Personal Loan acquisition using Naive Bayes Classifier based Learning Model
Project Duration : 21-Jan-2024 – 10-Feb-2024
Submission Information : (via) CSE-Moodle

Objective:

There is a bank with an expanding customer base. The vast majority of these customers are liability customers (depositors) with varied deposit amounts. The number of clients who are also borrowers (asset customers) is fairly modest, and the bank is keen on swiftly increasing this base in order to bring in more loan business and, as a result, generate more money through loan interest. The management wants to look into ways to convert its liability customers to personal loan customers (while keeping them as depositors).

Last year, the bank executed a campaign for liability customers that resulted in a solid conversion rate of more than 9% success. This has prompted the retail marketing department to develop campaigns to improve target marketing in order to raise success ratios with a limited budget.

The department wishes to create a model that will assist them in identifying possible clients who are more likely to purchase the loan. This will boost the success ratio while decreasing the campaign's cost.

Training Data Description: The **attribute** Information is given as follows.

The dataset contains various attributes of bank customers and whether they accepted a personal loan offered during the bank's previous campaign. It includes the following features:
- ID: Unique identifier for each customer.
- Age: Customer's age in completed years.
- Experience: Number of years of professional experience.
- Income: Annual income of the customer in thousands of dollars ($000).
- ZIP Code: Home address ZIP code of the customer.
- Family: Family size of the customer.
- CCAvg: Average monthly spending on credit cards in thousands of dollars ($000).
- Education: Level of education classified as 1: Undergrad; 2: Graduate; 3: Advanced/Professional.
- Mortgage: Value of house mortgage if any, in thousands of dollars ($000).
- Personal Loan: Binary attribute indicating whether the customer accepted the personal loan offered during the last campaign (1: Accepted, 0: Not Accepted).
- Securities Account: Indicates if the customer has a securities account with the bank (1: Yes, 0: No).
- CD Account: Indicates if the customer has a certificate of deposit (CD) account with the bank (1: Yes, 0: No).
- Online: Indicates if the customer uses internet banking facilities (1: Yes, 0: No).

- CreditCard: Indicates if the customer uses a credit card issued by the bank (1: Yes, 0: No).

1. Based on the dataset (described earlier), you will write a program to learn a **Gaussian Naive Bayes Classifier.**
2. Compare the results with the results generated by the Naive Bayes classifier learning algorithm from a pre-created package such as scikit-learn.

Note: The program can be written in C / C++ / Java / Python programming language from scratch. No machine learning /data science /statistics package / library should be used.

DataSets:
Filename: loan.csv

Tasks to be done:

1. In the dataset given the features are continuous,hence use Gaussian Naive Bayes for solving this problem.
2. Randomly split the dataset as 80/20 split i.e., 80% for training and 20% for testing.
3. Naive Bayes Classifier Model:
   a. Implement naive bayes algorithm in your code and mention the same in the report. Do NOT use scikit-learn for this part.
   b. Test the implementation of the Classifier from scikit-learn package.
4. Classification Report:
   a. Create a classification report in tabular form.
   b. You need to calculate precision, recall, f1-score and accuracy of the model.

Submission Details: (to be submitted under the specified entry in CSE-Moodle)
1. ZIPPED Code Distribution in CSE-Moodle
2. A brief (2-3 page) report/manual of your work
   (with your hyperparameter tuning results also presented in that report)

Submission Guidelines:

1. You may use one of the following languages: C/C++/Java/Python.
2. Your Programs should run on a Linux Environment.
3. You are **not** allowed to use any library apart from these (Also explore all these libraries if doing in Python, or equivalent of these):
   import numpy # linear algebra
   import csv # data processing, CSV file I/O
   import pandas # data processing, CSV file I/O
   from sklearn.model_selection import train_test_split
   from sklearn.metrics import accuracy_score
   from sklearn.metrics import classification_report

```
from sklearn.model_selection import KFold
from sklearn.tree import DecisionTreeClassifier # sklearn Decision Tree
from sklearn.naive_bayes import MultinomialNB # sklearn Naive Bayes
from sklearn.naive_bayes import GaussianNB # sklearn Naive Bayes
import operator
from math import log
from collections import Counter
```

Your program should be standalone and should **not** use any *special purpose* library for Machine Learning for the naive bayes algorithm. Numpy and Pandas may be used. And, you can use libraries for other purposes, such as generation and formatting of data.

4. You should submit the program file and README file and **not** the output/input file.
5. You should name your file as <GroupNo_ProjectCode.extension>.
6. The submitted program file *should* have the following header comments:
   # Group Number
   # Roll Numbers : Names of members (listed line wise)
   # Project Number
   # Project Title
7. Submit through CSE-MOODLE only.
   Link to our Course page: https://moodlecse.iitkgp.ac.in/moodle/course/view.php?id=561

You should not use any code available on the Web. Submissions found to be plagiarized or having used ML libraries (except for parts where specifically allowed) will be awarded zero marks.

For any questions about the assignment, contact the following TA:
Kajori Ghosh
(Email: kajorighosh4@gmail.com)