradient boosted tree model will also result the training and testing in a more time efficient manner.**



06

Conclusion

Recommendations

1. Share Feature Importance of identified features contributing to churn
2. Personalized Marketing Strategies to increase customer retention and satisfaction
3. Dynamic Pricing Modeling to align with customer preferences

Thanks!

Do you have any questions?

Group 9
Mahaam Ahmed, Natalie Nguyen & Natalie Rath