# Bank Loan Approval Prediction using Artificial Neural Network

## Problem Statement

The primary goal of this project is to **create an ANN model to predict whether a bank customer** **will be approved for a loan based on various features like income, credit history, and demographics.**

## Dataset Description

The dataset used in this project, 'UniversalBank.csv,' which is taken from “https://github.com/shayansoh/Bank-Loan-Prediction-using-AI” consists of 5,000 records with 14 features. Key features include:

* ID: Customer ID
* Age: Customer Age
* Experience: Amount of work experience in years
* Income: Amount of annual income (in thousands)
* Zipcode: Zipcode of where customer lives
* Family: Number of family members
* CCAvg: Average monthly credit card spendings
* Education: Education level (1: Bachelor, 2: Master, 3: Advanced Degree)
* Mortgage: Mortgage of house (in thousands)
* Securities Account: Boolean of whether customer has a securities account
* CD Account: Boolean of whether customer has Certificate of Deposit account
* Online: Boolean of whether customer uses online banking
* CreditCard: Does the customer use credit card issued by the bank?
* Personal Loan: This is the target variable (Binary Classification Problem)

**Preprocessing Steps**  
1. The “Experience” column had negative values which isn’t possible, so we have replaced the negative values in the `Experience` column with 0 as that’s more logical and has shown improvement in the overall model.  
2. Dropped irrelevant columns (`ID`, ‘zipcode’ and `Education label`) as Customer ID and zip code showed no correlation with variables hence dropping these columns to remove clutter and overfitting issues also to improve overall model performance which was evident.  
3. Standardized numeric features using `StandardScaler`.  
4. Split data into training (60%), Validation (20%) and testing (20%) subsets.

## Methodology

We have employed Artificial Neural Network (ANN) for binary classification. The ANN’s structure and hyperparameters were selected to optimize model performance while maintaining computational efficiency.

### Experimental Details

**Architecture Description**

* Training: 60% ,Validation : 20% , Testing: 20%
* Loss Function: Binary Crossentropy
* Optimizer: Adam.
* Callbacks: Validation set monitoring.
* Batch size: 32.
* Regularization through standardized data scaling.
* Epochs: 50.
* Used validation data for early stopping.

The model architecture is summarized in the table below:

|  |  |  |
| --- | --- | --- |
| Layer | Neurons | Activation |
| Input | 11 Features | - |
| Hidden 1 | 64 | ReLU |
| Hidden 2 | 32 | ReLU |
| Hidden 3 | 16 | ReLU |
| Output Layer | 1 | Sigmoid(Binary Classification) |

Performance Metrics: Accuracy, Precision, Recall, F1-score

### Results

Training and Validation Accuracy:  
Training and Validation Loss:  
Loss decreased steadily which indicates none to minimum overfitting in the model.

* Test Accuracy: 98.20%
* Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted 0 | Predicted 1 |
| Actual 0 | 890 | 14 |
| Actual 1 | 4 | 92 |

* Classification Report:  
   Precision (0): 100%, Recall (0): 98%, F1-score (0): 99%  
   Precision (1): 87%, Recall (1): 96%, F1-score (1): 91%

**F1-Score:** The F1-score balances precision and recall, and a value of 0.91 indicates good performance in identifying both classes. This is a strong indicator of a reliable model.

**Precision:** The model predicts a loan approval (class 1), it is correct 86.79% of the time, which is good but there is room for improvement in minimizing false positives.

**Recall:** The model is very effective at catching most of the true loan approvals(96%).

### Improvements

1. Use PCA for dimensionality reduction which can have a big impact on overall performance of the model.
2. Imbalance between classes can be corrected using SMOTE technique for better performance.

## Conclusion The over model is fit for predicting bank loan approval with these recommendations:

* If the bank is focused on stability and minimizing risks, **Precision** should be the primary metric to consider.
* If the bank wants to grow the business aggresively by approving as many qualified customers as possible, **Recall** should take precedence.
* If the bank is aiming for growth but is not willing to take too many risks, then the **F1 score** provides a balanced approach to decision-making.