Losing Customers?

By Collin Loo

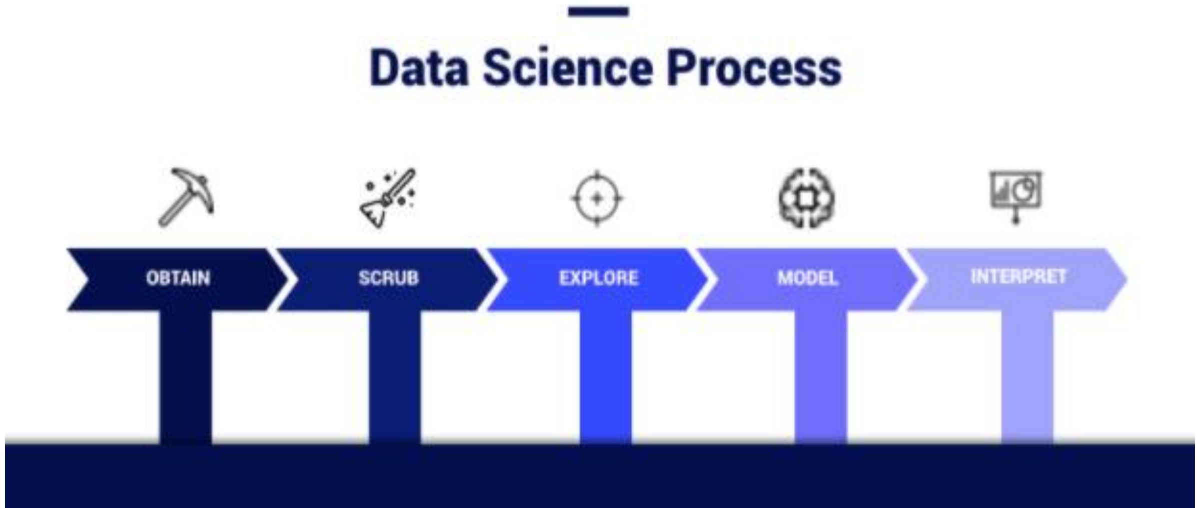
Why Should We Care?

- In general, it cost 5 to 7 times more to acquire new customers.
- Studies show it is easier to sell cross products to existing customers than to new subscribers.
- Losing customers hinders growth and represents lost investments.

Our Objectives

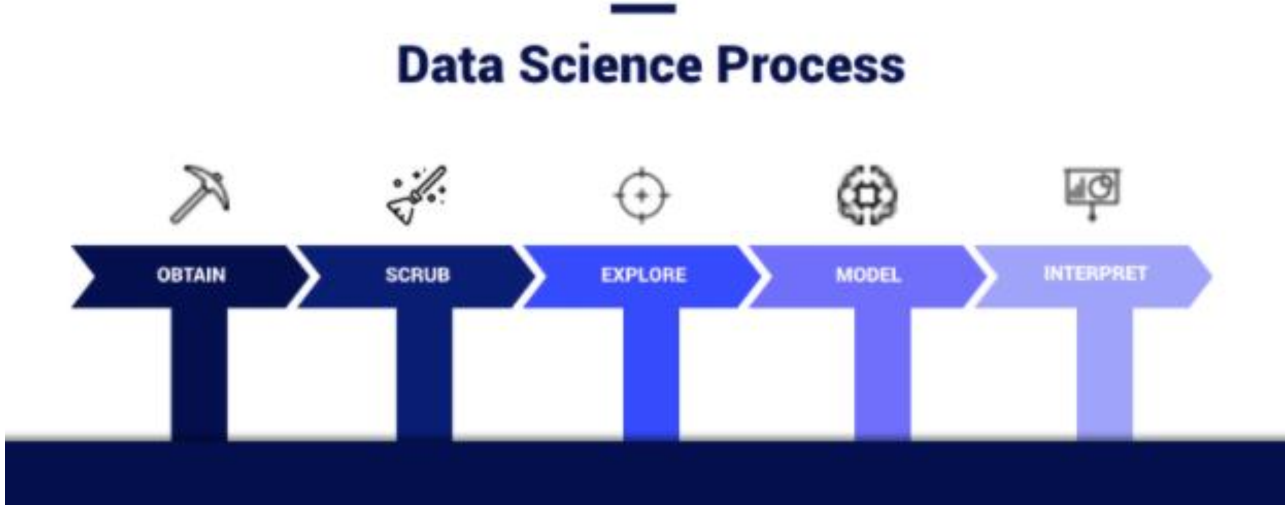
- Use Machine Learning (ML) to identify the features that contribute to customer churn.
- Present to the Customer Service, Marketing and Sales departments.
- Make recommendations.

The OSEMN Framework



- The OSEMN framework provides high level guidance on handling a data science project.
- Following the processes outlined by OSEMN allows for efficient data modeling.

- Data come from within the company.
- In the Scrub process:
  - Examine columns to determine if they are text or numbers?
  - Remove columns that do not contribute values to our ML model, such as phone numbers.
  - Handle missing values in columns, if any.



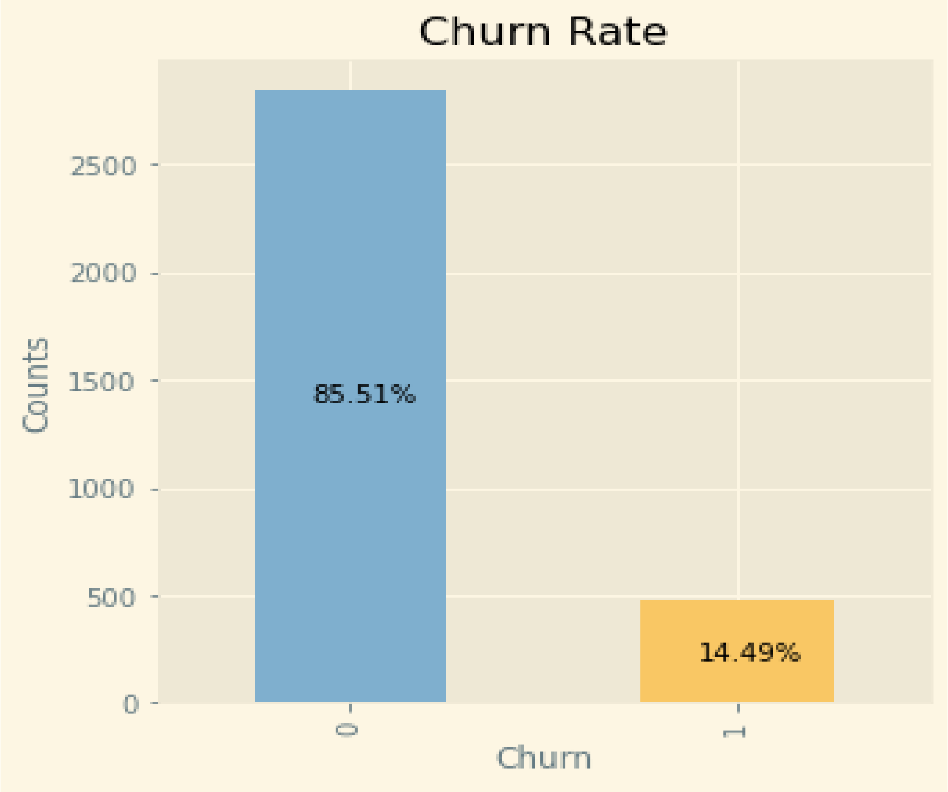
The OSEMN Framework

Obtain and Scrub

The OSEMN Framework

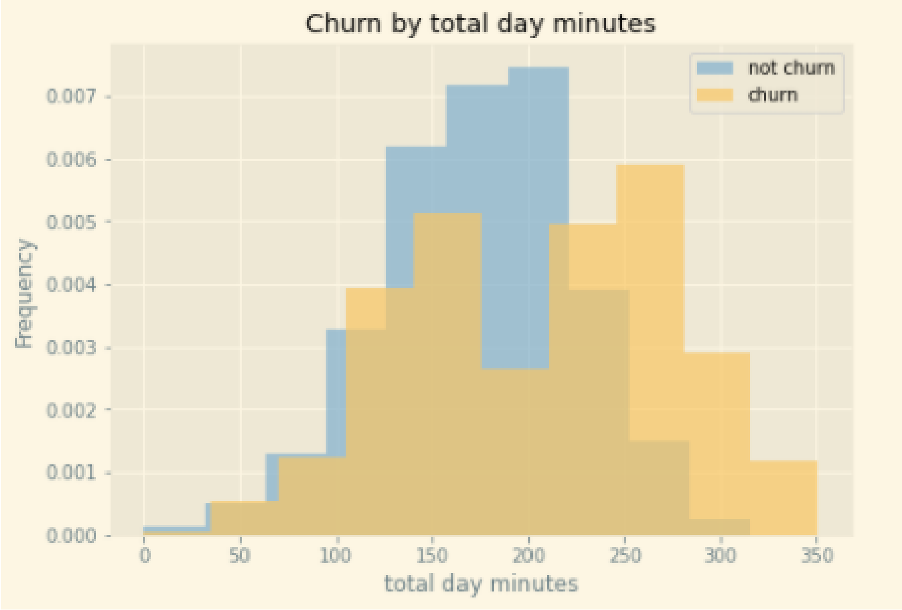
Explore Process

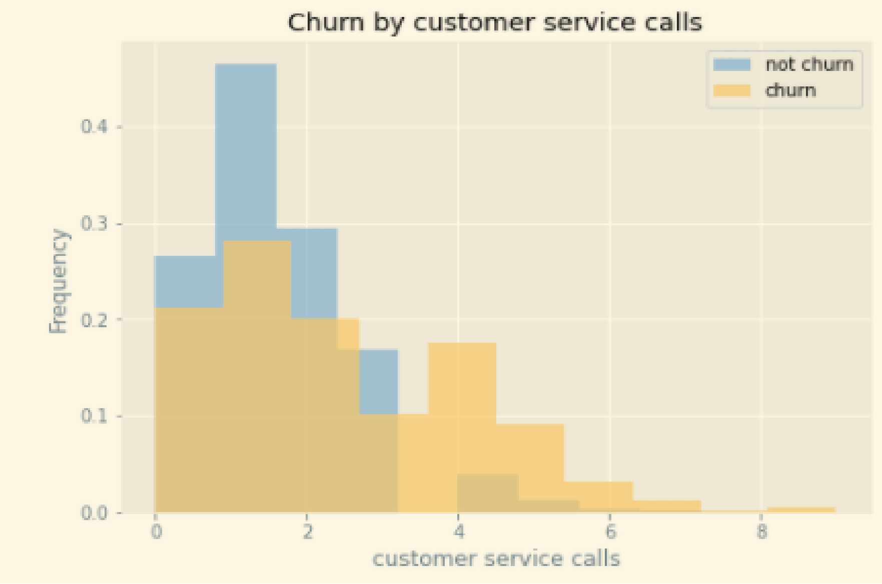
- There are 3,333 records in the dataset.
- The oldest account in the dataset is 8 months old. The average account age is 3.5 months.
- The dataset is composed of mainly new customers.
- We've lost 14.49% or 500 fairly new customers.



The OSEMN Framework

Explore Process



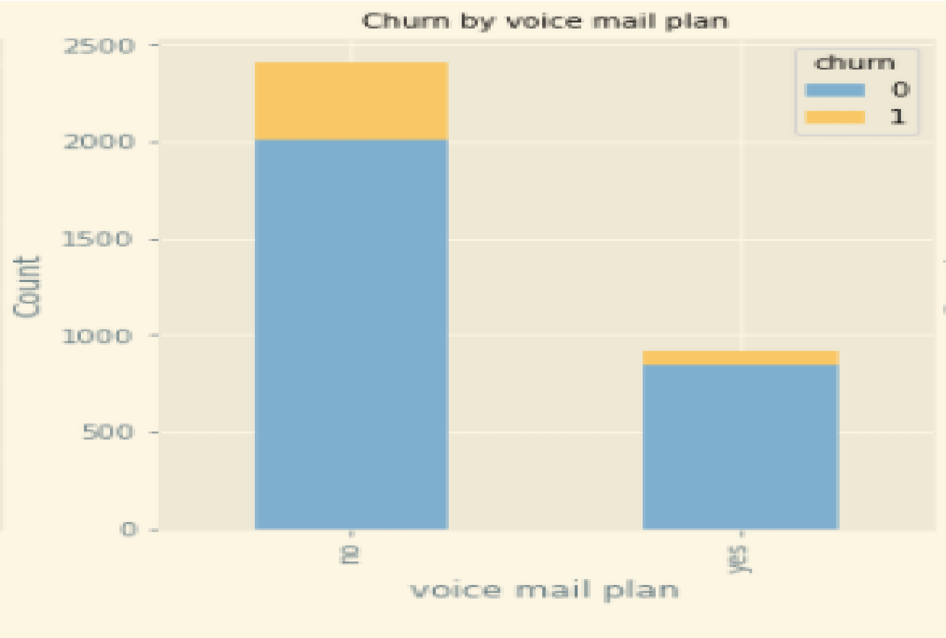


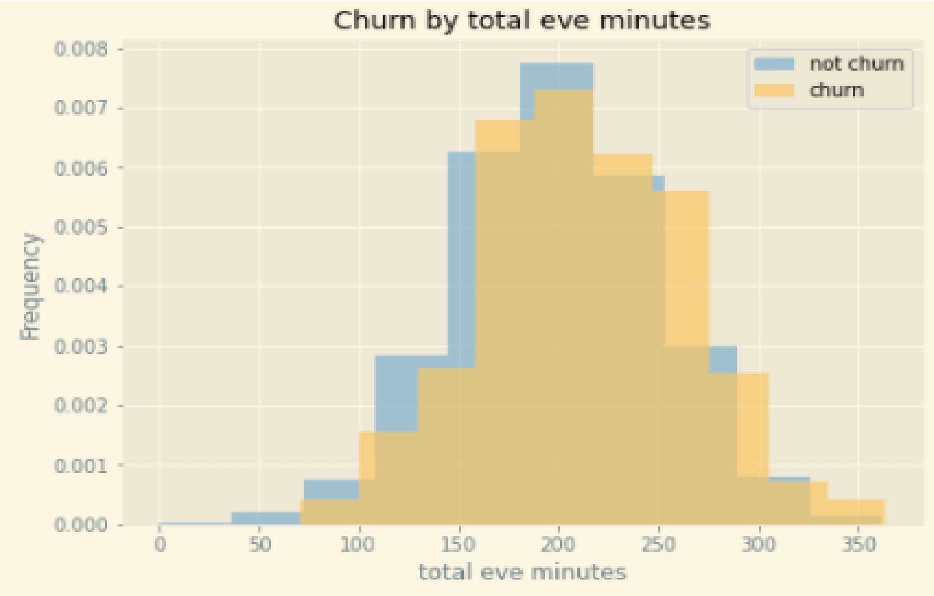
- Customers with heavy daily usage are more likely to leave than customers with less total daily usage .
- We tend to lose customers when they make a high number of calls to our customer service lines.

The OSEMN Framework

Explore Process

- Customers without a voice mail plan leave more often than those that have a voice mail plan.
- For customers with heavy usage during evening hours, the higher the risk they will leave.





Classification Algorithms

- In predicting customer Churn, we are basically predicting a Yes or No outcome. In data science, this is commonly known as a classification problem.
- There are several methods at our disposal to address this issue. We will evaluate four algorithms at this stage, namely Logistic Regression, K-Nearest Neighbors, Random Forest and XGBoost.
- The method that can predict more true positives and less false negatives will be our final model.

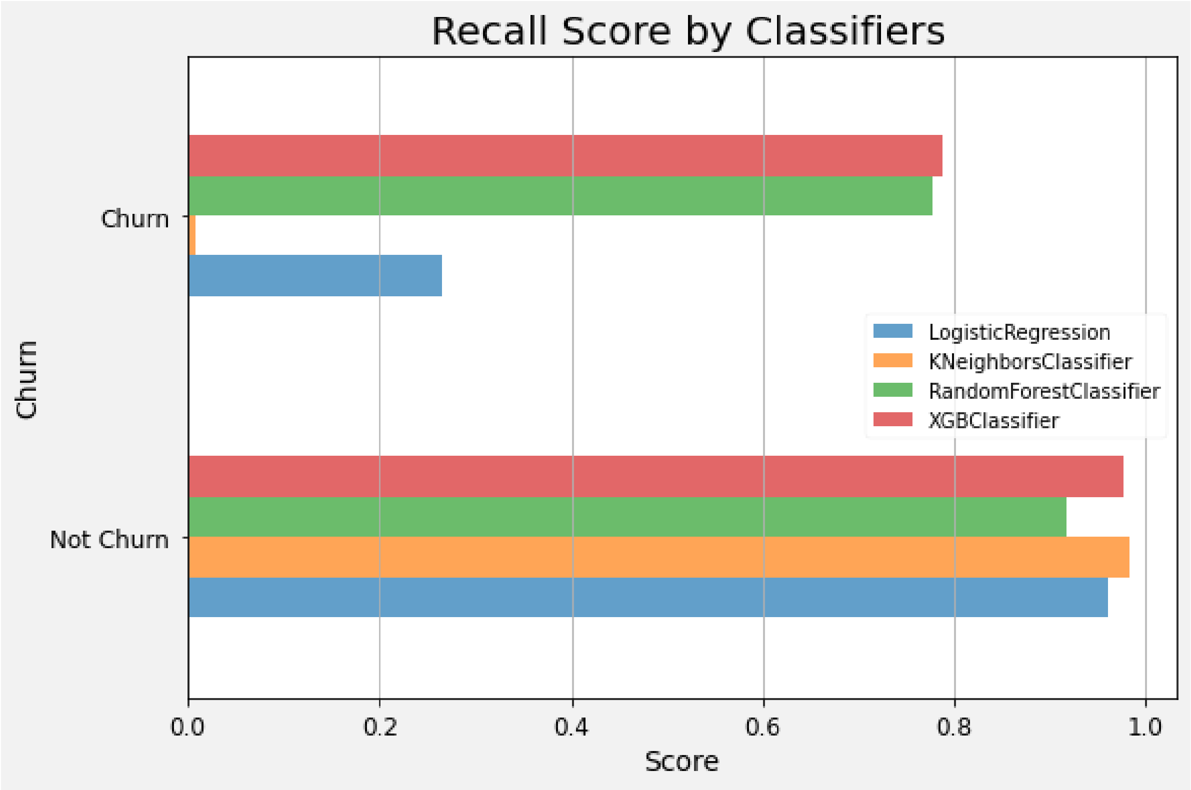
The OSEMN Framework

Model Process

The OSEMN Framework

Model Process – Final Model

- As our objective is to maximize model prediction for true positives and minimize false negatives, this information is captured in the recall score.
- The XGBoost algorithm delivers the highest recall scores in predicting both the ‘Churn’ and ‘Not Churn’ labels.

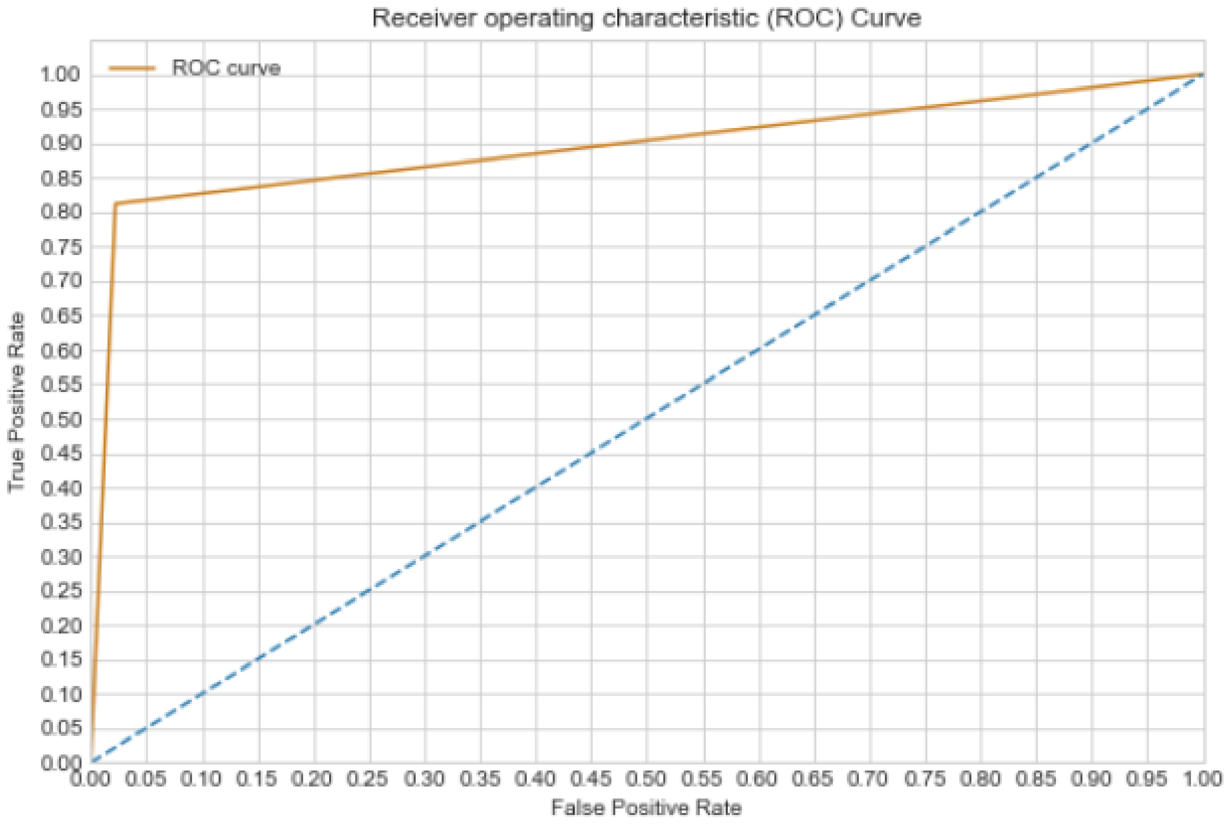


The OSEMN Framework

Model Process – Evaluation

- A useful model not only works well on controlled data but also need to perform equally well for unseen data.
- Our final model yields a 99.92 accuracy score on controlled data and 95.2 accuracy score on unseen data. Therefore, it will work well with new data in the future.
- The AUC score is another measure for model performance. Our model has an AUC score of 89.48% which means it has a probability of 89.48% chance of correctly distinguishing between the 'Churn' and 'Not Churn' labels.

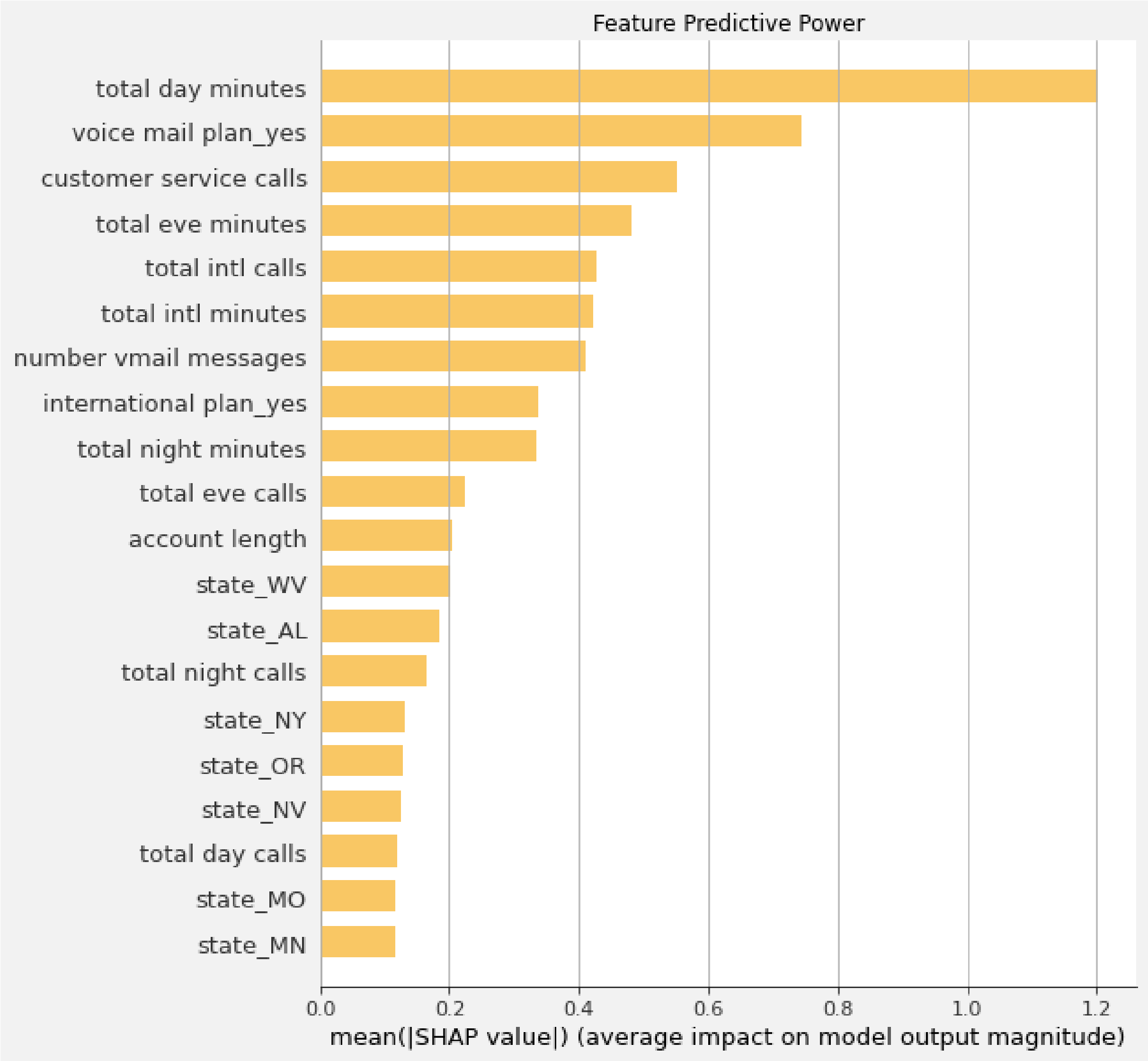
***** AUC Score: 89.48% *****



The OSEMN Framework

Model Interpretation

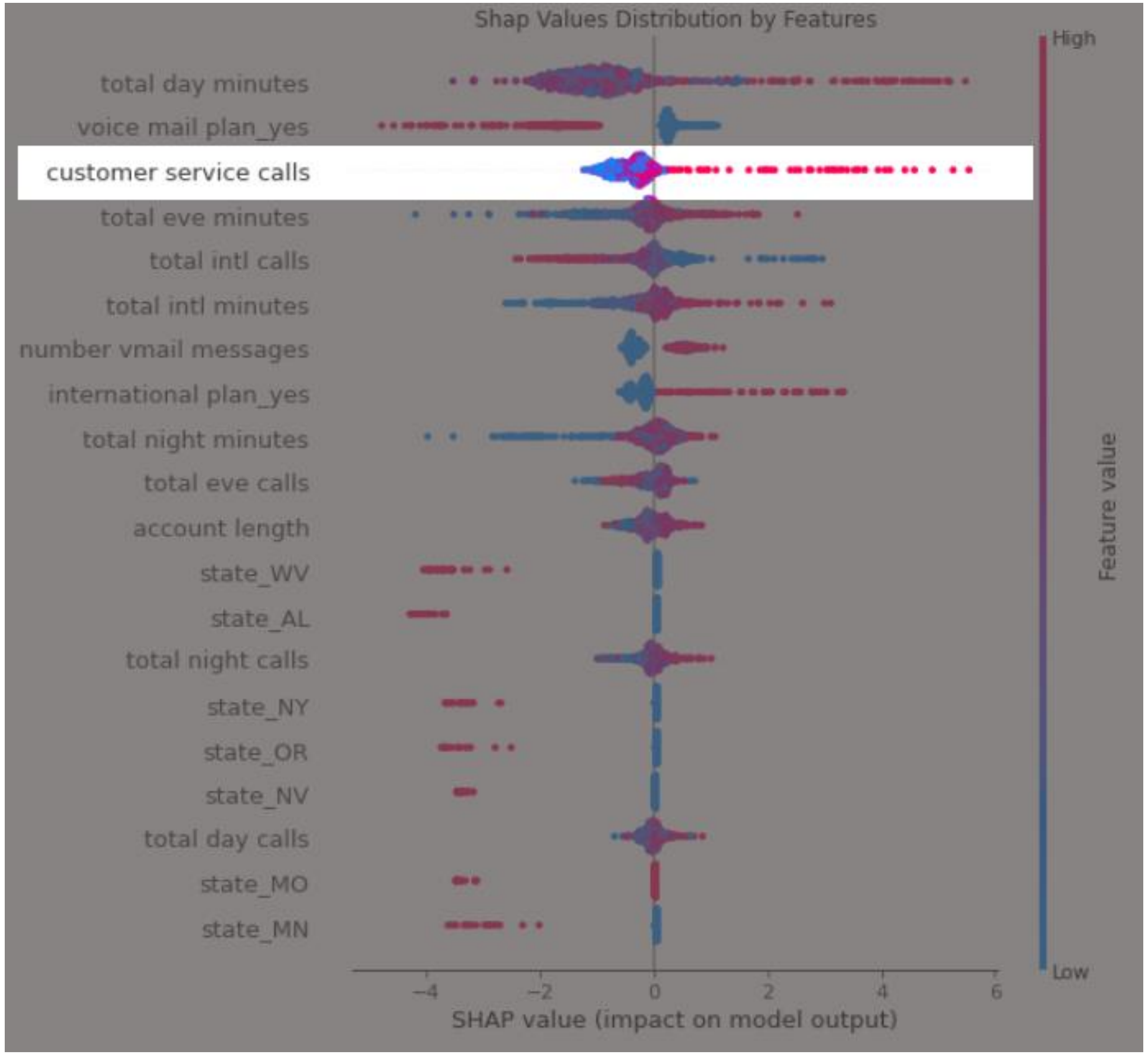
- We utilize a SHAP package to help us explain our model.
- The feature that impacts our model prediction for Churn the most is ‘total day minutes’.
- ‘Voice mail plan_yes’ comes in second, followed by ‘customer service calls’ and ‘total eve minutes’.
- ‘Total day minutes’ has twice the influence over the model outcomes than ‘customer service calls’.



The OSEMN Framework

Model Interpretation

- This plot explains how an individual customer with a given feature contributes to the model outcome.
- Customers without a voice mail plan or with low voice mail messages (blue dots) have a tendency to leave.
- In the case of number of call to service lines, the more frequent a customer contacts our service lines, the more likely he/she will leave.



Recommendations

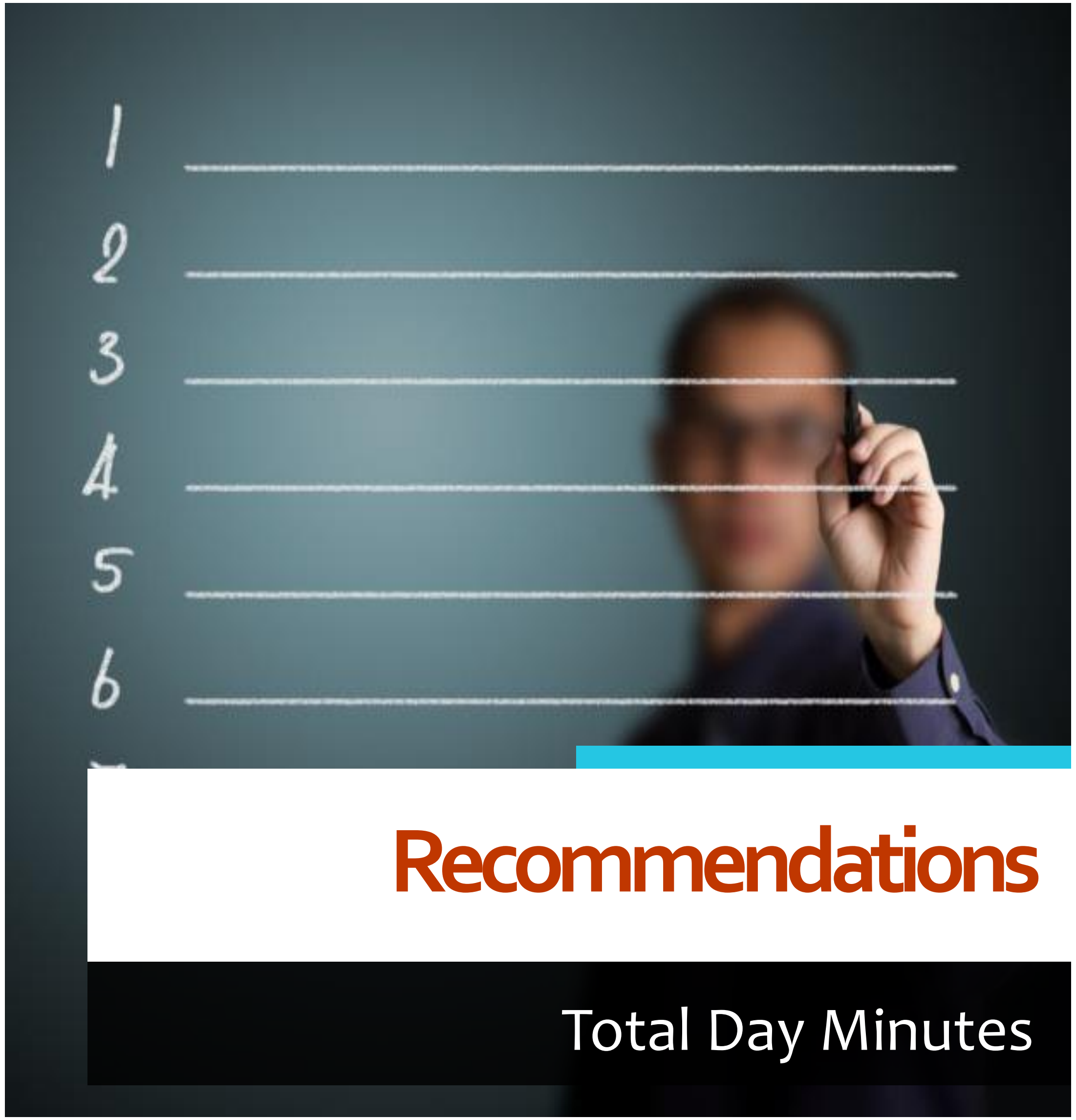
- We will use the top four most influential features to make recommendations on how to stop customer churn.

Issue

Customer with heavy daily usage are most likely to leave.

Proposal

- Micro segment day time minutes rate. Review current day time rate structures to see if additional rate levels make sense.
- ‘Frequent-flyers’ approach. Create VIP class for heavy usage users and reward them with benefits.



Recommendations

Total Day Minutes

Issue

Customers without a voice mail plan are most likely to leave.

Proposal

- Undertake comparative analysis to see what our competitors are doing.
- Match or beat their offerings.

Recommendations

Voice Mail Plan - Yes

Recommendations

Customer Service Calls

Issue

Customers with more calls to the customer service lines are most likely to leave.

Proposal

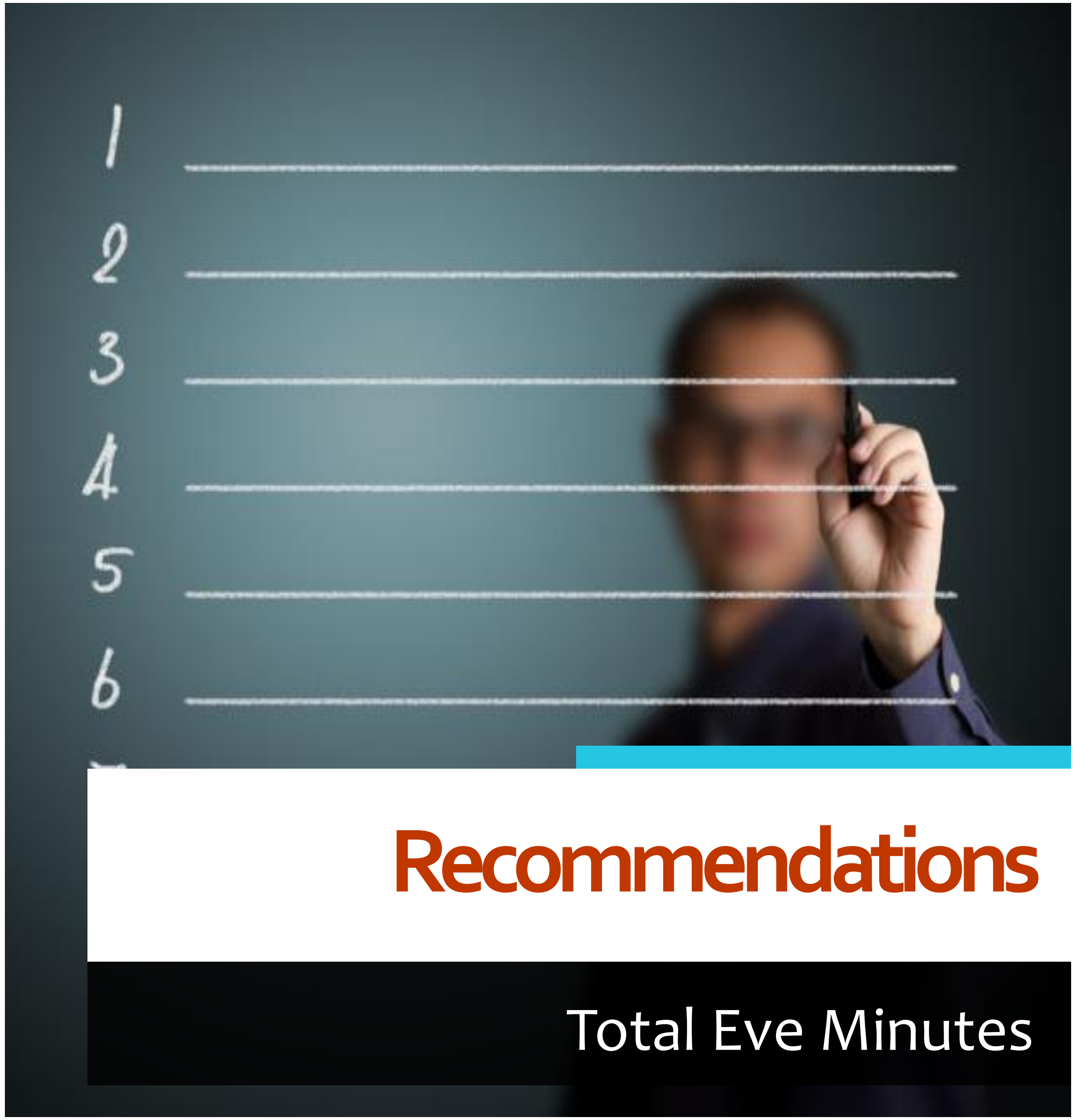
- Review nature of the calls.
- Conduct customer satisfaction surveys.
- Evaluate performance of customer service agents.

Issue

Customers with heavy evening usage are most likely to leave.

Proposal

- We recommend taking the same proposal for total day minutes, that is micro segment evening call rates and VIP status.



Recommendations

Total Eve Minutes

Future Work

- Include other algorithms like Bayesian Classification and Support Vector Machines in our model.
- Employ different data sampling techniques.
- Introduce additional algorithm tuning parameters.

Appendix

Model Selection Process

- [Model Baseline]
- [Model Tuning vs Baseline]
- [Model Tuning Results]

Change things

At Flatiron School you learn how the future is being built, so you can change anything, starting with a new career in code, data science, or cybersecurity.

Thank You

Collin Loo

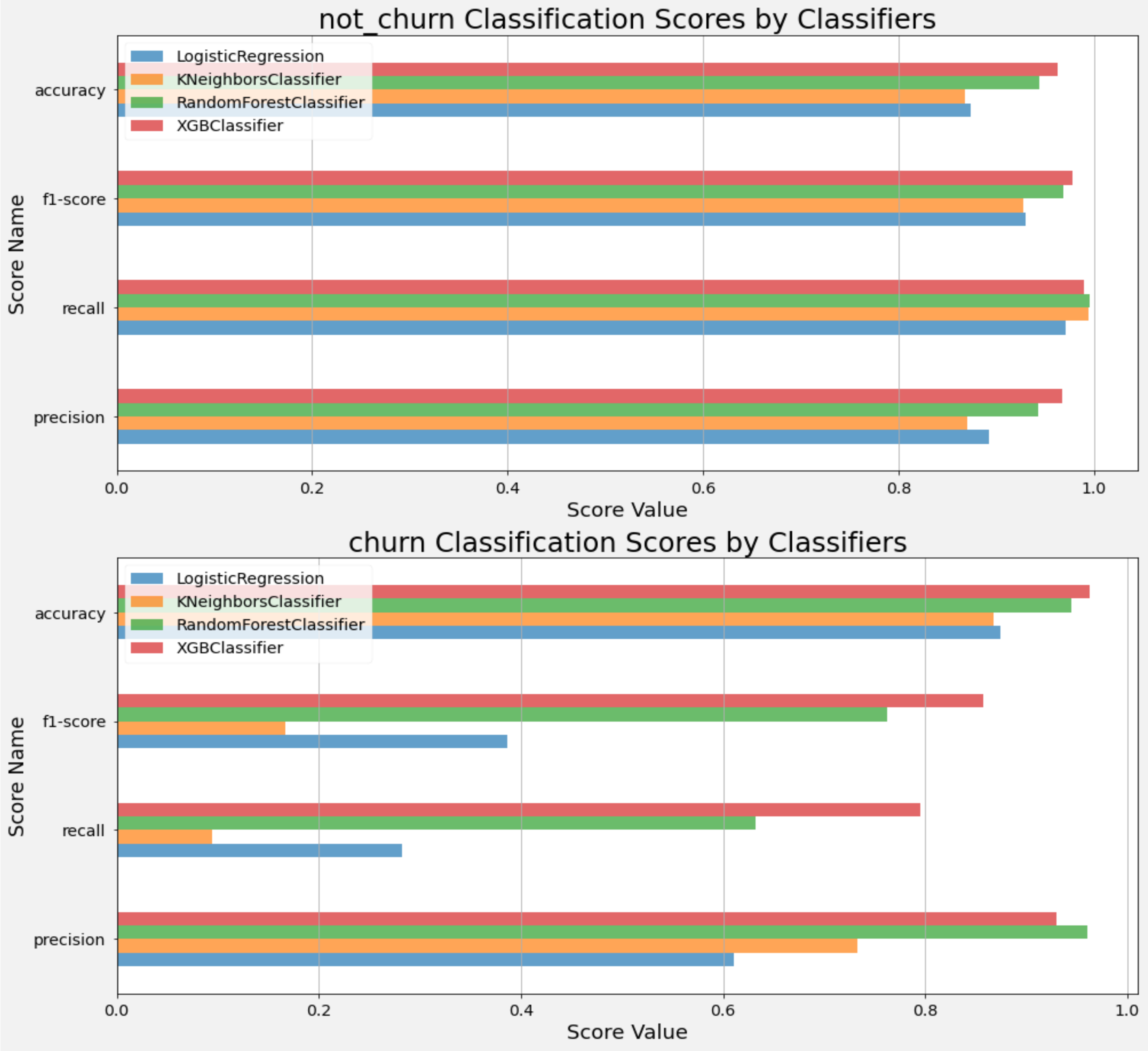
collinloo@yahoo.com

[Flatiron School]

The OSEMN Framework

Model Process – Baseline

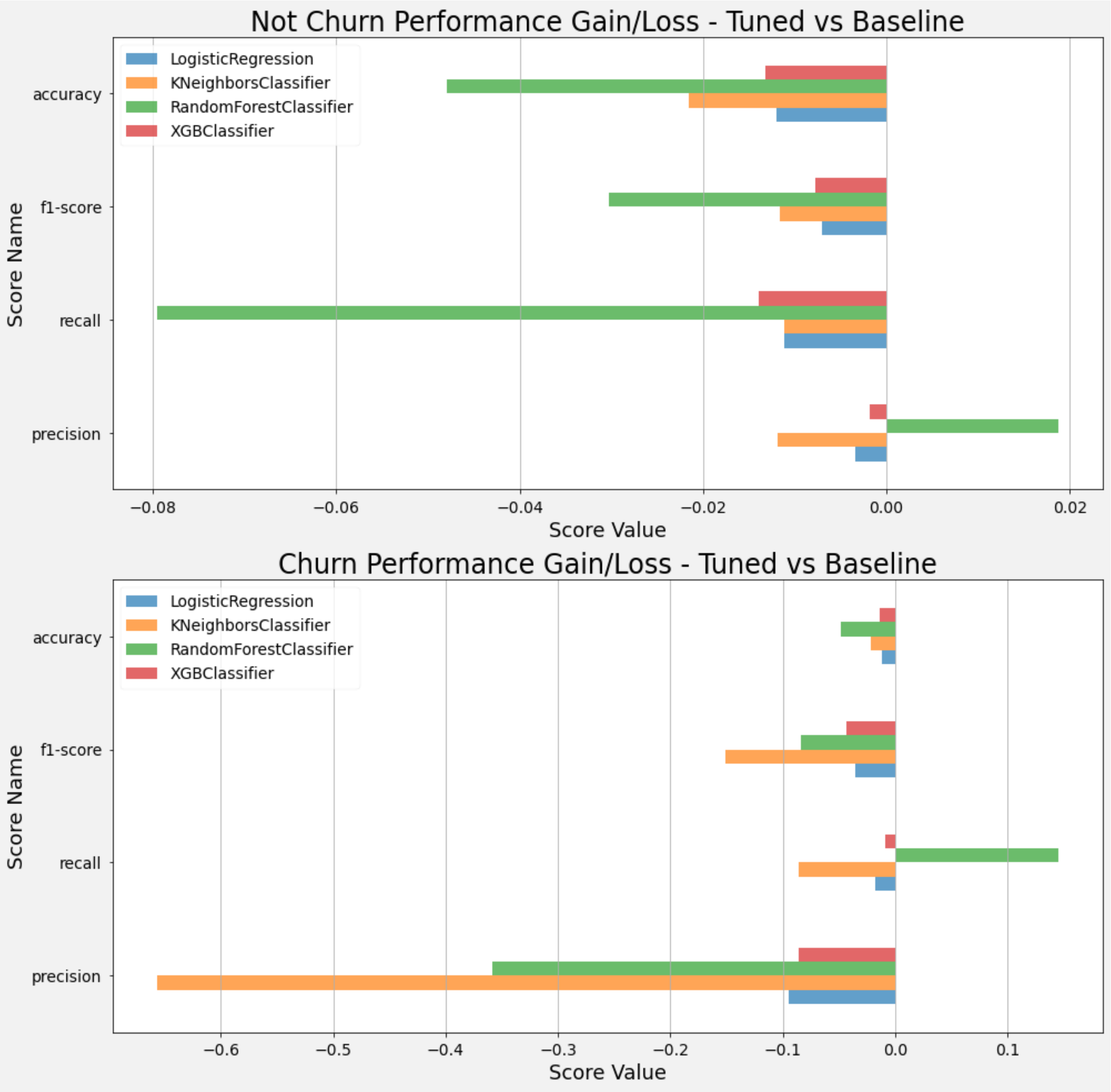
- To judge how well a model performs, one approach is to look at the various scores it produces, such as the precision score, the recall score, the F1_score and the accuracy score.
- As our objective is to maximize model prediction for true positives and minimize false negatives, this information is captured in the recall score.
- The XGBoost algorithm delivers the highest recall scores in predicting both the 'Churn' and 'Not Churn' labels.



The OSEMN Framework

Model Process – Tuning vs Baseline

- Due to the fact that the data contains less number of customers leaving, this can cause an issue in our model. We address this by resampling our data.
- The models themselves come with various parameters for fine tuning. We ran our models through different combinations of parameters to find the best one.



The OSEMN Framework

Model Process – Tuning Results

- XGBoost method outperformed the other three methods in most scoring. It will be our model of choice.