Course 2: Python Project 05

Week 05: Practice Project

Bank Personal Loan Modelling:- Identifying the Potential Customers for Loans.

Domain:

Banking and Marketing.

Context:

This case is about a bank (Thera Bank) which has a growing customer base. Majority of these customers are liability customers (depositors) with varying sizes 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.

Objective:

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

Dataset Description:

The dataset 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.

Attribute Information:

Slno.	Attribute	Description
1.	ID	Customer ID
2.	Age	Customer's age in completed years.
3.	Experience	Years of professional experience.
4.	Income	Annual income of the customer($000).
5.	ZIPCode	Home address zip code.
6.	Family	Family size of the customer.
7.	CCAvg.	Avg. spending on credit cards per month($000).
8.	Education	Education level. 1: Undergrad, 2: Graduate, 3: Advanced/Professional
9.	Mortgage	Value of house mortgage if any ($000)
10.	Securities account	Does the customer have any securities account with the bank?
11.	CD account	Does the customer have a certificate of deposit (CD) account with the bank?
12.	Online	Does the customer use internet banking facilities?
13.	Credit Card	Does the customer use a credit card issued by Universal bank?
14.	Personal loan (Target Variable)	Did this customer accept the personal loan offered in the last campaign?

Questions:

1. Import the required libraries and read the dataset.

2. Check the first few samples, shape, info of the data and try to familiarize yourself with different features.

3. Check if there are any duplicate records present in the dataset? If yes, drop them. and Drop the columns which you feel are redundant.

4. Display the Five Point Summary and write your key findings.

5. There are negative values in the variable 'Experience'. Convert them to non-negative values. (Hint: .abs function)

6. Get the target column distribution and comment on the class distribution.

7. Store the target column (i.e.Personal Loan) in the y variable and the rest of the columns in the X variable.

8. Split the dataset into two parts (i.e. 70% train and 30% test). and standardize the columns using the z-score scaling approach.

9. Train and test a Logistic Regression model to predict the likelihood of a liability customer buying personal loans. Display the train and test accuracy scores.

10. Print the confusion matrix and classification report for the model and write your conclusions on the results.