Loan Prediction

This presentation explores the application of machine learning techniques to predict loan eligibility. Housing Finance companies can leverage these models to automate the loan approval process, reduce Non-Performing Assets, and identify target customer segments.
We will cover the problem statement, algorithms used, implementation details, results, applications, and conclusion.

By Neeraj Rijhwani
Rollno. 64

Introduction to Loan Prediction

Finance and Lending

Finance raising and lending for real estate, consumer, mortgage and companies‘ loans is the central part of almost every bank‘s business model. Lending money to inappropriate customers forms the major source of credit risk.

The Challenge

Banking companies face a dual challenge: distinguishing deliberate defaulters and addressing bias from bank employees influenced by developers of defaulting companies.

The Goal

The primary goal is to safely invest capital. Machine learning automates authentication, predicting applicant security for housing finance companies.

Problem Statement: Loan Eligibility Prediction

Dream Housing Finance

Dream Housing Finance automates loan eligibility based on online application details like Gender, Marital Status, Education, Dependents, Income, Loan Amount, and Credit History.

Business Goal

The business goal is to process user data in real-time and identify customer segments eligible for loan amounts, enabling targeted marketing.

Objectives

Build a machine learning model with high accuracy and assist housing finance companies in customer segmentation for successful loan acquisition and repayment.

Algorithm Description

Logistic Regression

Models the probability of a discrete outcome given an input variable, commonly used for binary outcomes.

Random Forest

Builds decision trees on different samples, using majority vote for classification and averaging for regression.

K-Nearest Neighbors

Classifies new data points based on similarity to available cases, categorizing them into the most similar category.

Gaussian Naive Bayes

An extension of Naive Bayes, using the Gaussian distribution to estimate data distribution.

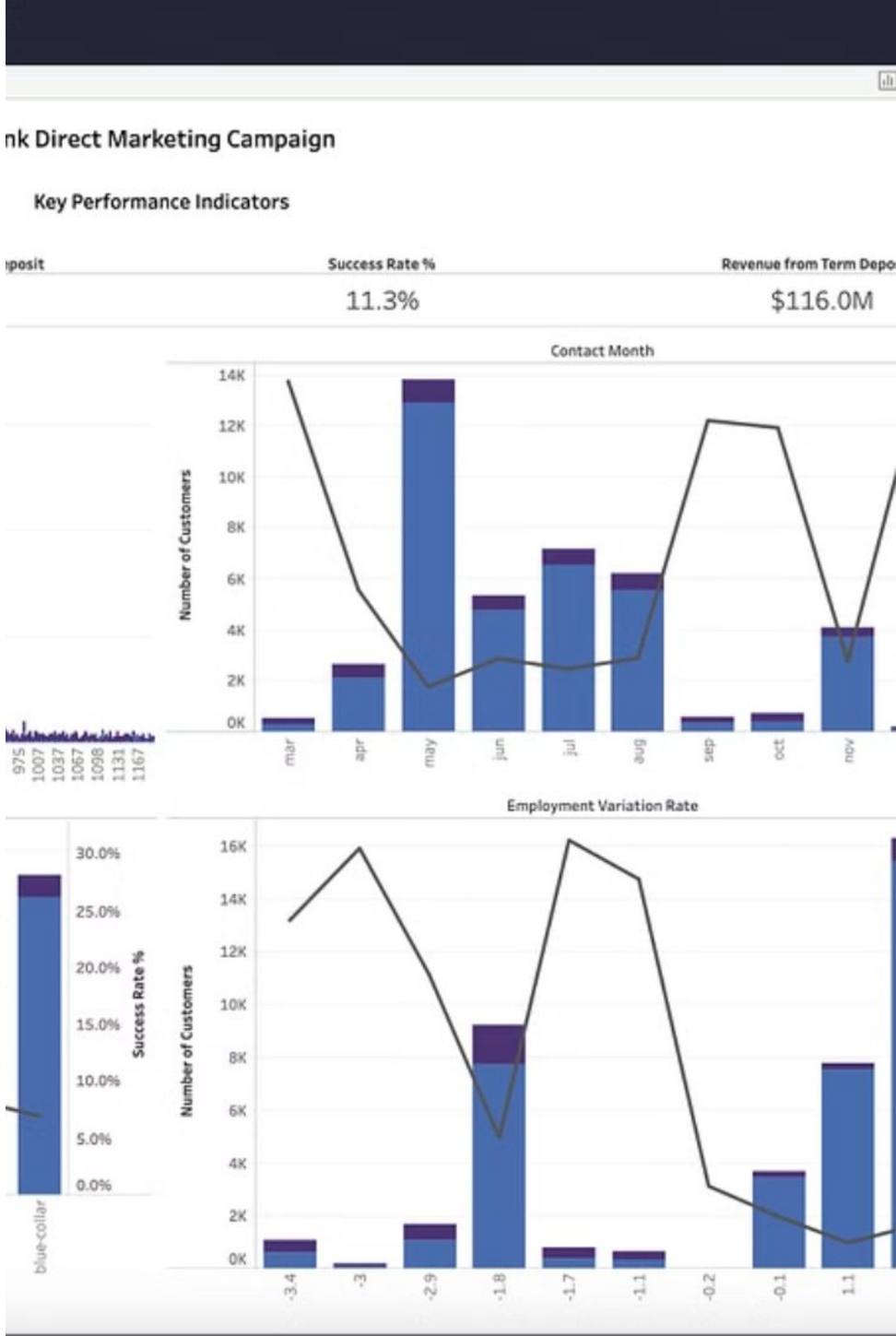
SVM Classifier

Creates an optimal line or decision boundary to segregate n-dimensional space into classes.

Gradient Boosting

Combines many weak learning models to create a strong predictive model, often using decision trees.

These algorithms are used to predict loan eligibility by analyzing various customer attributes and historical data.



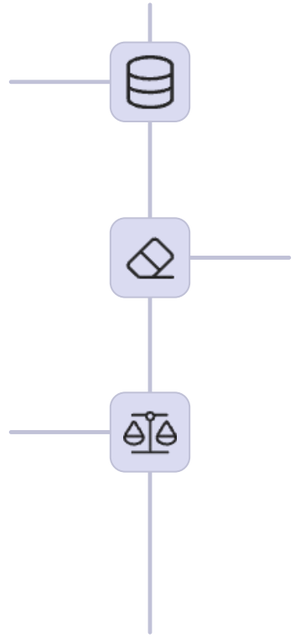
Implementation: Pre-processing and Feature Scaling

Dataset Description

The dataset includes details like Gender, Marital Status, Education, Dependents, Income, Loan Amount, and Credit History.

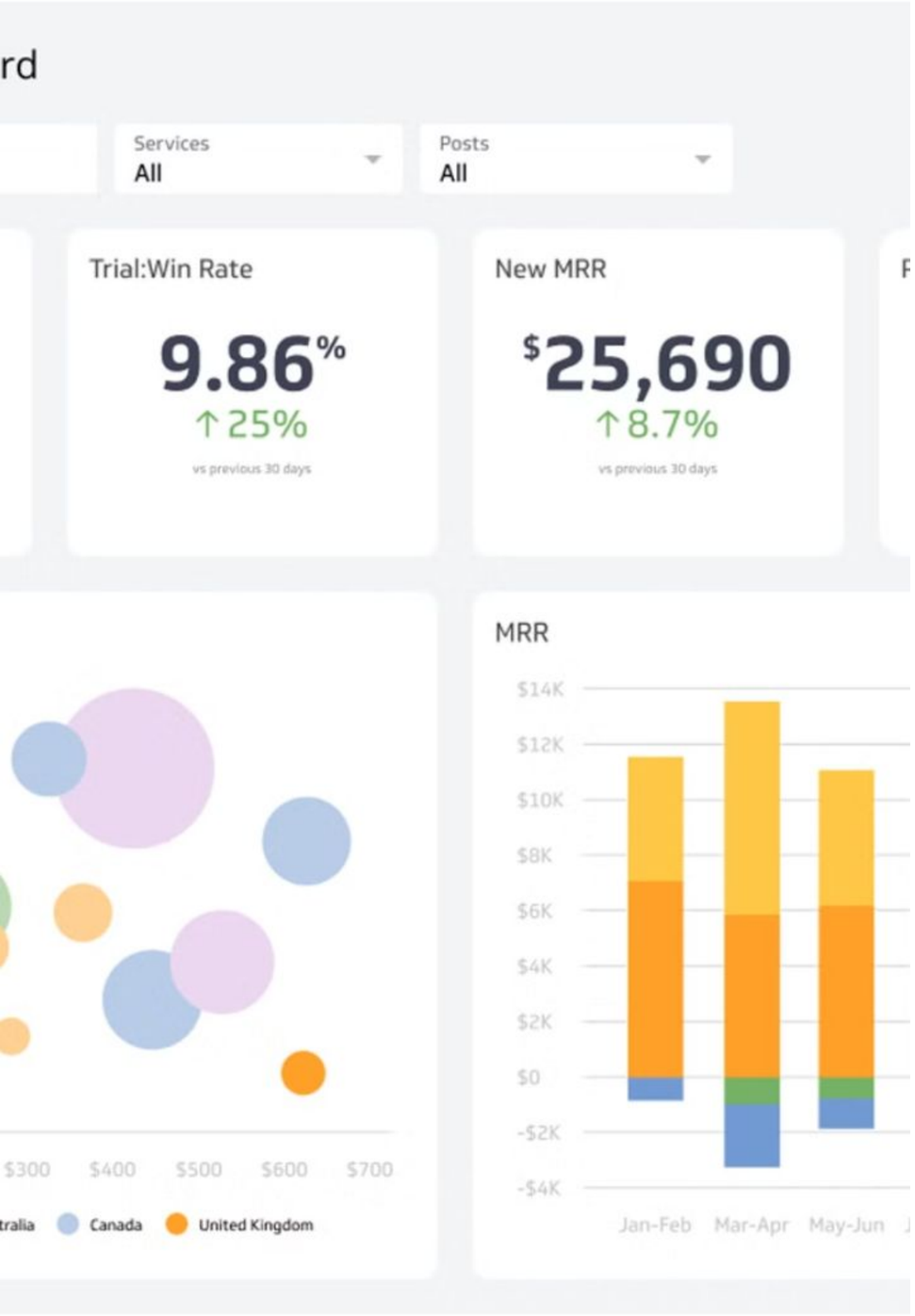
Feature Scaling

Categorical attributes are converted into numerical counterparts for easier classification.



Handling Missing Data

Missing numerical values are imputed using the median, while categorical values are filled using the mode.



Results and Analysis

1

Data Modeling Summary

Logistic Regression achieved the highest accuracy at 80.00%.

2

Hypothesis Testing

A paired t-test showed no significant difference between Logistic Regression and Naive Bayes.

3

Feature Importance

Credit History, Applicant's Income, Loan Amount, Dependents, and Property Area are key factors.

Results: Performance Metrics and Feature Importance

Performance Metrics

- Accuracy: 0.8
- Confusion Matrix: \[\[16 36] \[1 132]]
- Classification Report: (See original document)

Feature Importance

Credit History has maximum importance, while Employment has the least.

Loan status is strongly related to Credit History, Applicant's Income, Loan Amount, Dependents, and Property Area.

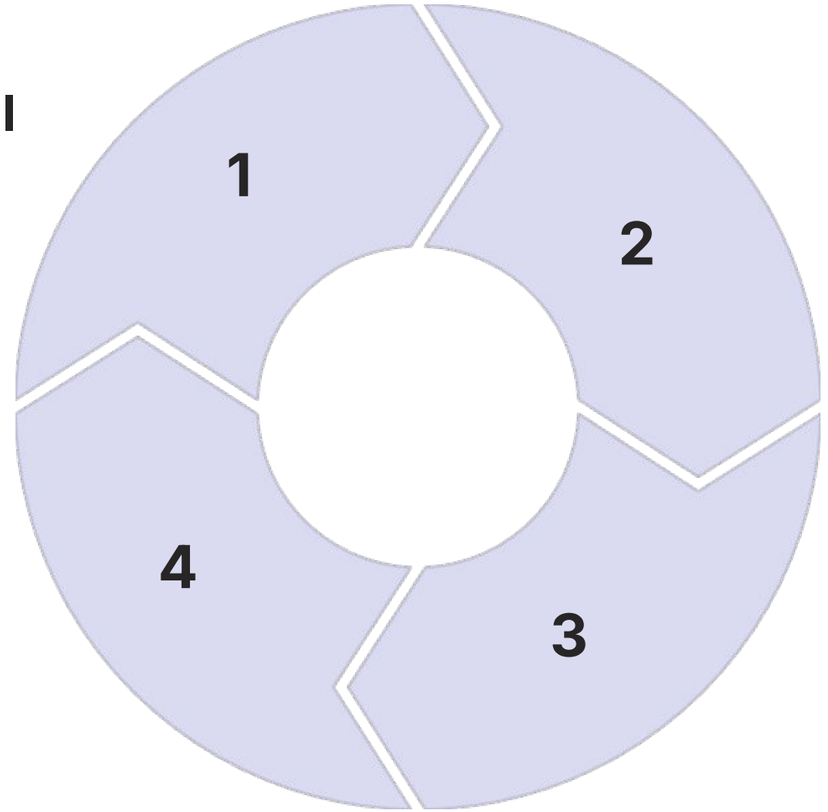
Applications of Loan Prediction

Automated Loan Approval

Streamline the loan approval process with real-time eligibility checks.

Personalized Offers

Tailor loan products to suit individual customer profiles.



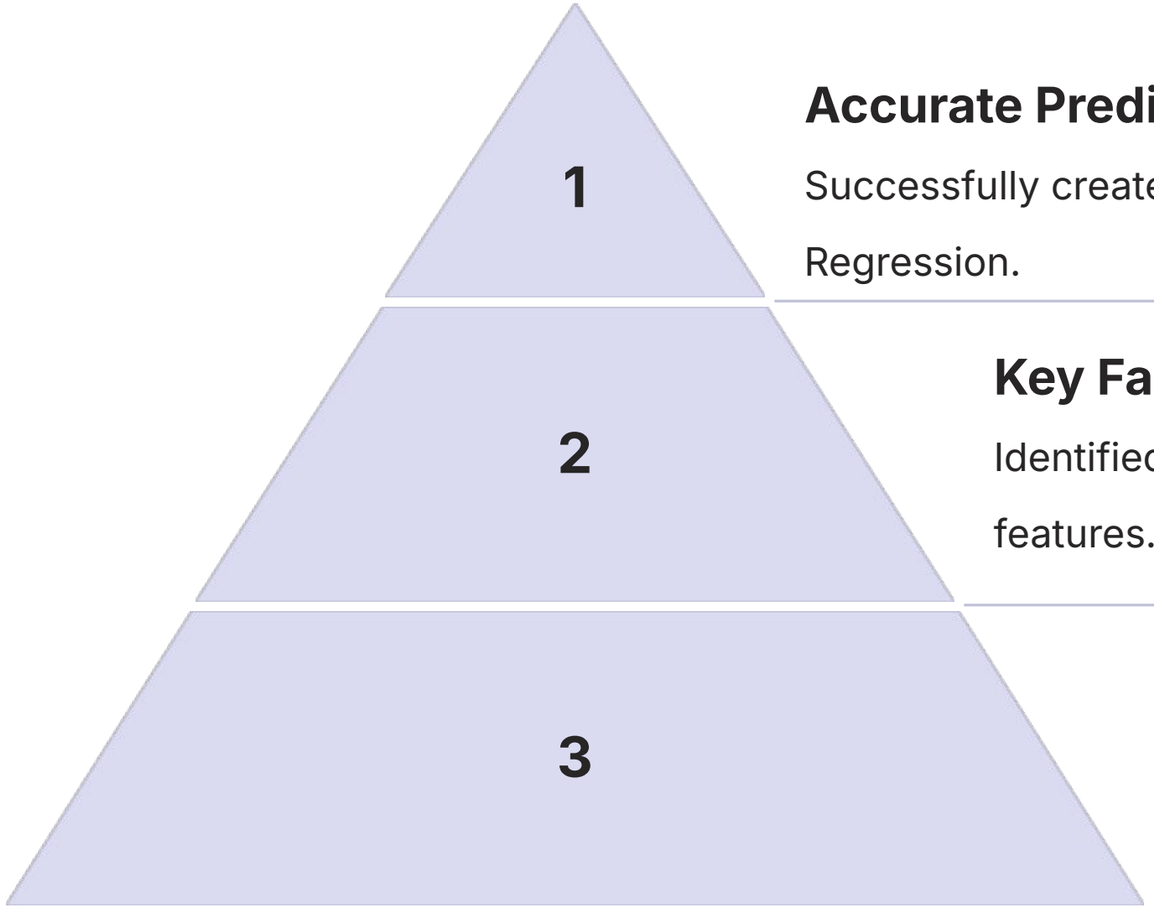
Customer Segmentation

Identify target customer segments for marketing efforts.

Risk Reduction

Minimize Non-Performing Assets by predicting loan defaulters.

Conclusion



Accurate Prediction

Successfully created a Classification ML Model using Logistic Regression.

Key Factors

Identified Credit History, Income, and Loan Amount as crucial features.

Practical Application

Enables automated loan eligibility and targeted customer segmentation.

The model achieved a score of 0.79166 on the submission, with an overall ranking of 944.