iFood Marketing Campaign Analysis

Group 27 – Final Presentation

MGT 6203 - Spring 2022

Suman Agrawal, Abhee Brahmnalkar, Christopher Dedow, Joshua DeWeese, Dinu Mathew

Georgia Tech

Dataset

- The dataset was provided by iFood, a leading food delivery app in Brazil similar to Doordash
- The dataset includes
  - Customer profile information
  - Customer product preferences
  - Previous marketing campaign successes/failure
- Summary
  - 28 columns
  - 2,240 rows (customers)
  - Key variables:
    - Demographic information such as age, income, marital status country, and Education level
    - Response to previous marketing campaigns
    - Amounts spent in various categories such as meat products, gold products etc.
    - Number of purchases by category

Georgia Tech

Objective, Problem Statement and Hypothesis

Objective:

- Our primary objective is to predict the success of a future marketing campaign (the Response variable) and be able to use that information to better target marketing to the customer base.
- Some questions we are interested in:
  - What is the relationship between customer attributes and response to product specific campaigns?
  - Does response to campaigns vary between customers that buy online vs in store?

Problem Statement:

- Can we predict the success of the last marketing campaign based on customer attributes utilizing probabilistic and supervised learning models?

Hypothesis:

- Our hypothesis is that a probabilistic machine learning model with high accuracy would allow iFood to better target campaigns to customers that are most likely to respond. This would lead to higher conversion rates and thus higher total sales.

Georgia Tech

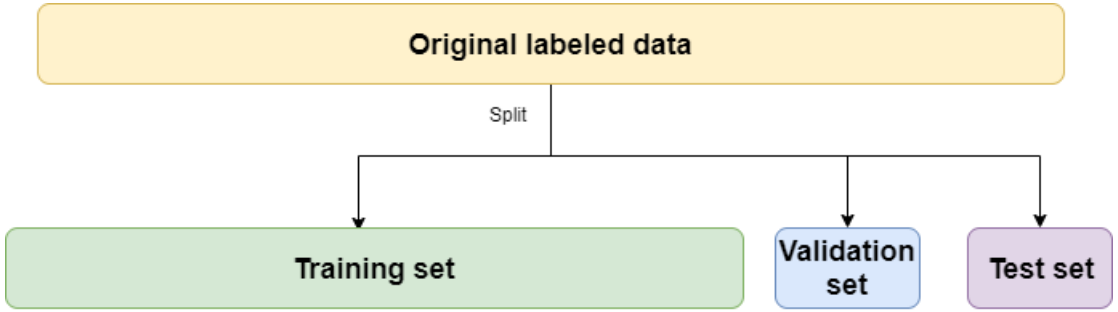
Approach and Methodology

Steps:

1. Assess model performance through following process flow
    1. Train (75% of dataset)
        1. Optimize for accuracy
        2. Test with and without cross-validation and hyperparameter tuning to attempt to reduce overfitting and ideally return higher performance
    2. Validation (12.5% of dataset)
        1. Accuracy
        2. Precision
        3. Recall
    3. Test (12.5% of dataset)
2. Identify prospective/current customers whom we think will respond to a future marketing campaign

Success Metrics:

1. Accuracy – correctly predicted responses / total responses
    1. Question answered: Out of all the responses, how many did we correctly predict?
    2. (TP + TN) / (TP + FP + FN + TN)
2. Precision – true positive values / total predicted true values
    1. Question answered: Out of all the responses we predicted would respond, how many responded?
    2. TP / (TP + FP)
3. Recall – also known as sensitivity, correctly predicted true positive responses / actual true positives
    1. Question answered: Out of all responses that responded, how many did we correctly capture?
    2. TP / (TP + FN)



Georgia Tech

Exploratory Analysis

- **Principal Component Analysis (PCA) -** Given large number of variables we wanted to see if we can reduce the dimensionality of the problem. This didn't yield significant results.

- **Multiple Regression Model -** This vanilla model gave us preliminary insights into the most significant variables of the dataset.

- **Logistic Regression -** This model was a natural choice when we wanted to see the outcome as a probability rather than a binary yes/no prediction on Response variable. Having the ability to tinker with the threshold was important to us to capture the risk tolerance of executives. This model yielded 90% accuracy, which was promising so we decided to take it a step further with stepwise regression in the subsequent analysis

Logistic Regression: Variable Impact

Georgia Tech

Baseline & Optimized Models

Baseline

- **Random Forest**
  - Given the substantial number of variables, Random Forest was the best choice between decision tree models.
- **Naïve Bayes**
  - Instead of solely relying on one classification model, we also wanted to try out another classification model which will provide us probabilities and as such we went with Naïve Bayes because it is a simple model to construct

Optimized

- **Naïve Bayes: Hyperparameter Tuning**
  - Use cv for parameter tuning from the baseline model, and added gridsearch (Caret package)
  - Accuracy was used to select the optimal parameters
- **Stepwise Regression with K-fold cross validation**
  - Used forward and backward stepwise regression to identify optimal set of variables
  - Created k-fold models for forward and backward, then predicted accuracy of each model
- **Assumptions**
  - We kept a 0.5 threshold value for predicting probability
  - In all models, we treated Response as a factor

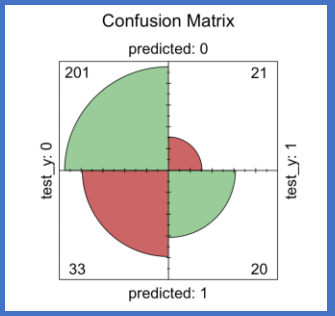
Georgia Tech

Results

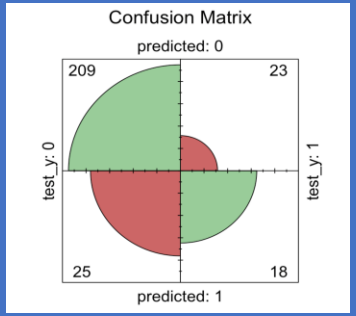
Results on Validation dataset

Model	Precision	Recall	Accuracy
Basic Naives Bayes	90.5%	85.9%	80.3%
Optimal Naive Bayes	90.0%	89.3%	82.5%
Stepwise Forward Logistic Regression	89.5%	98.3%	88.72%
Random Forest	88.93%	96.15%	86.54%

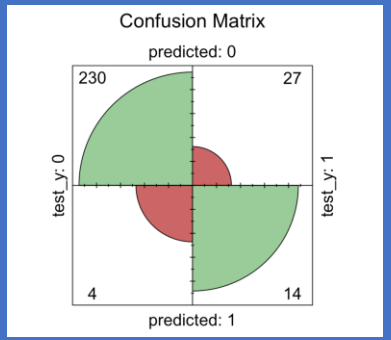
Basic Naive Bayes



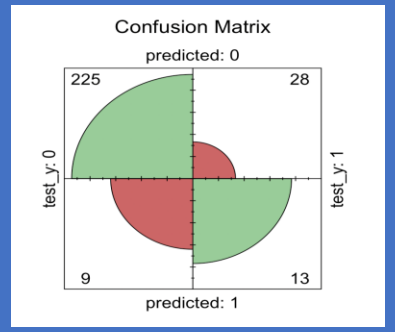
Optimal Naive Bayes



Stepwise Forward Log Reg



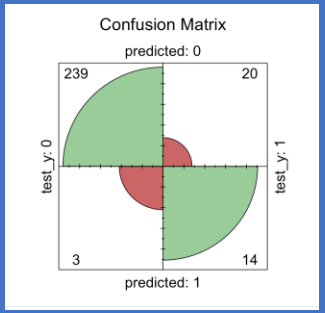
Random Forest



Results on Test dataset for the final model

Model	Precision	Recall	Accuracy
Stepwise Forward Logistic Regression	92.27%	98.76%	91.67%

Stepwise Forward Log Reg



Georgia Tech

Interpretation

- **Overall response to past campaigns** - The odds of responding to a new campaign increase by 556% with one unit increase in overall campaign acceptance in the past, holding all other predictors constant.
- **Tenure of customers** - The odds of responding to a new campaign increase by 0.44% when the customer has been with the company for one more day, holding all other predictors constant.
- **Recency of last purchase** - The odds of responding to a new campaign decreases by 2.84% with one day increase in the days since last purchase was done, holding all other predictors constant.
- **Teens at Home** - The odds of responding to a new campaign decreases by 64.76% with one additional teen, holding all other predictors constant.
- **Marital Status** - The odds of responding to a new campaign decreases by 77.86% for customer who is married compared to customer who is not married, holding all other predictors constant.
- **Customer living with a partner** - The odds of responding to a new campaign decreases by 71.49% for customer who has a partner compared to customer who does not have a partner, holding all other predictors constant.

Georgia Tech

Interpretation

- **Customer with PhD** - The odds of responding to a new campaign decreases by 174% for customer who has a PhD compared to customer who does not, holding all other predictors constant.
- **Total number of deals purchased** - The odds of responding to a new campaign increase by 27.73% with one additional unit of deal purchased, holding all other predictors constant.
- **Response to campaign# 4** - The odds of responding to a new campaign decreases by 63.86% for customer who responded to campaign# 4 compared to customer who did not, holding all other predictors constant.
- **Amount of meat products purchased** - The odds of responding to a new campaign increase by 0.24% with one additional unit of meat purchased, holding all other predictors constant.
- **Total number of purchase done in store** - The odds of responding to a new campaign decrease by 14.11% with one additional purchase done in store, holding all other predictors constant.

Georgia Tech

Conclusion

Based on our analysis:

- The company should be spending most of the campaign budget on customers who have a longer tenure, are not married and are without teenage dependents.
- Customers with advanced education should be part of the target population.
- The campaigns should be focused on offering discounts for ordering meat products through the app.
- The campaigns should be targeted at customers who do not visit a physical store frequently.
- The company does not need to create opportunity cost by creating test and control groups for customers who fall in the above segment and instead should switch offer tactics for every campaign.
- Marketing should focus on doing an A/B test for customers who are outside of the above segmentation to validate the kind of offers that will attract them towards their app and to opt for food delivery. Multiple variations of the content, design, and placement of the campaigns can be tested to see the variation in responses.

Georgia Tech