

Home Loan Data Analysis

Project Write-Up

Project Overview

This project focuses on predicting whether a loan applicant will repay the loan or default, using historical home loan data. Since loan default can cause financial loss to banks, this model helps in making safer and smarter lending decisions. The main challenge in this dataset is that it is highly imbalanced—most people repay loans, while very few default.

Problem Statement

To ensure safe lending, past loan data must be analyzed. The task is to build a deep learning model that can predict the probability of default for future loan applicants. The dataset contains many features and is imbalanced, which makes prediction more challenging.

Objective

To create a deep learning model that predicts whether an applicant will be able to repay a loan or not using historical data.

Domain

Finance

Methodology

1. **Data Loading**
The dataset was loaded using Python libraries such as Pandas and NumPy.
2. **Null Value Check**
All columns were checked for missing values. Necessary cleaning and handling of null values was done to avoid errors in training.
3. **Target Column Analysis**
The TARGET column was analyzed to find the percentage of defaulters (1) and non-defaulters (0).
This showed that the data was highly imbalanced.
4. **Data Balancing**
Since defaulters were very few, data balancing techniques like oversampling or undersampling were used so that the model could learn both classes properly.
5. **Data Visualization**
Bar graphs and plots were created to show class distribution before and after balancing.
6. **Encoding Categorical Data**
Text-based columns were converted into numeric form using encoding methods so that the deep learning model could process them.
7. **Model Building**
A deep learning model was built using Keras and TensorFlow.
The model was trained on the processed and balanced dataset.
8. **Model Evaluation**
 - **Sensitivity (Recall):** Used to measure how well the model detects actual defaulters.
 - **ROC-AUC Score:** Used to measure overall model performance.

Results

The model successfully learned to predict whether a person will repay a loan or default.

Using sensitivity and ROC-AUC, the model showed good performance in identifying risky applicants.

Conclusion

This project demonstrates how deep learning can be used in the finance domain to reduce loan risk.

By analyzing past loan data and handling class imbalance properly, the model helps banks make better and safer loan approval decisions