**Author:** Divya Pardeshi  
 **Date:** 14-11-2023  
**House Loan Data Analysis Project**

**Objective:**

The project aims to leverage historical loan data to construct a robust deep learning model for predicting loan default probabilities, crucial in maintaining secure lending practices.

**Key Steps:**

1. **Data Loading and Overview:**
   * Loaded the dataset, meticulously examined for null values across crucial columns.
   * Conducted an analysis of the target variable, highlighting the default percentage within the dataset.
2. **Data Preprocessing:**
   * Handled null values via appropriate imputation strategies based on data types (mean for numeric, mode for categorical).
   * Executed encoding on categorical columns essential for model construction.
3. **Handling Imbalanced Data:**
   * Balanced the imbalanced dataset using random undersampling, ensuring equitable representation of default and non-default cases.
   * Applied feature scaling via StandardScaler on the balanced dataset to standardize features.
4. **Model Development:**
   * Constructed a deep learning model architecture leveraging TensorFlow's Keras API.
   * Integrated multiple dense layers and dropout regularization to combat overfitting.
   * Trained, evaluated, and optimized the initial model, utilizing metrics like accuracy, precision, recall, and ROC-AUC.
5. **Hyperparameter Tuning and Model Refinement:**
   * Implemented Stratified K-Fold cross-validation to fine-tune crucial hyperparameters such as epochs and batch sizes.
   * Conducted comprehensive error analysis and refinement techniques to augment the model's predictive capacity.
6. **Model Evaluation and Conclusion:**
   * Computed a spectrum of model performance metrics (accuracy, precision, recall, F1-score, and ROC-AUC) for the final model.
   * Noted an improvement in AUC-ROC from 0.7301 pre-hyperparameter tuning to 0.7320 post-tuning, underscoring the model's enhancement potential.

**Conclusion:**

This project effectively harnessed deep learning methodologies and rigorous data analysis to create a predictive model for loan default probabilities. Despite the initial dataset complexities, the model demonstrated a slight but noteworthy performance boost post hyperparameter tuning, showcasing promise for future loan risk assessments.