PREDICTING BANK LOAN ELIGIBILITY USING MACHINE LEARNING MODELS

ABSTRACT

- The prediction of loan eligibility is a critical task for financial institutions, directly impacting their earnings and the accuracy of credit scores. With the increasing volume of loan applications, traditional methods of eligibility assessment are becoming inefficient and may **fail to correctly identify suitable loan recipients**, leading to higher financial risks.
- This study investigates the application of Artificial Intelligence (AI) techniques, specifically Machine Learning (ML) to predict loan eligibility more accurately and efficiently.
- We explore the performance of several ML classification algorithms (**Gaussian Naive Bayes, AdaBoost, Gradient Boosting, K Neighbors Classifier, Decision Trees, Random Forest, and Logistic Regression**).
- Additionally, the study integrates Ensemble Methods and SMOTE with SMOTE-TOMEK techniques to address class **imbalance**, improving model effectiveness. The models are evaluated using four key metrics: accuracy, precision, recall, and F1-measure.
- These findings demonstrate the potential of AI-driven predictive models to enhance loan eligibility assessments, reduce financial risks, and optimize credit scoring in the financial industry.

Contents

Introduction

Overview of Machine Learning in Banking

Importance of ML in Financial Services

Machine Learning (ML) plays a crucial role in financial services by improving decision- making accuracy, enhancing customer experiences, and increasing operational efficiency through automated processes and predictive analytics.

Historical Perspective on Loan Prediction

Traditionally, loan prediction relied heavily on manual assessment and static models. The evolution of ML has significantly advanced this field, offering dynamic and data- driven prediction models that increase reliability and reduce biases.

Objectives and Scope

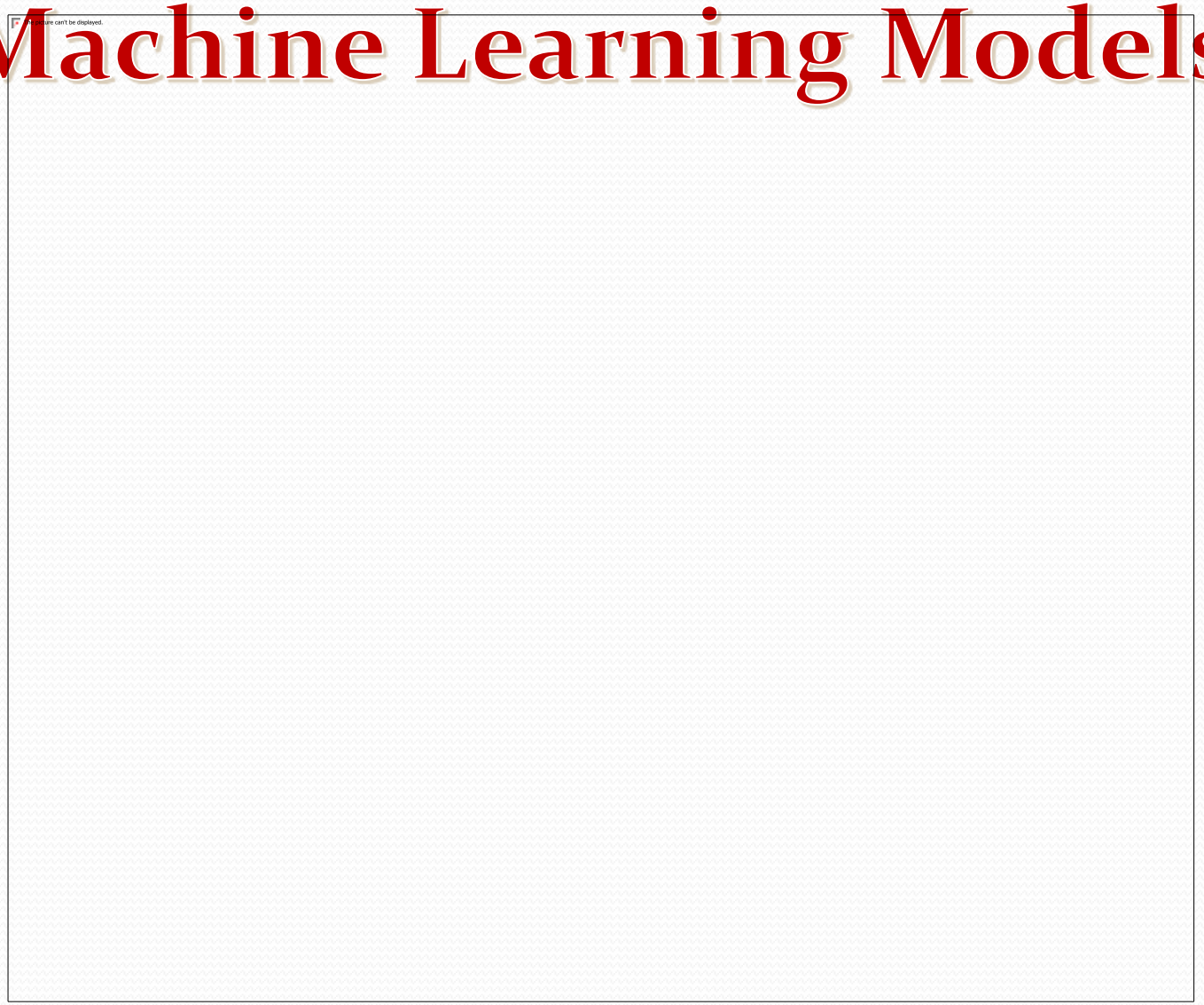
Presentation Goals

The primary goals of this presentation are to introduce the application of ML in banking, explain its importance, and provide insights into loan prediction through data- driven methods.

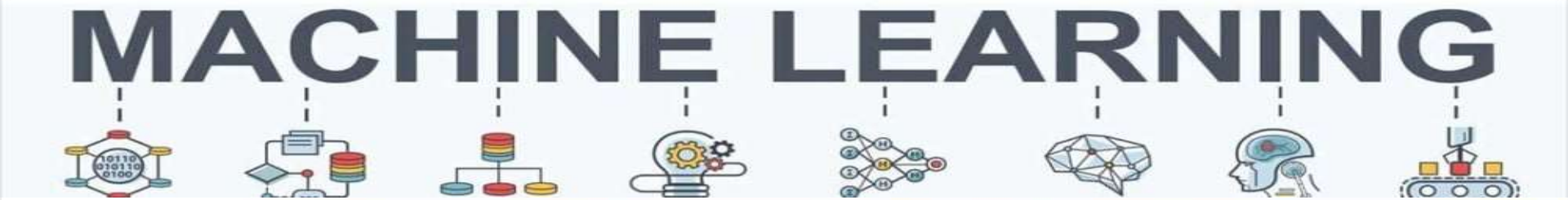
Outline of Key Topics

This presentation will cover an overview of ML in the banking sector, the significance of ML for loan prediction, a historical perspective, and the goals and scope of ML applications in financial services.

Machine Learning Models



Supervised Learning Models



01

Logistic Regression

Logistic Regression is a statistical model used to predict the probability of a binary outcome, relying on a logistic function to model the data.

02

Decision Trees

Decision Trees are a non- parametric supervised learning method used for classification and regression by learning simple decision rules from the data.

03

Support Vector Machines

Support Vector Machines are supervised learning models that analyze data for classification and regression by finding the hyperplane that best separates classes.

Ensemble Learning Methods

Random Forests

Random Forests are an ensemble learning method that construct multiple decision trees during training and output the class that is the mode of the classes.

Gradient Boosting

Gradient Boosting is a machine learning technique used for regression and classification tasks, producing a prediction model in the form of an ensemble of weak prediction models.

Usage Scenarios and Applications

Ensemble methods like Random Forests and Gradient Boosting are commonly used in various applications such as fraud detection, medical diagnosis, and stock market prediction.

Steps involved in creating Loan Eligibility Prediction Model

Data Collection and Preprocessing

Data Sources

Data for loan eligibility prediction can be gathered from multiple sources such as financial institutions, credit bureaus, and customer self- reported information. It is essential to ensure data accuracy and relevance.

Handling Missing Values

When dealing with loan application data, missing values must be addressed. Techniques such as mean imputation, median imputation, or using predictive algorithms to estimate missing data ensure model reliability.

Feature Engineering

Feature engineering involves creating new variables or transforming existing ones to better represent the underlying patterns in the data. This may include normalizing income, encoding categorical variables, and creating interaction terms.

Model Training and Evaluation

Splitting Data into Training and Test Sets

The dataset is divided into training and test sets to evaluate model performance. Typically, 70- 80% of data is used for training, while the remaining 20- 30% is reserved for testing, ensuring unbiased evaluation.

Metrics for Model Evaluation

Common metrics for evaluating loan eligibility models include accuracy, precision, recall, F1- score, and the area under the receiver operating characteristic (ROC) curve. These metrics help in assessing the model's predictive power.

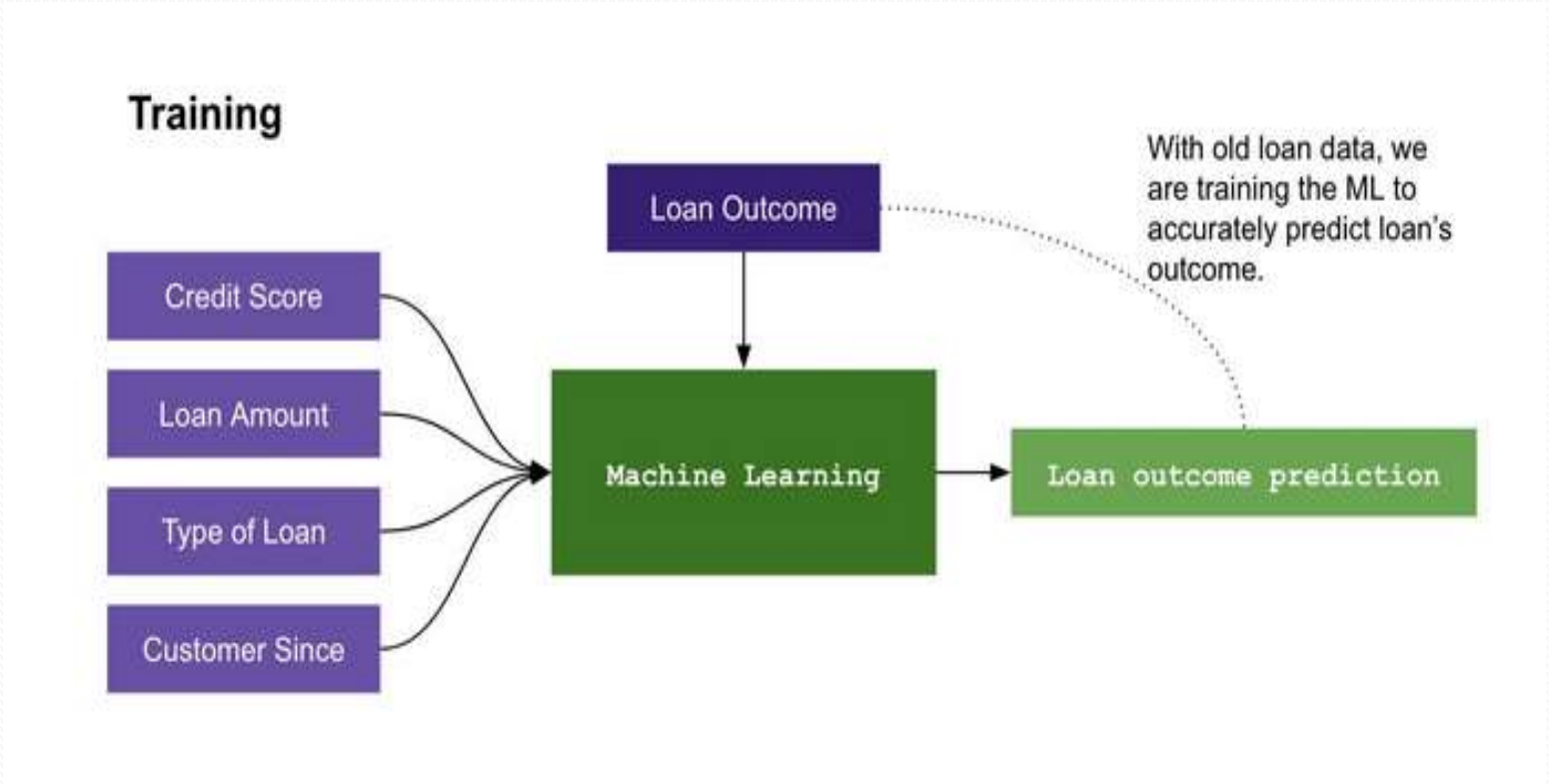
Cross-Validation Techniques

Cross- validation, such as k- fold cross- validation, is employed to assess the model's performance across different subsets of the data. This technique reduces the risk of overfitting and provides a more generalized performance estimate.

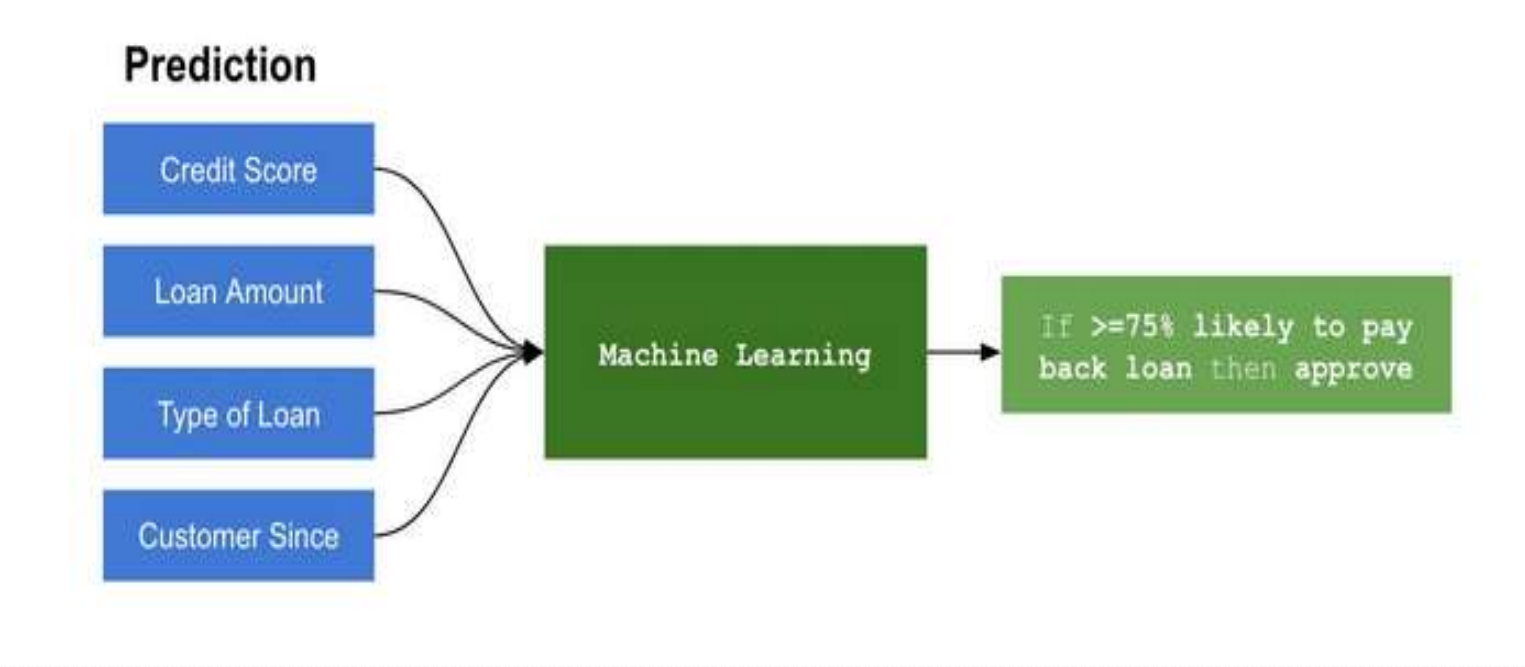
Hyperparameter Tuning

Hyperparameter tuning involves adjusting the parameters of the learning algorithm to improve performance. Techniques like grid search, random search, and Bayesian optimization are used to find the optimal hyperparameters set.

Model Creation



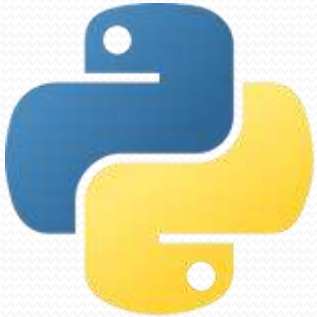
Deployment Stage



Advantage:

- Easy to Select the Eligible Candidate to Issue Loans Without any Major Loan Default Risk
- Reduces Traditional Manual Human Effort

Software Requirement



- Python
- Django (Web Framework)

Library Used

- Scikit-learn
- Pickle
- Numpy
- Pandas

THANK YOU

Project Done By : THANIGAIVEL G

Project Guided By :

- DR. A .MANIKANDAN
- Mrs. KALPANA