Personal Loan Campaign

by: Garey Salinas

greatlearning
Power Ahead

Contents

Business Problem

The number of customers who are also borrowers (asset customers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process, earn more through the interest on loans. In particular, the management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors).

Objective

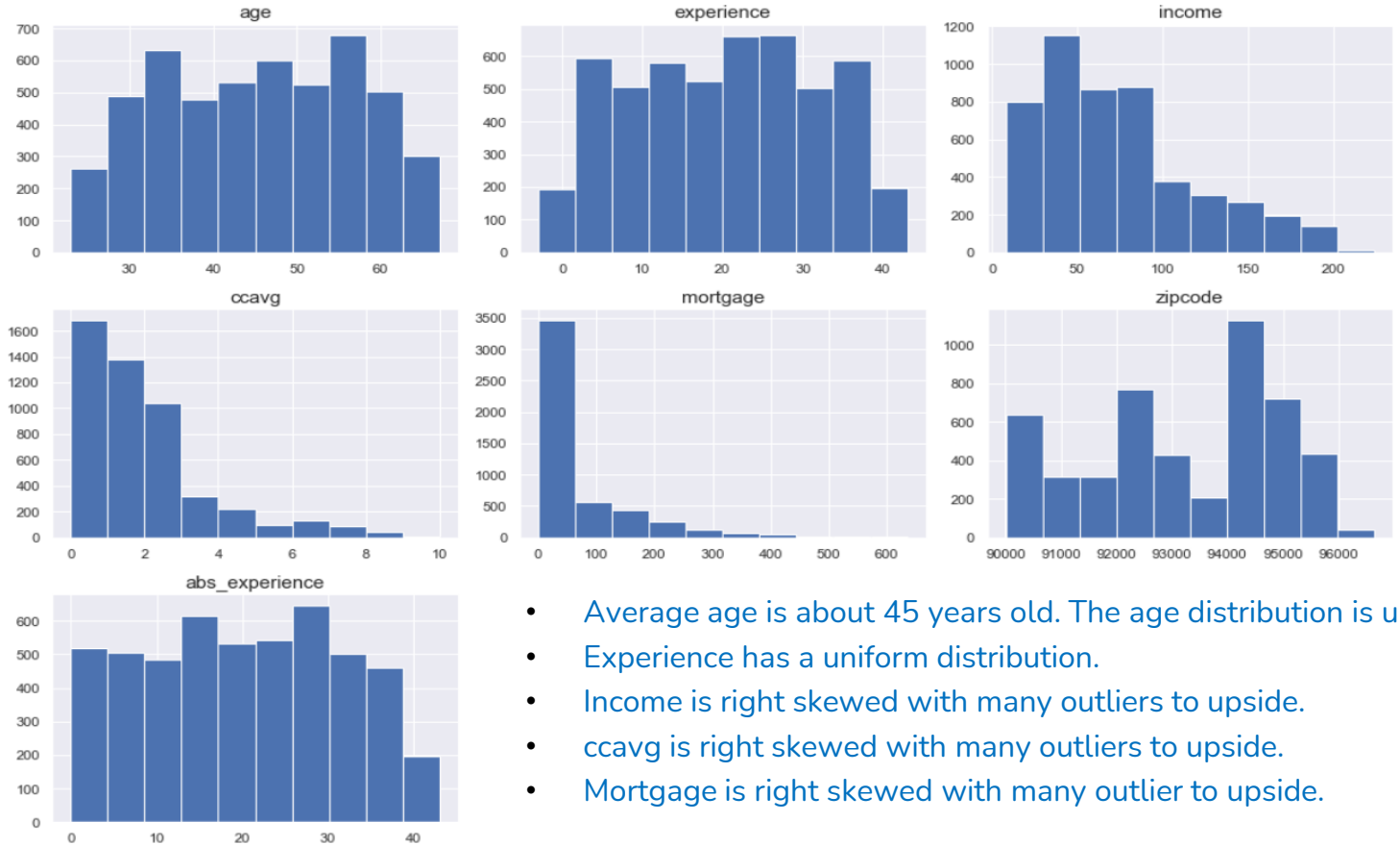
1. Predict whether a liability customer will buy a personal loan.
2. Which variables are most significant.
3. Which segment of customers should be targeted more.

Data Overview

Labels	Description
Id	Customer ID
age	Customer's age in completed years
experience	#years of professional experience
income	Annual income of the customer (in thousand dollars)
zipcode	Home Address ZIP code.
family	the Family size of the customer
ccavg	Average spending on credit cards per month (in thousand dollars)
education	Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional
mortgage	Value of house mortgage if any. (in thousand dollars)
personal_loan	Did this customer accept the personal loan offered in the last campaign?
securities_account	Does the customer have securities account with the bank?
cd_account	Does the customer have a certificate of deposit (CD) account with the bank?
online	Do customers use internet banking facilities?
credit_card	Does the customer use a credit card issued by any other Bank (excluding All life Bank)?

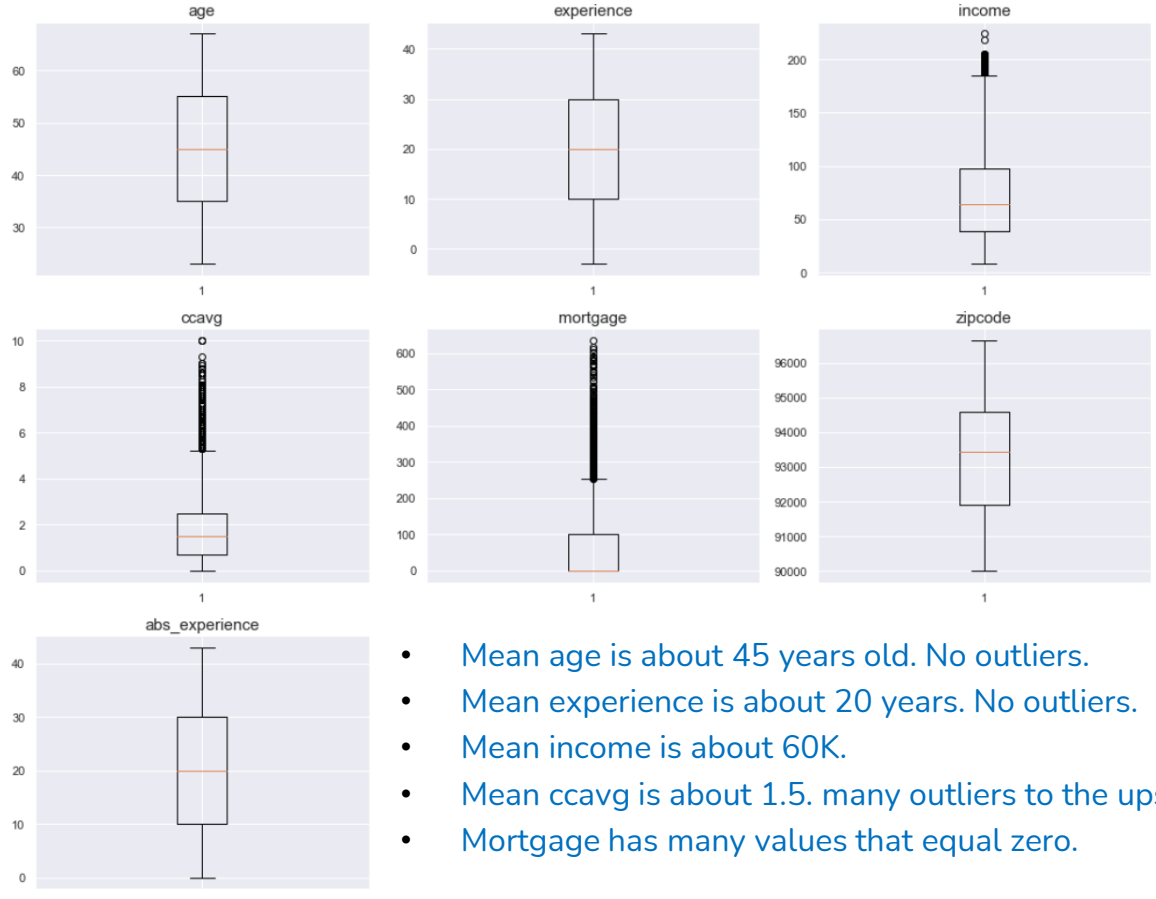
Observations	Features
5000	14

EDA – Univariate Analysis



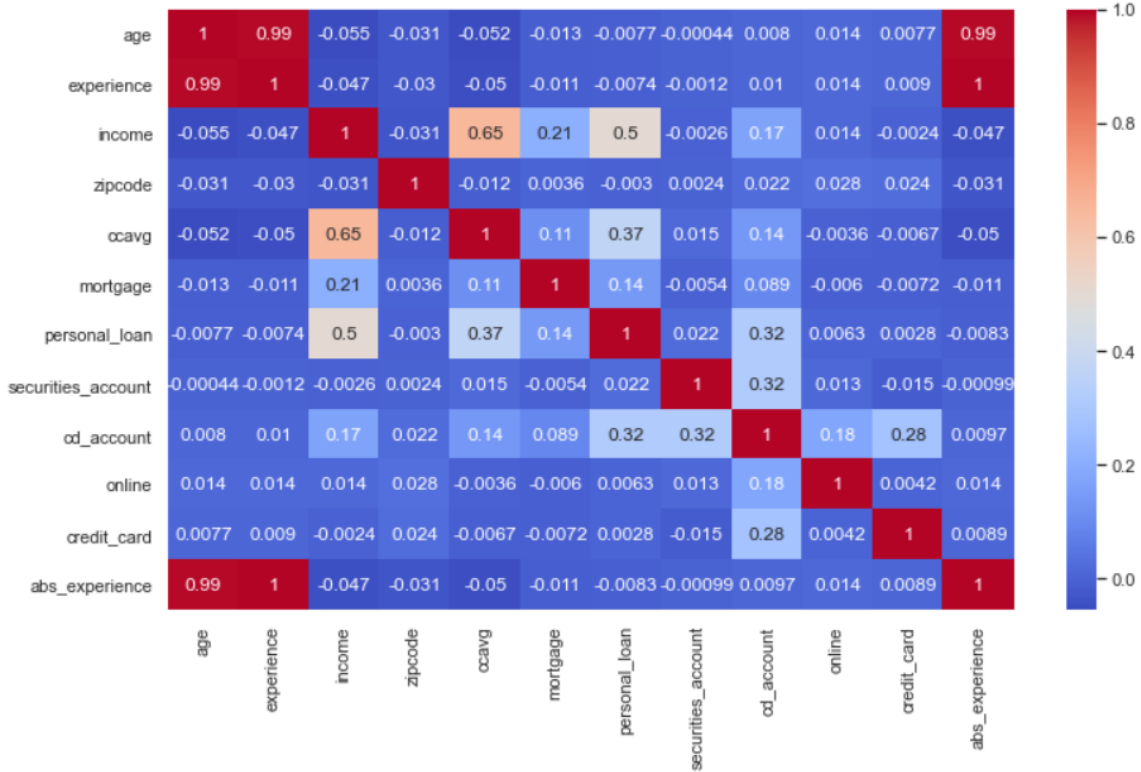
- Average age is about 45 years old. The age distribution is uniform.
- Experience has a uniform distribution.
- Income is right skewed with many outliers to upside.
- ccavg is right skewed with many outliers to upside.
- Mortgage is right skewed with many outlier to upside.

EDA – Univariate Analysis



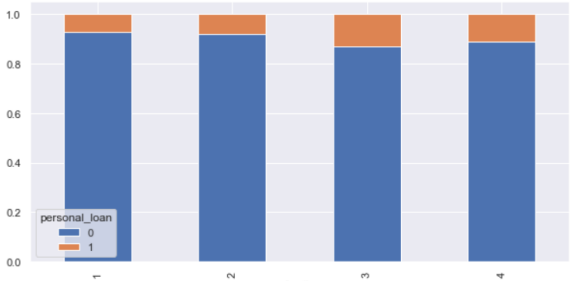
- Mean age is about 45 years old. No outliers.
- Mean experience is about 20 years. No outliers.
- Mean income is about 60K.
- Mean ccavg is about 1.5. many outliers to the upside.
- Mortgage has many values that equal zero.

EDA – Bivariate Analysis

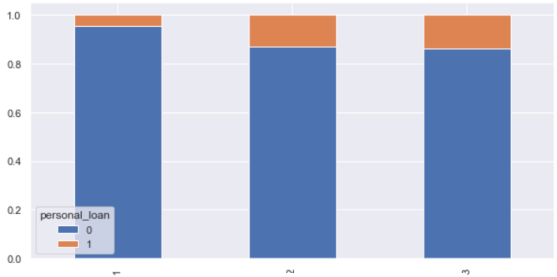


- Age and experience are heavily correlated with each other.
- Credit card usage and income are heavily correlated with each other.

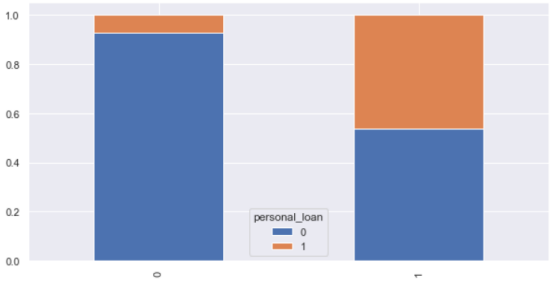
EDA – Bivariate Analysis



personal_loan vs family



personal_loan vs education



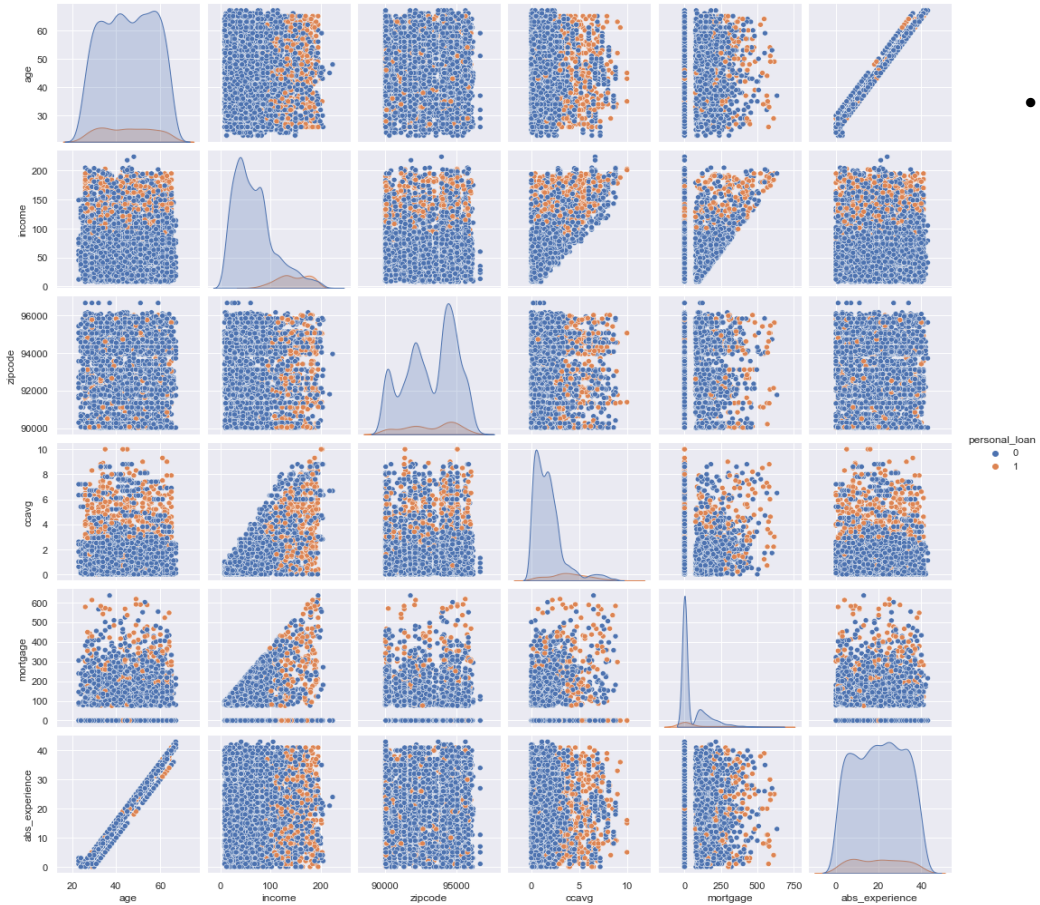
personal_loan vs cd_account

- Those customers with a family of 3 have a greater percentage of personal loans than the other family categories.
- A family of 4 has the second greatest percentage of personal loans followed by a family of 2, and 1.

- Those customers with an education of '2' and '3' hold a greater percentage of personal loans that those customer with an education of '1'.

- Those customers with cd accounts have a greater percentage of personal loans than those customer without a cd account.

EDA – Bivariate Analysis

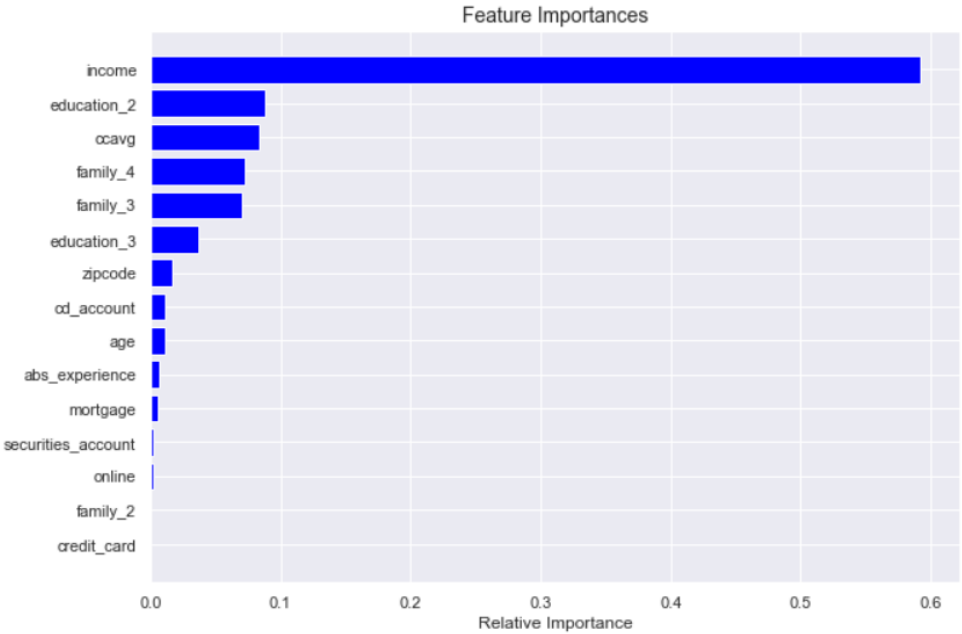


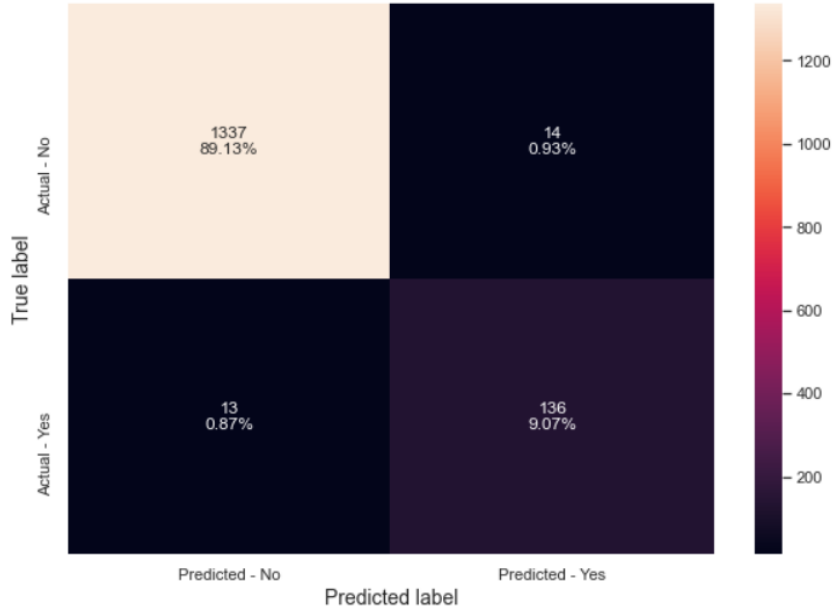
- The plot shows income is higher among those customers with personal loans.

EDA – Bivariate Analysis

- On average, those customers with higher incomes have personal loans.
- On average, those customers with higher credit card usage have personal loans.

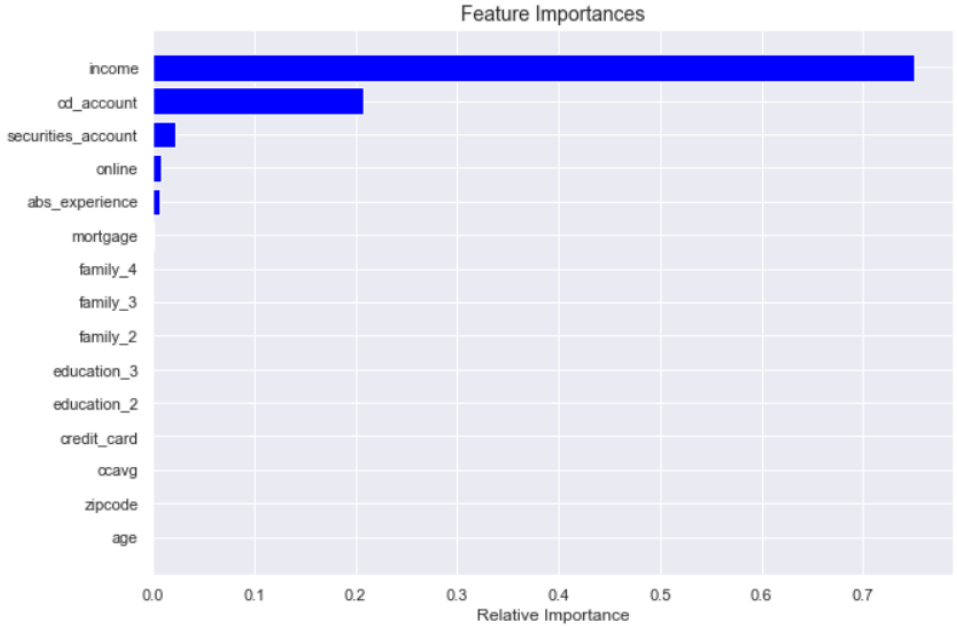
Model 1: Initial

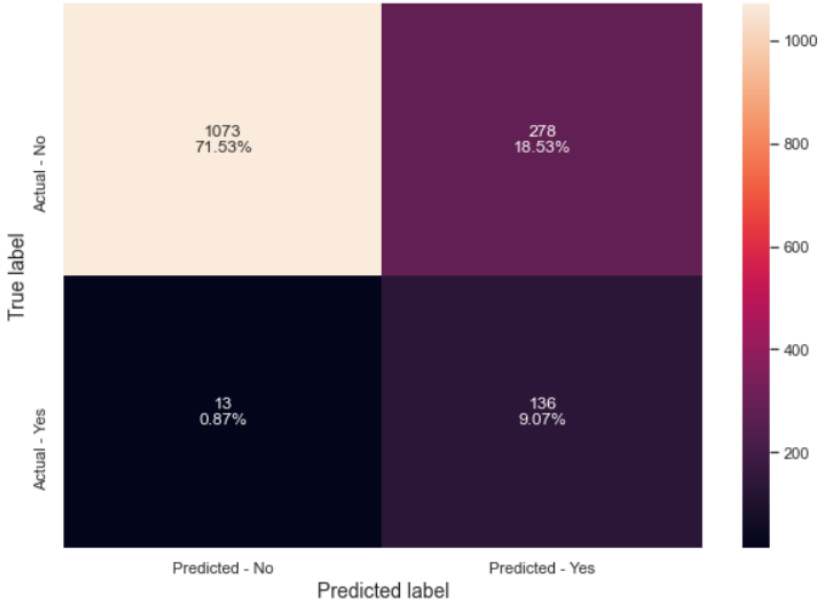




- We will build our model using the DecisionTreeClassifier
- We only have ~10% of positive classes, so if our model marks each sample as negative, then also we'll get 90% accuracy, hence accuracy is not a good metric to evaluate.
- Recall on training set is 1 and recall on test set is .91. Model is too complex – we are overfitting.

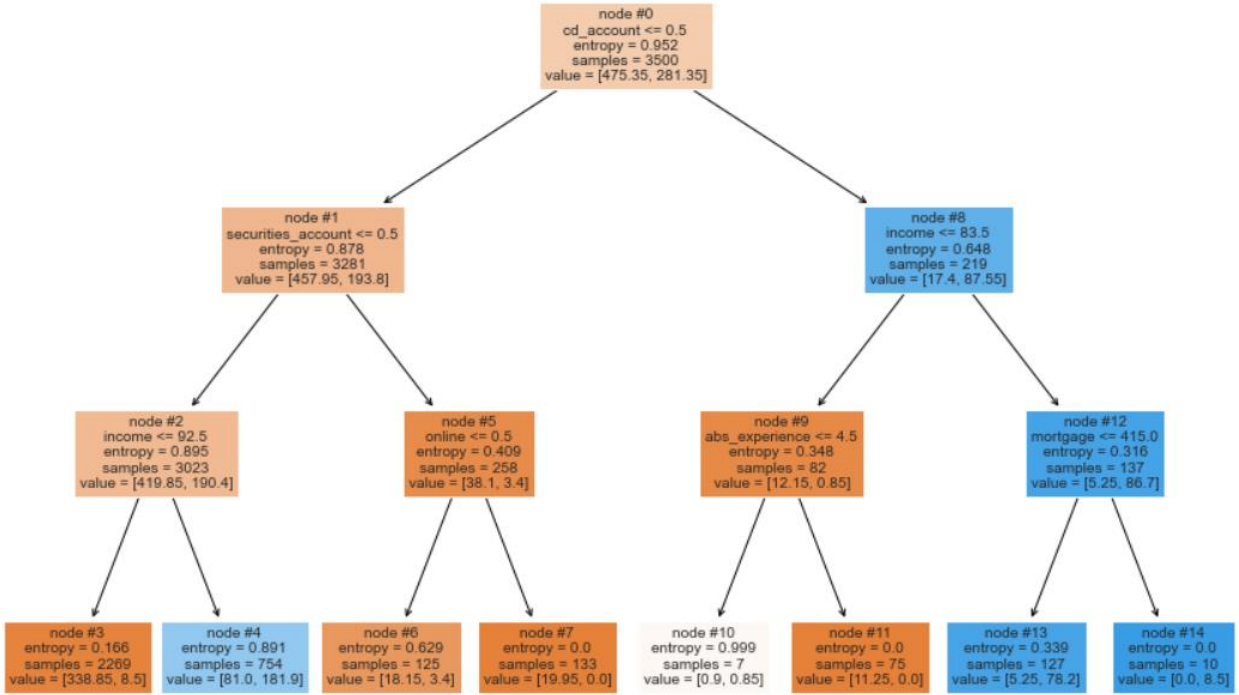
Model 2: Hyperparameter Tuning





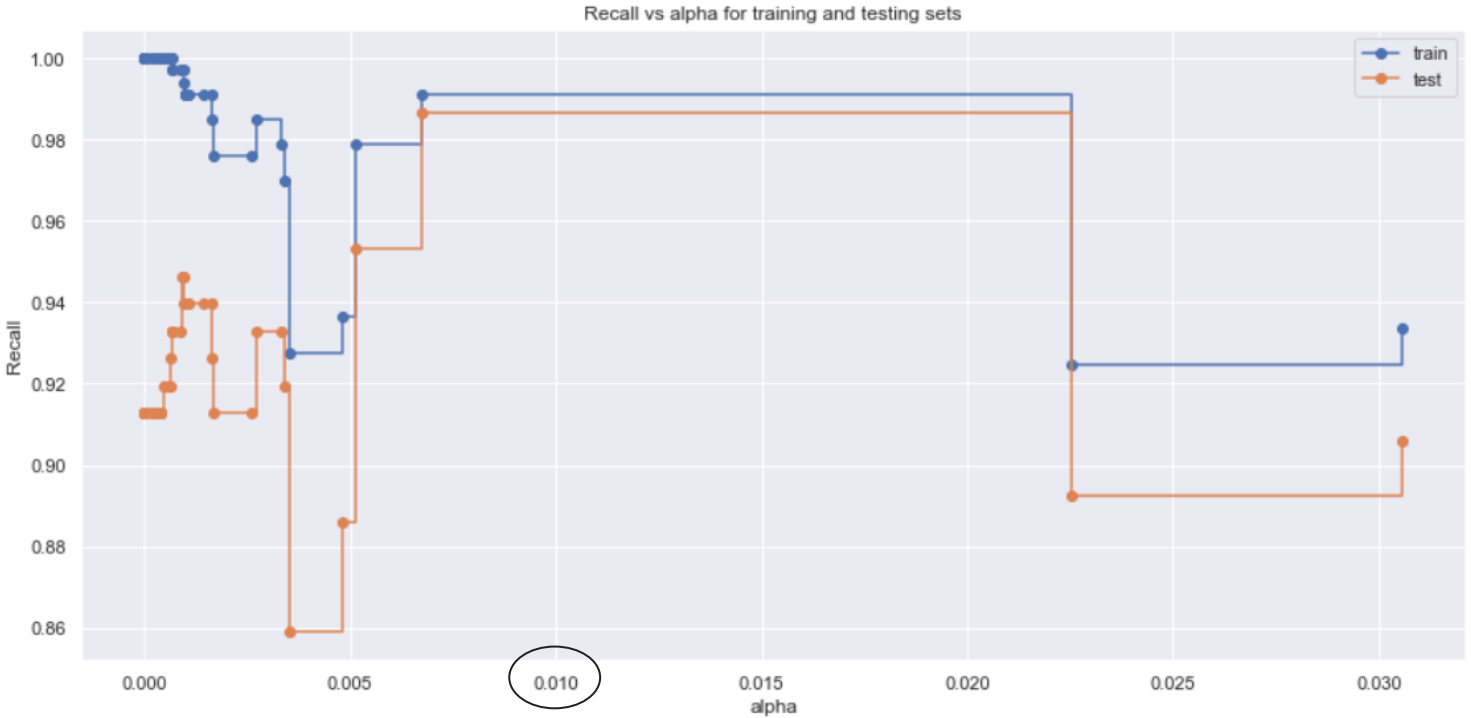
- We used GridSearchCV for hyperparameter tuning of our tree model.
- Income shows to be the most important feature in our hyperparameter tuning model.
- The recall scores for the training and test set is .955 and .913.

Model 2: Hyperparameter Tuning



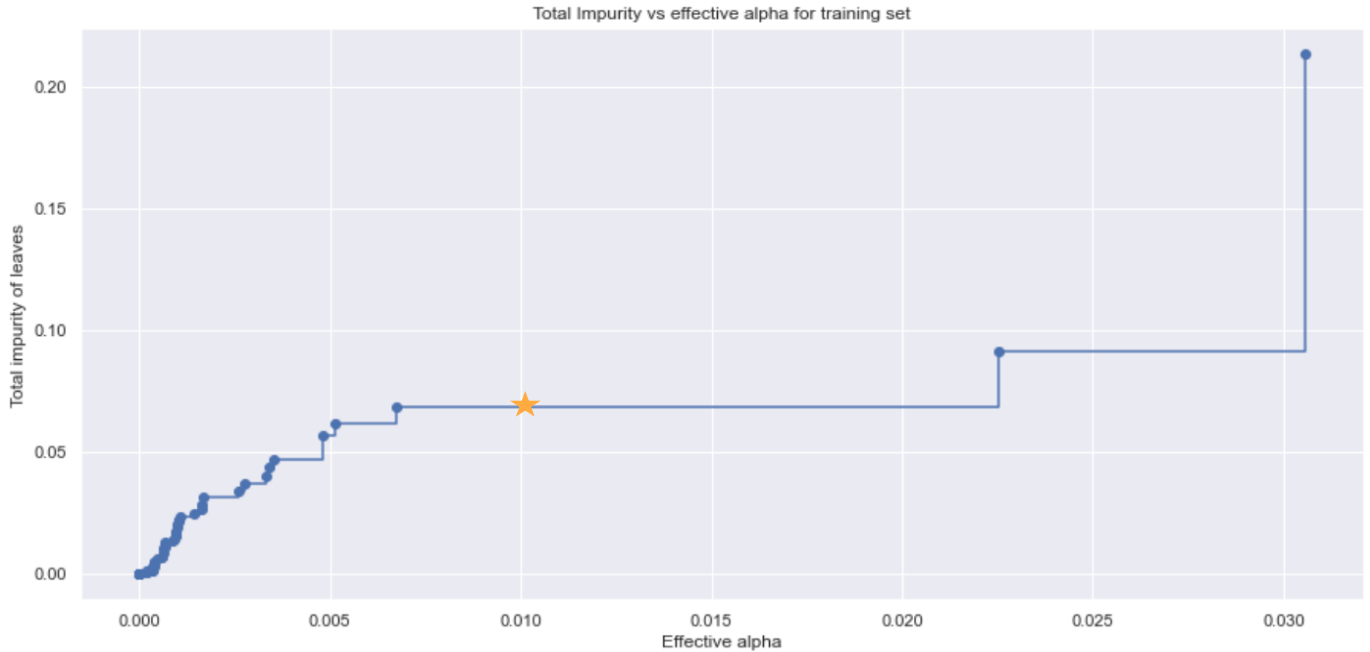
- Tree model is easy to interpret and doesn't overfit.

Model 3: Cost Complexity Pruning



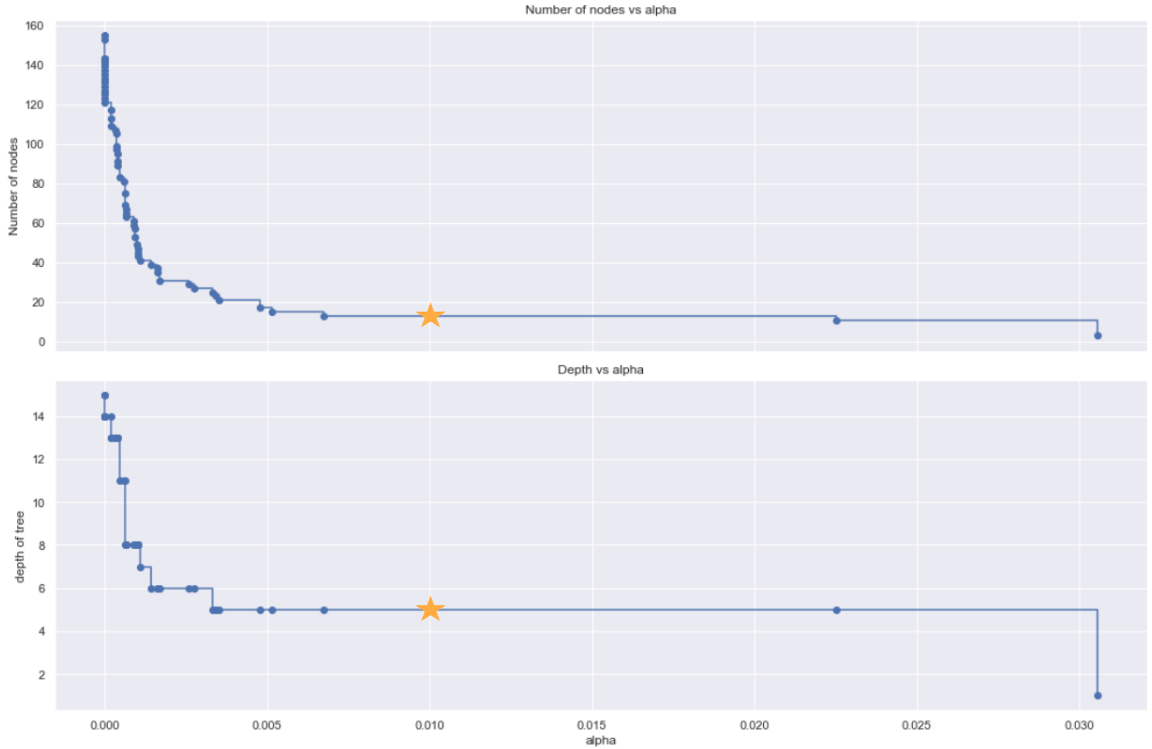
- When alpha is between ~.006 and ~.0225, the recall train and test scores are the best. For our model we choose alpha to be 0.01. We will confirm our recall train and test scores are close to .99.

Model 3: Cost Complexity Pruning



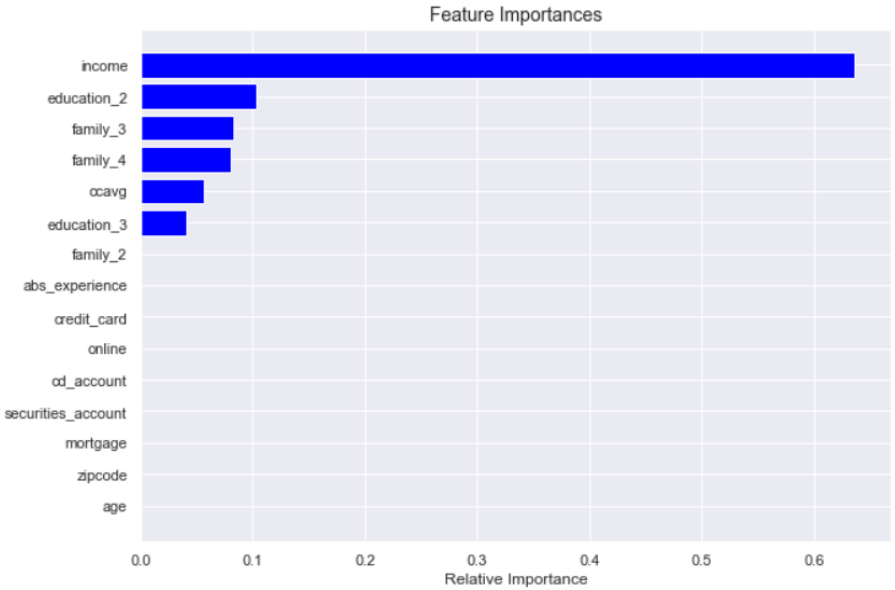
- Nodes with the effective alpha are pruned first. With an effective alpha of 0.01, our total impurity will be about .065.

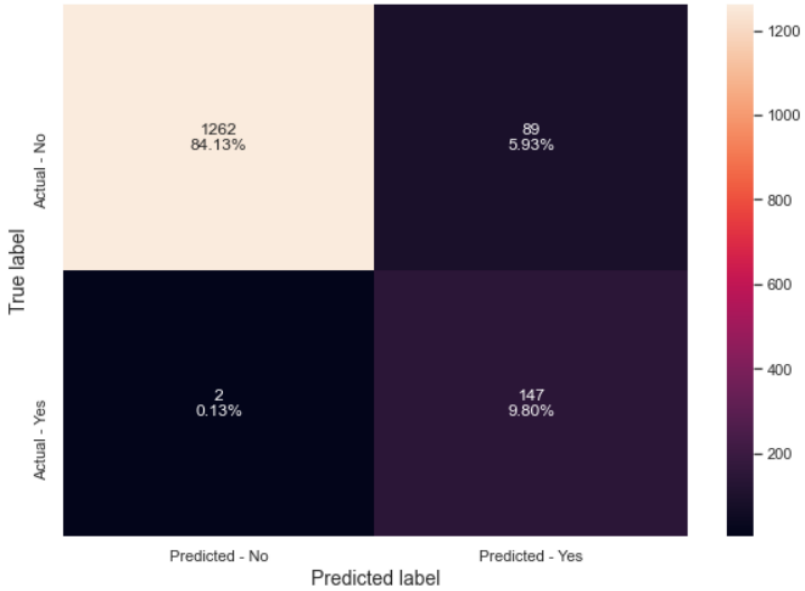
Model 3: Cost Complexity Pruning



- As alpha increases more nodes are pruned, and the depth of tree could decrease. With an effective alpha of 0.01, the depth of our tree will be 5, and total nodes close to 15. Our model is easy to interpret.

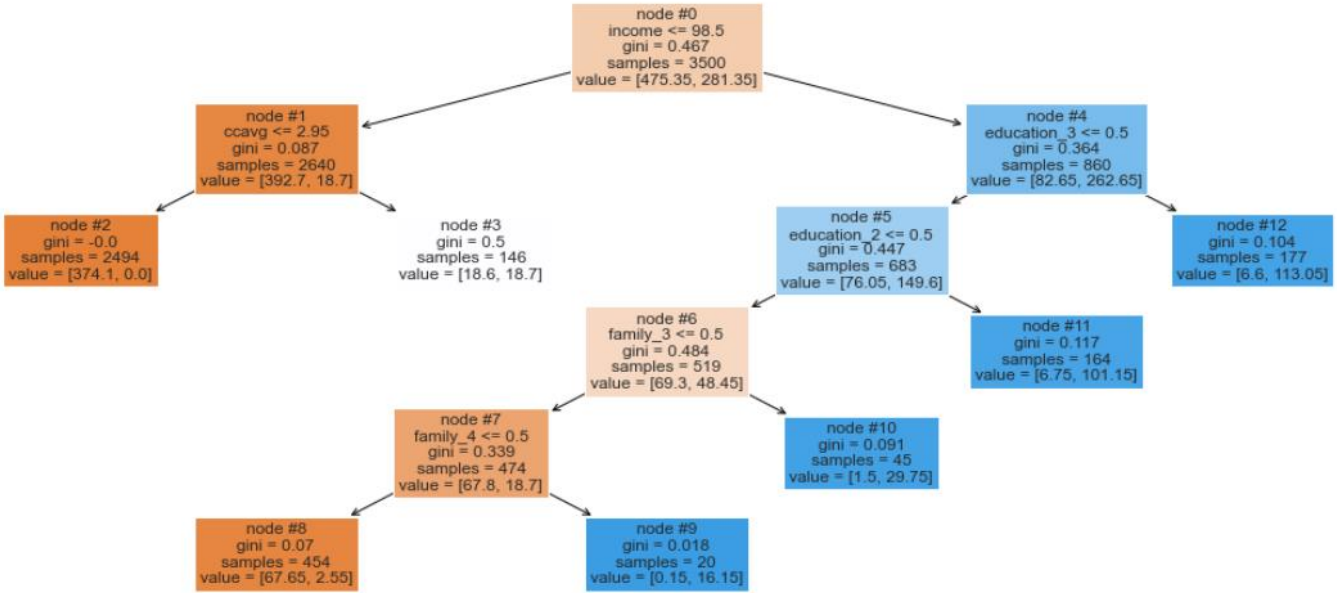
Model 3: Cost Complexity Pruning





- We built our model using the DecisionTreeClassifier and set our ccp_alpha parameter to 0.01.
- Important features are income, education, family, and ccavg.
- The recall for the training set was .99 and the test set was .98
- The model is easy to interpret and doesn't overfit the training set.

Model 3: Cost Complexity Tuning



- Tree model is easy to interpret.

Model Performance Summary

Model	Train Recall	Test Recall
Initial	1.00	.91
Hyperparameter tuning	0.95	0.91
Cost complexity tuning	0.99	0.98

Business Insights and Recommendations

- Income, education, family, and credit card usage are the most important features in predicting potential loan customers.
- From the decision tree model, income is the most important feature. If our customer's yearly income is less than 98.5K, there is a good chance the customer won't have a personal loan.
- From the model, those customers with an income greater than 98.5 and with an education level greater than or equal to 3 (Advanced/Professional) were most likely to have a personal loan. Recommend to target customers that have incomes lower than 98K.
- It was observed that those customers who use the online facilities were more likely to have personal loans. Make the site more user-friendly and encourage those customers who don't use the facilities to use the online facilities. Make the application process to get personal loans easy with a better user experience. Those customers with separate securities and cd accounts are more likely to get a personal loan. Customers who use the bank's online facilities are more likely to get a personal loan versus those customers who don't use the online facilities.