Project-Bank Loan

Campaign for selling personal loans.

This case is about a bank (Thera Bank) which has a growing customer base. Majority of these customers are liability customers (depositors) with varying size of deposits. 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). A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio with minimal budget.

The department wants to build a model that will help them identify the potential customers who have higher probability of purchasing the loan. This will increase the success ratio while at the same time reduce the cost of the campaign.

The file Bank.xls contains data on 5000 customers. The data include customer demographic information (age, income, etc.), the customer's relationship with the bank (mortgage, securities account, etc.), and the customer response to the last personal loan campaign (Personal Loan). Among these 5000 customers, only 480 (= 9.6%) accepted the personal loan that was offered to them in the earlier campaign.

**Column descriptions**

  ID : Customer ID

 Age: Customer's age in completed years

 Experience:  #years of professional experience

 Income:  Annual income of the customer ($000)

 ZIPCode: Home Address ZIP code.

 Family: Family size of the customer

 CCAvg: Avg. spending on credit cards per month ($000)

 Education: Education Level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional

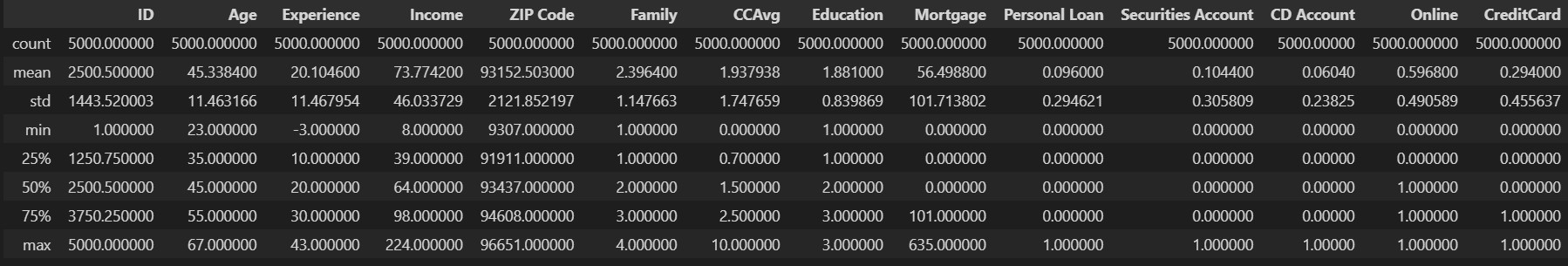
 Mortgage:  Value of house mortgage if any. ($000)

**Business**

* to identify the leads who can be converted into existing depositor into personal loan
* which will classify the account holder have more potential customers to avail personal loan
* our dateset contains 14 attributes and 5000 records with 9% success rate

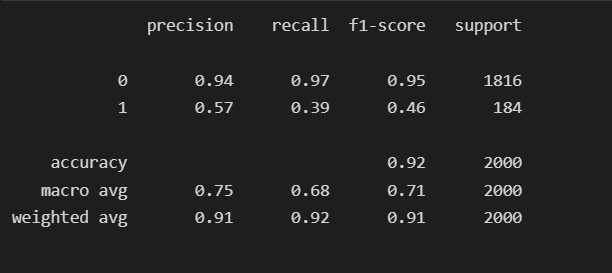
**Approach**

* there are no missing values in the dateset
* we have outliers for the attributes such as Income, CCAvg, Mortgage
* we have skewness for the following attributes as



Data description

**logistic regression**

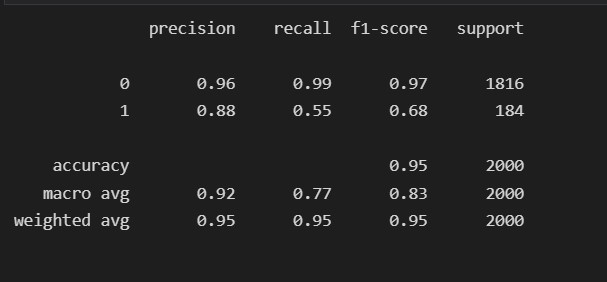
* we have implemented logistic regression with
* the accuracy with 92%
* the recall with 39% where we need to focus on recall(model is predicting) to boost 

logistic regression

**KNN**

we have implemented KNN

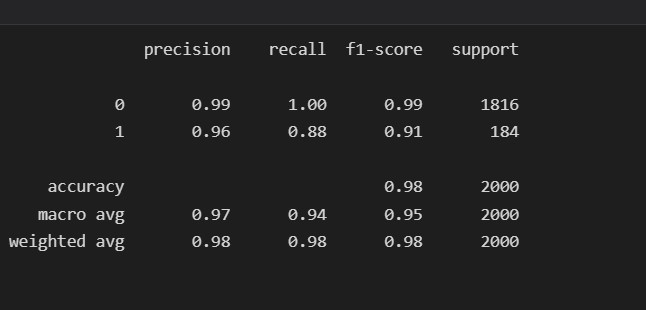
* we got recall with 8% compare with the base model it is decreasing
* majorly we have to focus on recall
* as the dataset has outliers and skewed
* we have done scaling down
* in which recall we got 55%, where the accuracy is comparable and recall is boosted to 55% where the false negative is reduced with 23% compare to base model logistic model
* we have compared with the train and test accuracy (over fit or under fit)



KNN

**Decision Tree**

* where we have implemented decision tree (where we don't any prepossessing like no scaling)
* initially we fit the decision tree, where the model is over-fitted
* so we have gone through hyper-parameter tuning some of the parameter like maximum\_depth, minimum\_sample\_depth, minimum\_sample\_leafs
* after tuning the parameters, the optimal values which we got in this tuning as been implemented in the decision tree
* we got the result, we boosted the recall value to 88% and accuracy are comparable, compare with base model we reduced 55% of errors

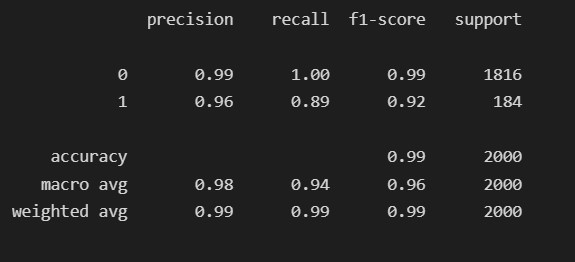


decision tree

**random forest**

next we have implemented final model as random forest with the best hyper-parameters we fit a model where are accuracy is comparable with 98%

finally we got the recall of 89% compare with base model 56% of false negative are reduced (where the random forest is good for prediction)



random forest

**Business Logic**

Result- so, the best model is the random forest to predict