

Predicting Customer Behavior

Data Crew:

Darren Hsieh, Taro Iyadomi, Ryan Kawamura, Axel Malvaez, Daniel O'Brien

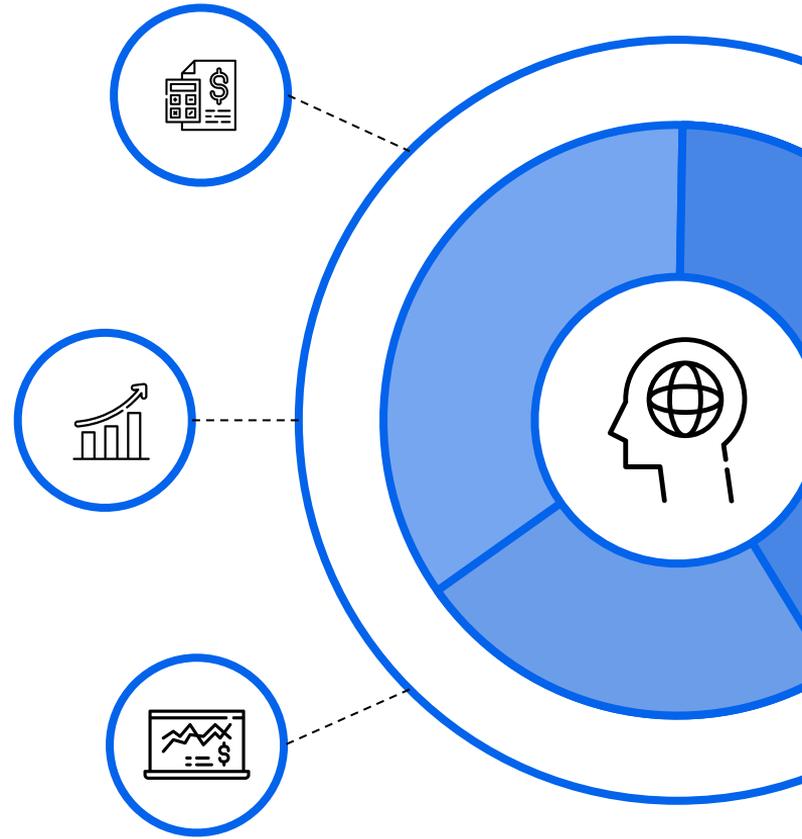


Table of Contents



Introduction



Data Preparation



Data Exploration



Classification Models



Results



Conclusion

Introduction

Our Team's Goal

Create a classification model capable of predicting whether a customer would make a purchase based on their first couple interactions on the website, providing Fingerhut a model to leverage to learn more about their customers and hence help drive business

Data Preparation

Data Preparation Steps

01

Duplicate Removal

With journey step
reindexing



02

Widening

Allocating each row
to a unique customer



04

Dimension Reduction

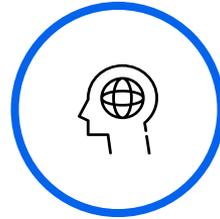
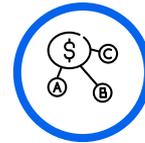
Reducing size while
retaining information



03

Feature Engineering

Extracting information
from the data



Duplicate Removal & Widening

Duplicate Removal:

1. Removed duplicate rows with identical timestamps.
2. Reindexed journey_steps_until_end column to retain sequential property.

Journey Steps Until End Reindexing

customer_id	J.S.U.E (Before)	J.S.U.E (After)
2787130	1	1
2787130	3	2
2787130	4	3
2787130	9	4

Widening:

1. Grouping data by customer, leading to a unique customer per row.
2. Merged data from multiple accounts.
3. first_n_events column: list of the first n (n=5, 20) events of the customer's first journey.
4. time_since_last_event column: list of the time deltas between each of the n events from first_n_events.

Feature Engineering

Total Features

27

1. Customer Related Variables:
 - No. of accounts, no. of journeys, avg. journey length, initial device, etc.
2. Event/Stage/Milestone Related Variables:
 - Has_x ($x \in$ events/stages/milestone), mode events, first n events, etc.
3. Time Related Variables:
 - Time_in_y ($y \in$ stages), total time, avg. time between events, etc.

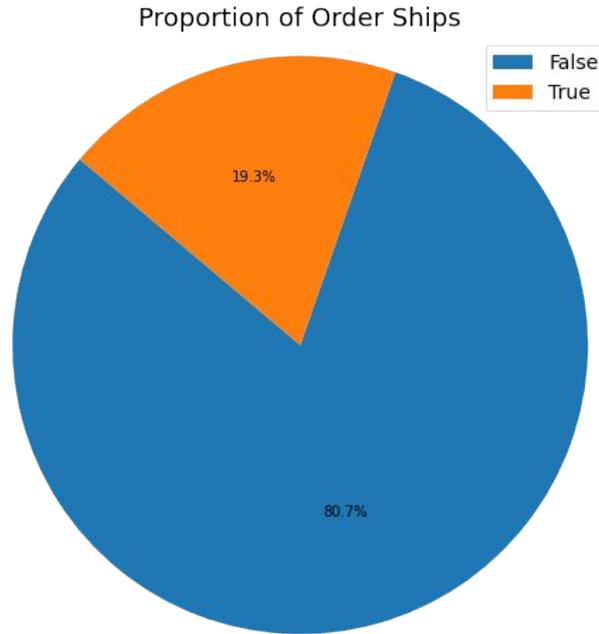
Dimension Reduction

1. Embedding:
 - Training LSTM to generate low dimensional vector representations of first_n_events and time_since_last_event variables.
2. Sampling:
 - Taking randomized subsets of customers (5%) for early analysis and modeling.

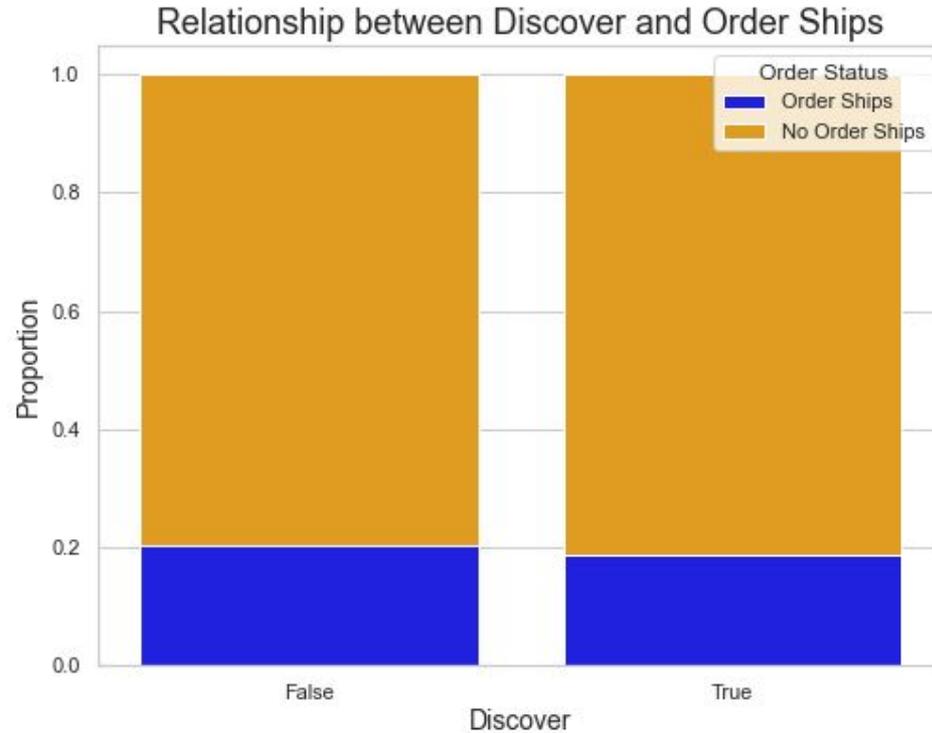


Initial Insights

Proportion of “Successful” Journeys

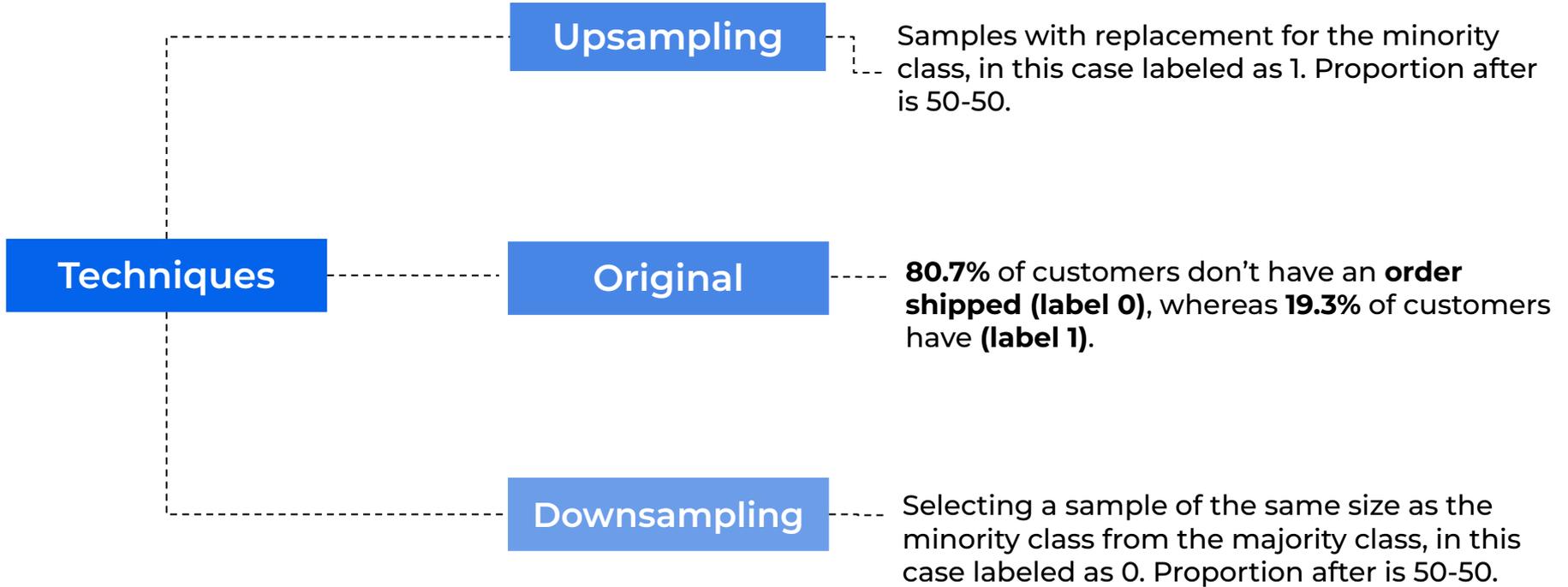


Importance of Discover Campaign

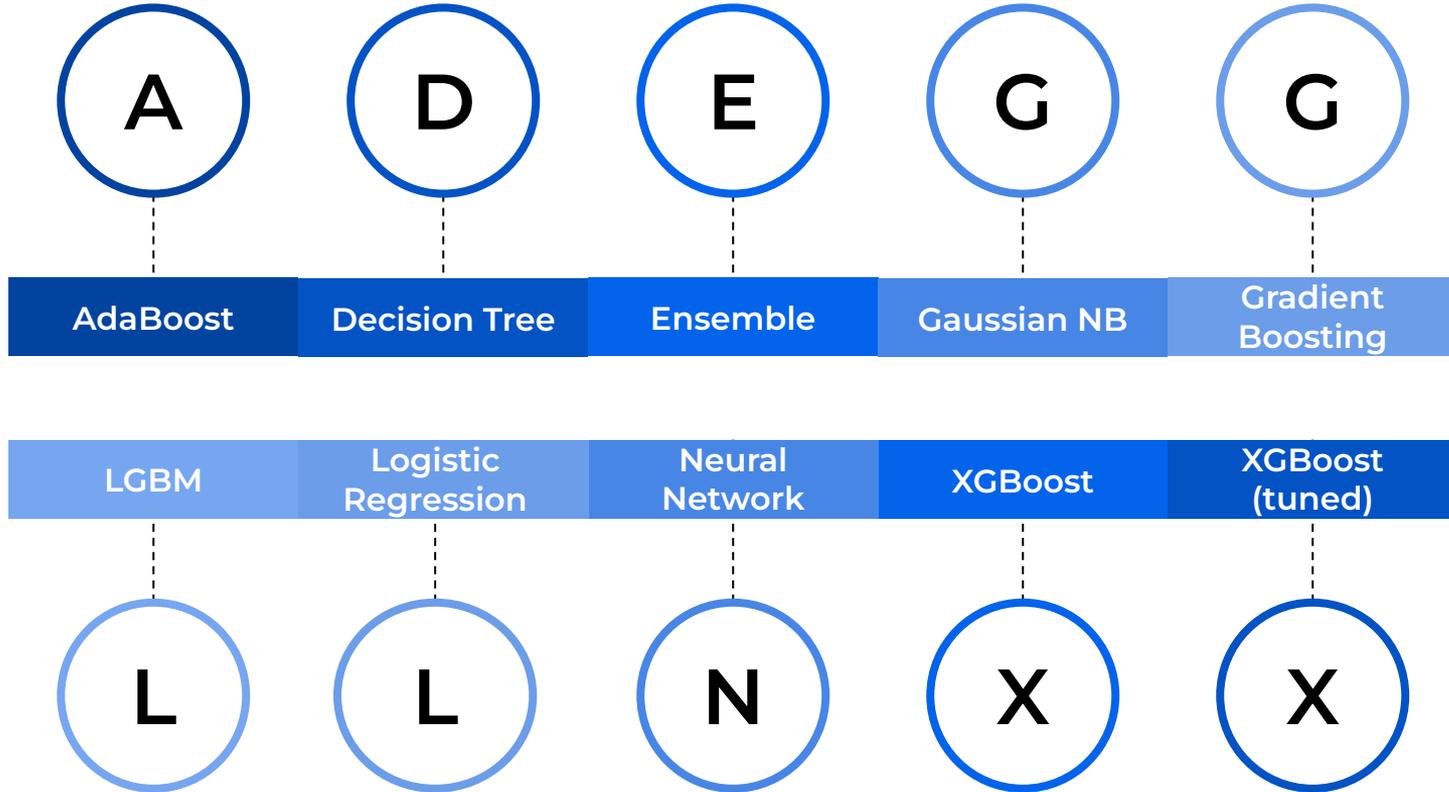


Classification Models

Approaches



Models



Advance Model Development

Bayesian Optimization (XGBoost)

We utilized Optuna, a Python library which uses Bayesian optimization algorithms to go through the search space.

In total, we tuned 9 parameters of the XGBoost model.

- Number of estimators
- Maximum depth
- Learning rate
- ... etc

Neural Network Architecture

Designed comprising five fully connected layers with ReLU activations, interspersed with dropouts to avoid overfitting and incorporating L2 regularization, culminating in a sigmoid activation function.

Additionally, prior to training data input, weights for each label were computed accordingly.

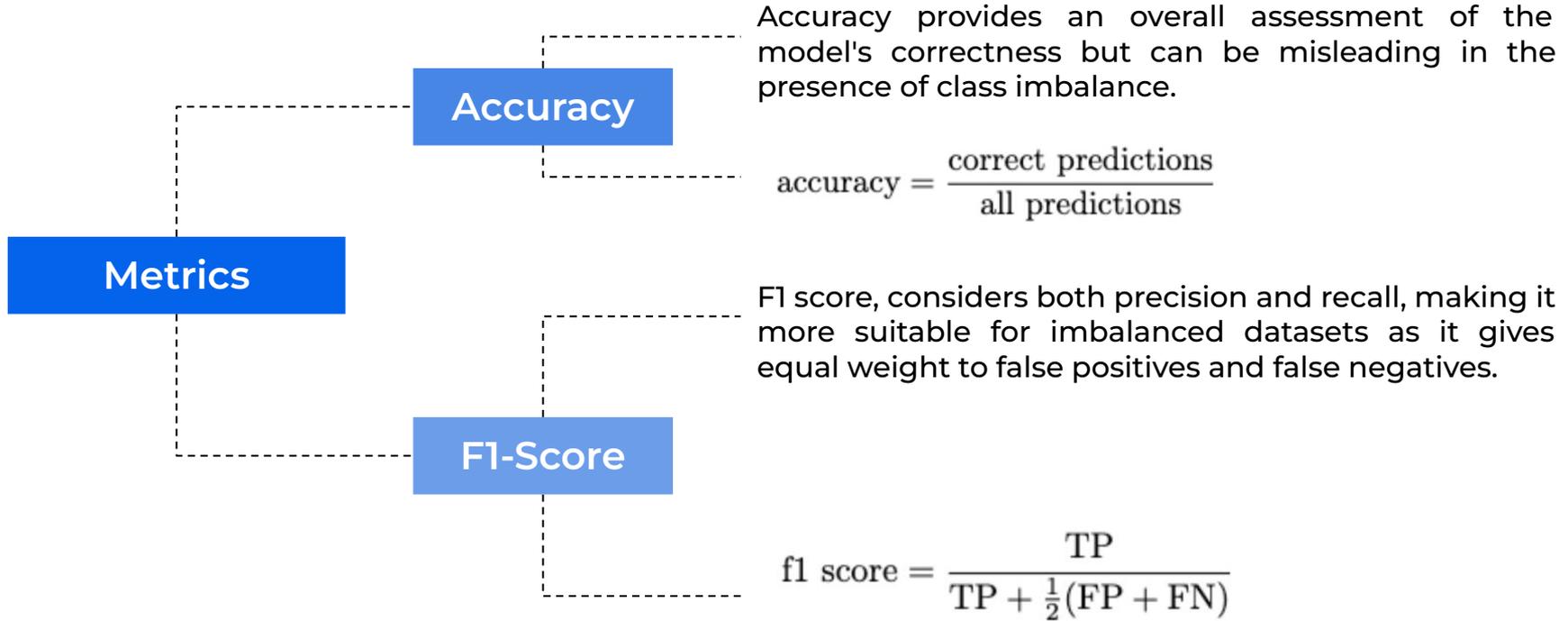
Ensemble

Our ensemble combines diverse models such as : Logistic Regression, Random Forest, Gaussian Naive Bayes, XGBoost, LightGBM, Gradient Boosting, AdaBoost, and Decision Tree classifiers.

Employing a Voting Classifier with soft voting, it aggregates predictions and enhances performance. Evaluation via classification report validates effectiveness.

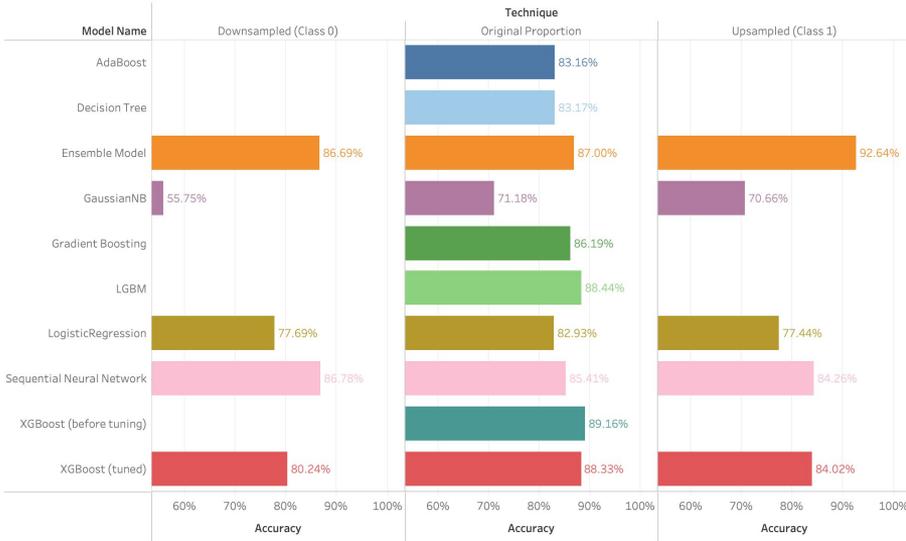
Results

Choose of the correct metric

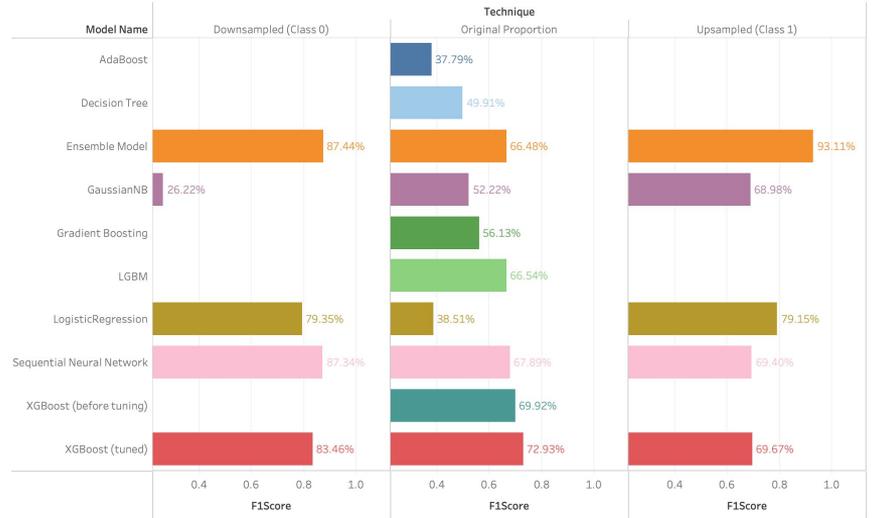


Metrics

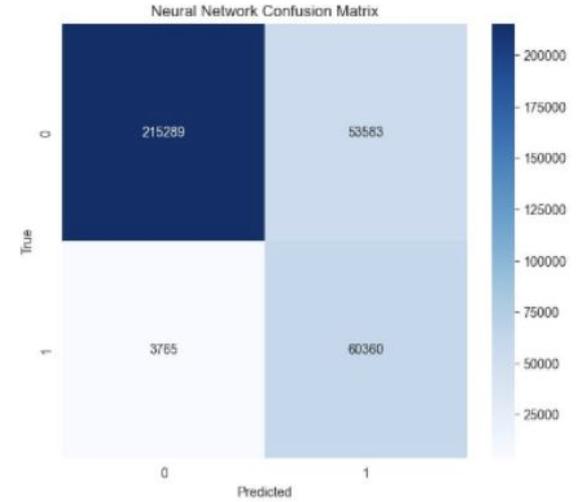
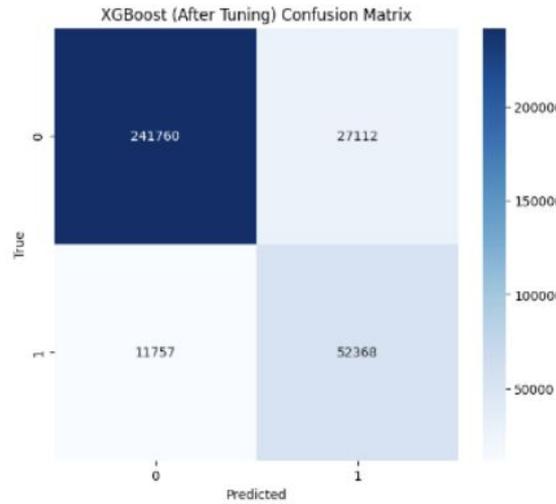
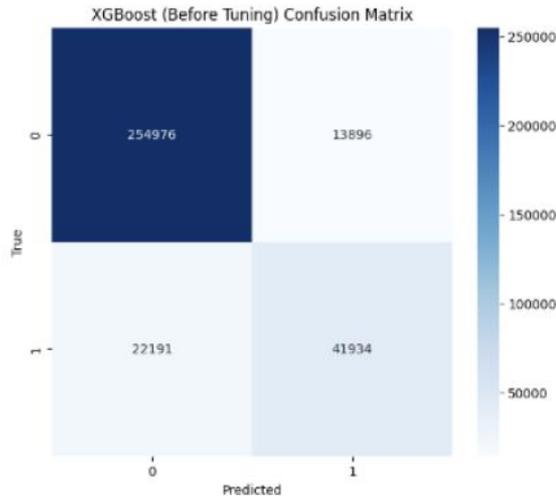
Accuracy of Models by Technique



F1 score of models by technique

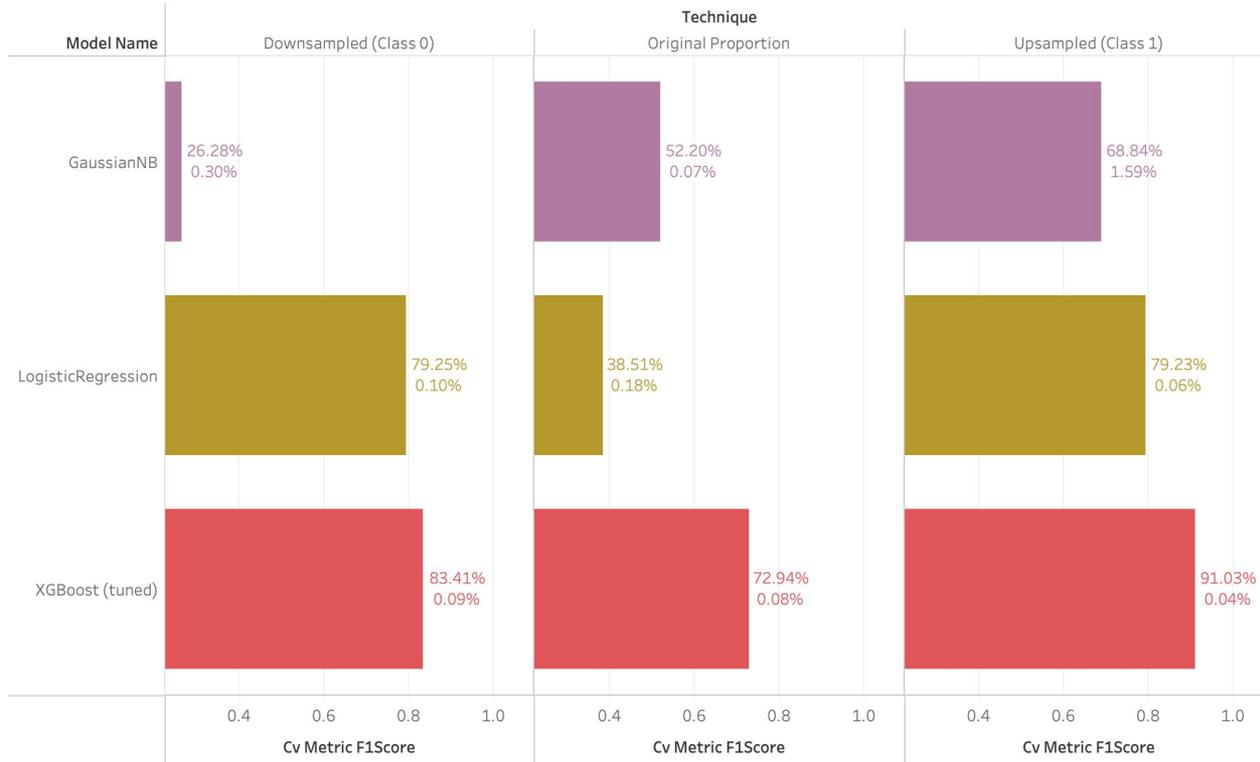


Confusion Matrices



Cross Validation with F1 Score

F1 scores (mean and std) in CV by Technique



Conclusion

Conclusion



Future Steps

- **Explore Clustering Models**
 - Potential to yield accurate results from customer segmentation
- **Fingerhut Use Cases for Model**
 - Analyze customer behavior to target customers that are predicted to quit
 - Customize website to guide customers towards purchase



Reflection

- **Tuned XGBoost Model**
 - F1 Score: 72.94%
 - Significant step towards understanding customer behavior
 - Opens the door to identify critical parts of Fingerhut's website and help them increase ordering rates

Questions