PROJECT 4- PERSONAL LOAN CAMPAIGN

DATA SCIENCE AND BUSINESS ANALYTICS - SUPERVISED LEARNING - CLASSIFICATION

KRITHIKA SRINIVASAN

BUSINESS OVERVIEW

- AllLife Bank is interested in expanding their customer base rapidly to bring in more loan business
- A campaign run last year encouraged 9% of customers to obtain personal loans
- They'd like to increase the success ratio by identifying potential customers who have a higher probability of purchasing the loan
- Objectives:
  - To predict whether a liability customer will buy a personal loan or not
  - Which variables are most significant
  - Which segment of customers should be targeted more

SOLUTION APPROACH

- This is a clear classification problem
- The approach I took was to solve the problem first with Logistics Regression and then with the Decision Tree algorithm
- The above methods of Supervised Learning seem apt for this problem as we need to predict if a customer will buy a loan or not (1 or 0)
- Prior to implementing these algorithms, in-depth EDA was carried out to identify patterns in the data

KEY FINDINGS AND INSIGHTS (EDA)

- Percentage of customers who took a personal loan = 9.6%
- There are a lot of customers with low annual income and few customers with high annual income
- 29% of the customer base is single, followed by 25% being a couple
- Most customers have a low monthly credit card expense
- 42% of the customers did not finish an undergraduate degree. 28% have an undergraduate degree and 30% have a professional degree

KEY FINDINGS AND INSIGHTS (EDA)

- 69% of the customers do not have a mortgage
- 89% of the customers do not have a securities account
- 94% of the customers do not have a CD account
- 60% of the customers have an online presence
- 30% of the customers have a credit card with another bank

KEY FINDINGS AND INSIGHTS (EDA)

- Customers who have taken the personal loan have an income of greater than $ 50K
- Fewer customers have taken the personal loan who have Education level as 1, compared to levels 2 and 3
- Customers who dont have a securities account mostly do not have a personal loan
- Most customers who have a CD account also have a personal loan
- Customers who have Education level 1 have higher income

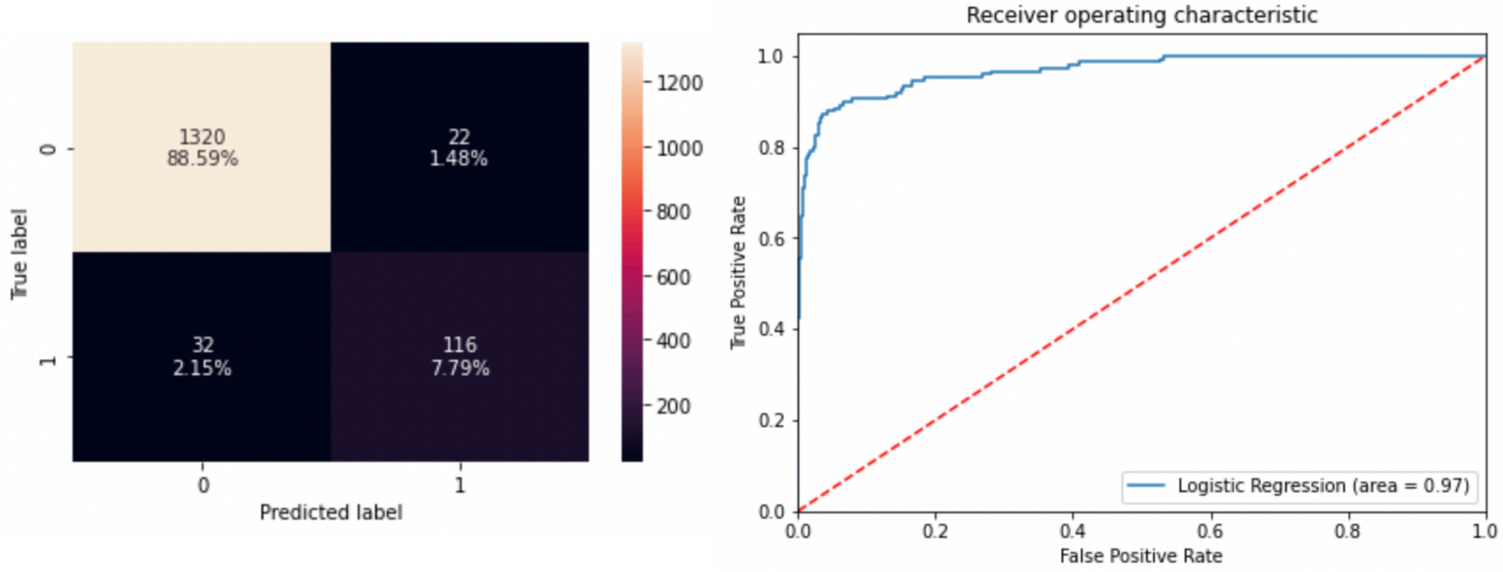
MODEL EVALUATION CRITERIA

- In terms of what classification error is worse for this problem, both false positives and negatives are bad for the bank
- Predicting a customer will buy a personal loan when they do not is bad as the bank would assume a certain profit level because of the incorrect prediction
- Predicting a customer will not buy when they end up buying might also be an unfavorable scenario for the bank as the bank now has a situation they did not account for
- Hence, the f1_score was maximized by adding/dropping features

MODEL OVERVIEW

Logistics Regression Model

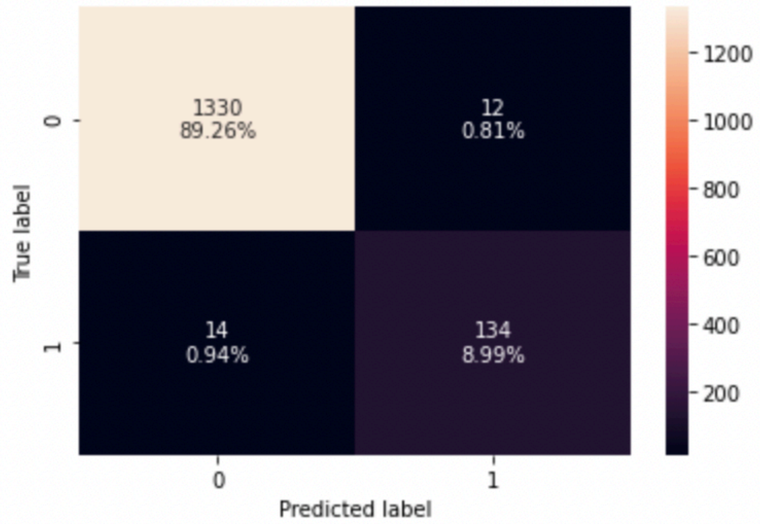
- Optimal threshold of 0.42
- f1_score on training set = 0.808
- f1_score on test data = 0.811

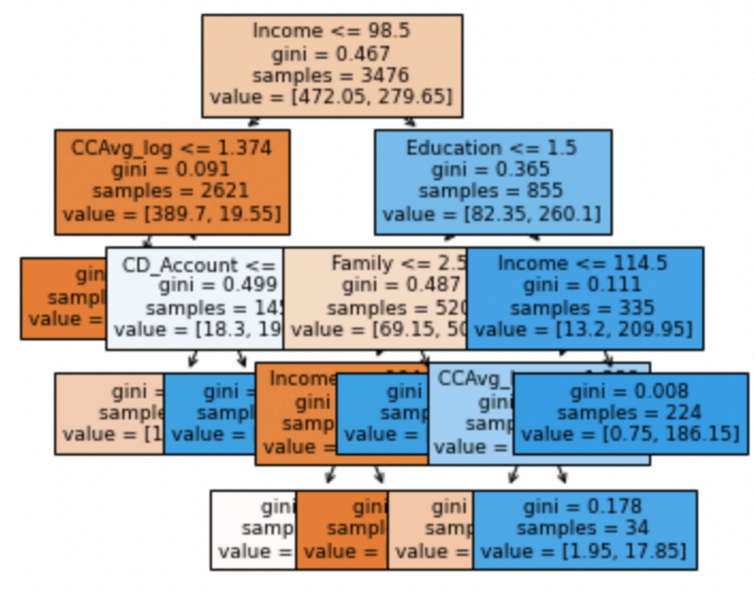


MODEL OVERVIEW

Decision Tree Model

- Cost complexity pruning was done to further maximize the f1_score
- f1_score on training set = 0.910
- f1_score on test data = 0.911





PERFORMANCE SUMMARY

- The Decision Tree model performs better than the Logistics Regression model
- The Decision Tree model gives more intuitive information on what customers should be targeted for the personal loan campaign
- However, both models have a high f1_score, which suggests that the error due to misclassification is minimized
- Both models can be used for this classification problem

BUSINESS RECOMMENDATIONS

- The bank should target customers who have:
  - Higher income
  - Not completed an undergraduate degree
  - A family of size greater than 2
- If they have completed an undergraduate degree, the bank should target those customers who have a higher average credit card spending per month