Machine Learning Use Case: Predicting Personal Loan Acceptance

SNB 2nd Co-op Project
Loan Eligibility

Department:	EDM
Prepared for:	Naeem AlMaramihi
Prepared by:	Rimas Alshehri Ramah Alharbi Miaad Alsulami Khalid Nimri
Date:	February 6, 2025
Version:	1.0

1 Introduction

The **Bank Personal Loan Dataset** is a structured dataset containing demographic, financial, and behavioral attributes of bank customers. The goal is to predict whether a customer will accept a personal loan offer from the bank.

2 Dataset Description

Source: Kaggle - Bank Personal Loan Modeling
Total Samples: 5,000 bank customers
Total Features: 14 columns (excluding the target variable)
Target Variable: *Personal Loan* (Binary: 0 = No, 1 = Yes)
Data Type: Structured tabular data

3 Feature Description

The dataset consists of the following features:

Feature	**Type**	**Description**
ID	Numerical	Unique customer identifier (Dropped)
Age	Numerical	Customer age in years
Experience	Numerical	Years of professional experience
Income	Numerical	Annual income (in thousands)
ZIP Code	Categorical	Customer's ZIP code (Dropped)
Family	Categorical	Number of family members (1 to 4)
CCAvg	Numerical	Average monthly credit card spending (in thousands)
Education	Categorical	1 = Undergraduate, 2 = Graduate, 3 = Advanced/Professional
Mortgage	Numerical	Mortgage value (in thousands)
Securities Account	Binary	1 = Has a securities account, 0 = No
CD Account	Binary	1 = Has a Certificate of Deposit (CD) account, 0 = No
Online	Binary	1 = Uses online banking, 0 = No
CreditCard	Binary	1 = Has a credit card issued by the bank, 0 = No
Personal Loan (Target)	Binary	1 = Accepted Loan, 0 = Did Not Accept

Table 1: Feature Descriptions in the Bank Personal Loan Dataset

This report provides an overview of customers with and without personal loans in the bank.

- **Customers with personal loans**: 480 (9.6%)
- **Customers without personal loans**: 4520 (90.4%)

Dashboard

The dashboard visualization for the dataset is shown below:

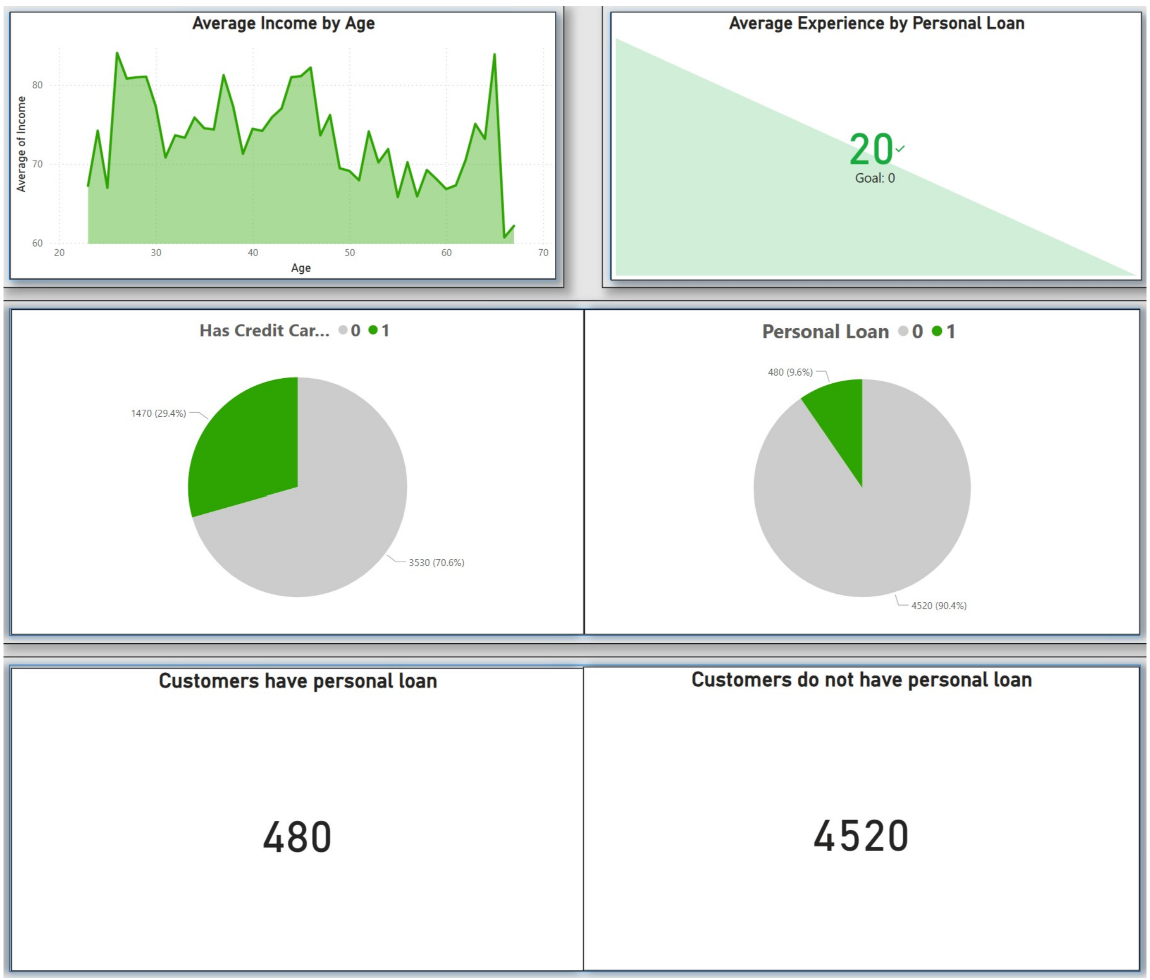

Figure 1: Bank Loan Dashboard Visualization

For an interactive dashboard with detailed insights, click here.

4 Data Preprocessing and Cleaning

Before model training, the following preprocessing steps were applied:

- **Dropped Non-Relevant Features:** Removed ID and ZIP Code.
- **Income Conversion:** Converted annual income to monthly by dividing by 12.
- **Handled Missing Values:** Filled missing values with median imputation.
- **Fixed Negative Values:** Converted negative experience values to absolute values.
- **Feature Scaling:** Applied *StandardScaler* to normalize numerical features.
- **Encoded Categorical Variables:** Converted education and binary attributes into numerical values.

5 Business Use Case

This dataset is useful for **predicting personal loan acceptance** and has key applications in the banking industry:

- **Targeted Marketing:** Identify customers likely to accept loan offers.
- **Credit Risk Analysis:** Assess loan eligibility based on financial history.
- **Customer Segmentation:** Classify customers based on income and spending behavior.
- **Fraud Detection:** Identify unusual financial activity.

6 Machine Learning Approach

To predict personal loan acceptance, we implemented:

6.1 Train-Validation-Test Split

The dataset was split into:

- **Training Set:** 70%
- **Validation Set:** 15%
- **Test Set:** 15%

6.2 Models Used

Four machine learning models were trained and evaluated:

- **Support Vector Machine (SVM) with RBF Kernel**
- **Support Vector Machine (Linear Kernel)**
- **Logistic Regression**
- **Random Forest**

6.3 Evaluation Metrics

The following metrics were used to assess model performance:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-score**
- **ROC-AUC Score**

7 Results

Model	Train Accuracy	Val Accuracy	Recall	Precision	F1-score	AUC Score
SVM (RBF Kernel)	0.9911	0.9853	0.8750	0.9692	0.9197	0.9974
SVM (Linear Kernel)	0.9517	0.9520	0.5556	0.9091	0.6897	0.9653
Logistic Regression	0.9517	0.9467	0.5972	0.7963	0.6825	0.9672
Random Forest	0.9800	0.9800	0.9200	0.9800	0.9500	0.9900

Table 2: Training and Validation Results for Different Models

7.1 Classification Reports

1. SVM (RBF Kernel)

```
              precision    recall  f1-score   support

           0       0.99      1.00      0.99       678
           1       0.97      0.88      0.92        72

    accuracy                           0.99       750
   macro avg       0.98      0.94      0.96       750
weighted avg       0.99      0.99      0.98       750
```

2. SVM (Linear Kernel)

```
              precision    recall  f1-score   support

           0       0.95      0.99      0.97       678
           1       0.91      0.56      0.69        72

    accuracy                           0.95       750
   macro avg       0.93      0.77      0.83       750
weighted avg       0.95      0.95      0.95       750
```

3. Logistic Regression

	precision	recall	f1-score	support
0	0.96	0.98	0.97	678
1	0.80	0.60	0.68	72
accuracy			0.95	750
macro avg	0.88	0.79	0.83	750
weighted avg	0.94	0.95	0.94	750

4. Random Forest

	precision	recall	f1-score	support
0	0.99	1.00	0.99	690
1	0.96	0.85	0.90	60
accuracy			0.99	750
macro avg	0.97	0.92	0.95	750
weighted avg	0.99	0.99	0.98	750

8 Conclusion

The **Bank Personal Loan Dataset** provides insights into customer behavior regarding loan acceptance. By applying machine learning classification models, banks can optimize their **loan approval strategies** and increase **customer engagement** through personalized financial services.
For the global application of these models, refer to the Hugging Face platform for advanced implementations and tools.
You can find the code implementations for the classifiers used in this project below:
Github

Feel free to explore and adapt the models for your projects!

9 References

- Kaggle Dataset: Bank Personal Loan Modeling