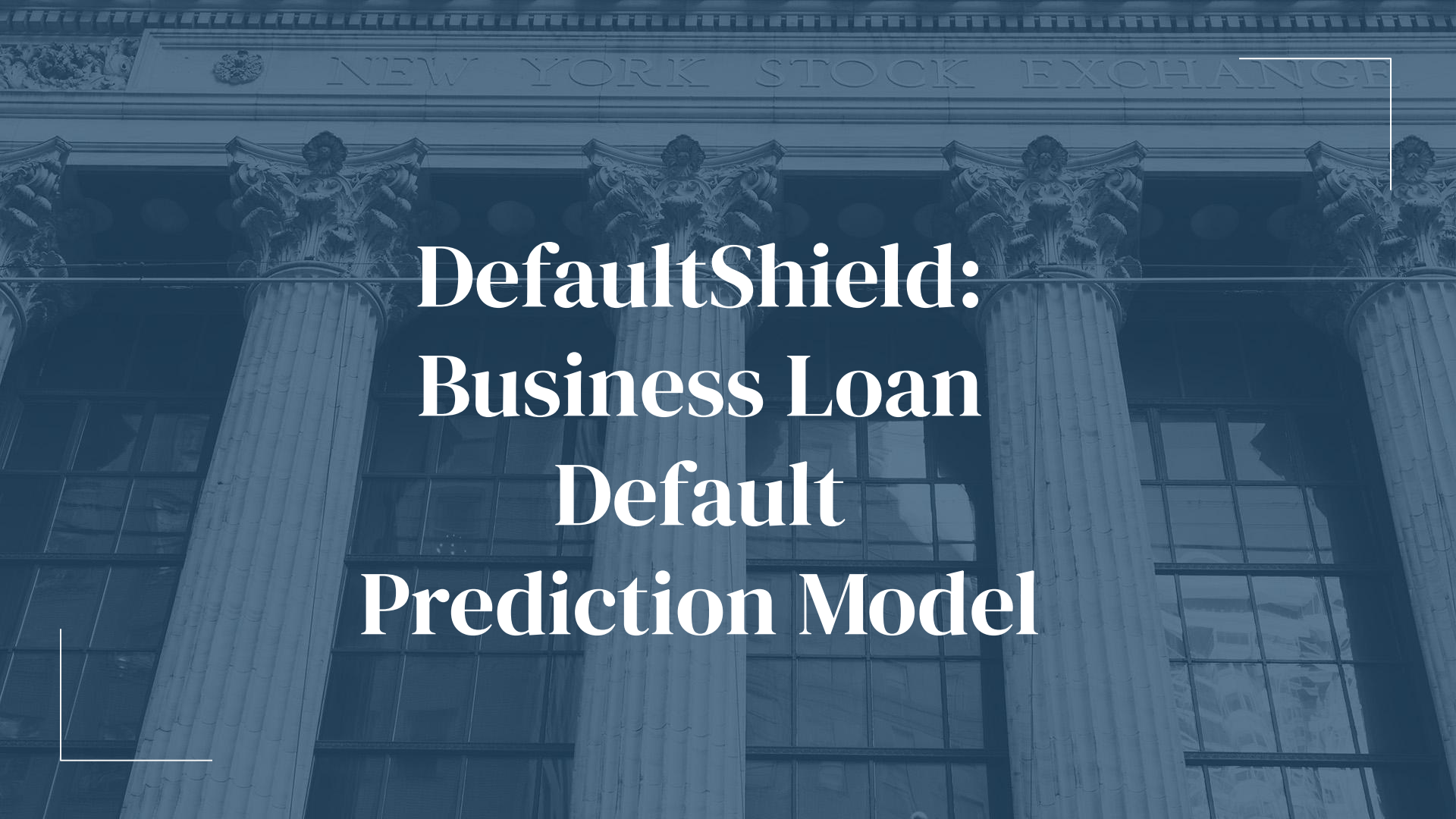


DefaultShield: Business Loan Default Prediction Model

Table of Contents

01 Project Overview

Problem : Business Loan Default

Critical Challenge in Finance: Accurately predicting business loan defaults.

Risks: Financial losses and limited credit provision due to prediction inaccuracies.

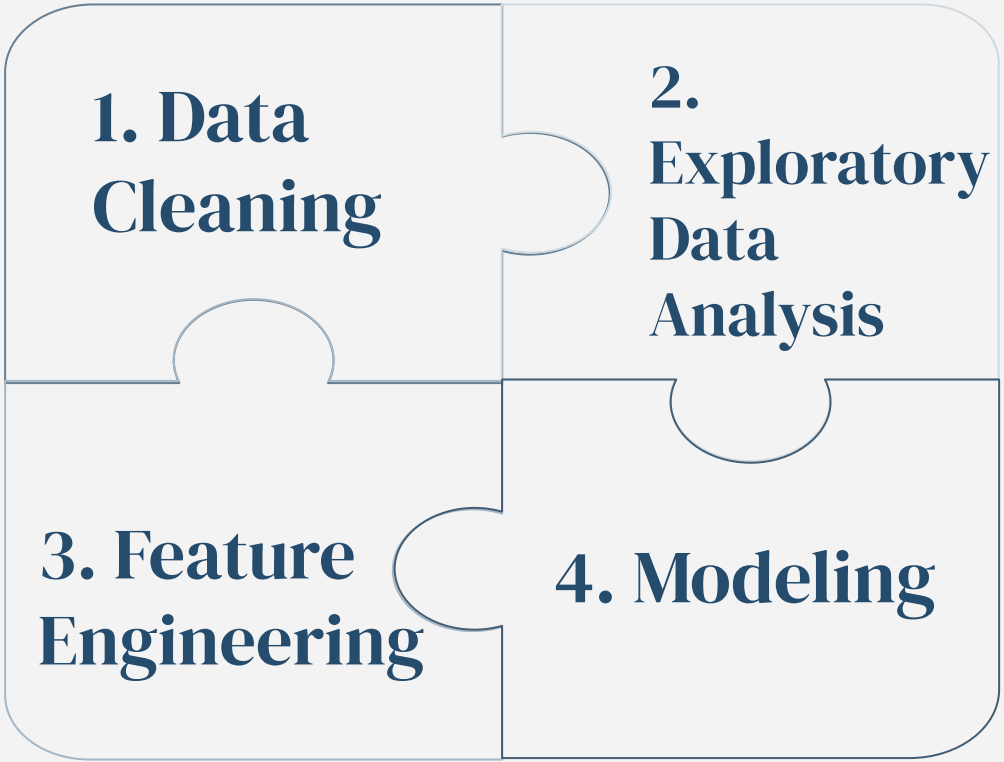
Solution : Prediction Model

For Lenders: Significantly reduces the risk of losing money on bad loans.

For the Economy: More accurate predictions mean more businesses can get the loans they need to grow, boosting the economy.

For Financial Health: Leads to a more stable and reliable lending environment.

Workflow


1. Data Cleaning
2. Exploratory Data Analysis
3. Feature Engineering
4. Modeling

Data Source

02

Data Source

Origin of Data: Our project utilizes data from the Small Business Administration (SBA), known for its extensive records on small business loans.

Focus: This dataset specifically targets small businesses seeking loans, providing a relevant and focused perspective for our analysis.

#	Variable Name	Description
1	LoanNr_ChkDgt	Identifier Primary key
2	Name	Borrower name
3	City	Borrower city
4	State	Borrower state
5	Zip	Borrower zip code
6	Bank	Bank name
7	BankState	Bank state
8	NAICS	North American industry classification system code
9	ApprovalDate	Date SBA commitment issued
10	ApprovalFY	Fiscal year of commitment
11	Term	Loan term in months
12	NoEmp	Number of business employees
13	NewExist	1 = Existing business, 2 = New business
14	CreateJob	Number of jobs created
15	RetainedJob	Number of jobs retained
16	FranchiseCode	Franchise code, (00000 or 00001) = No franchise
17	UrbanRural	1 = Urban, 2 = rural, 0 = undefined
18	RevLineCr	Revolving line of credit: Y = Yes, N = No
19	LowDoc	LowDoc Loan Program: Y = Yes, N = No
20	ChgOffDate	The date when a loan is declared to be in default
21	DisbursementDate	Disbursement date
22	DisbursementGross	Amount disbursed
23	BalanceGross	Gross amount outstanding
24	MIS_Status	Loan status charged off = CHGOFF, Paid in full =PIF
25	ChgOffPrinGr	Charged-off amount
26	GrAppv	Gross amount of loan approved by bank
27	SBA_Appv	SBA's guaranteed amount of approved loan

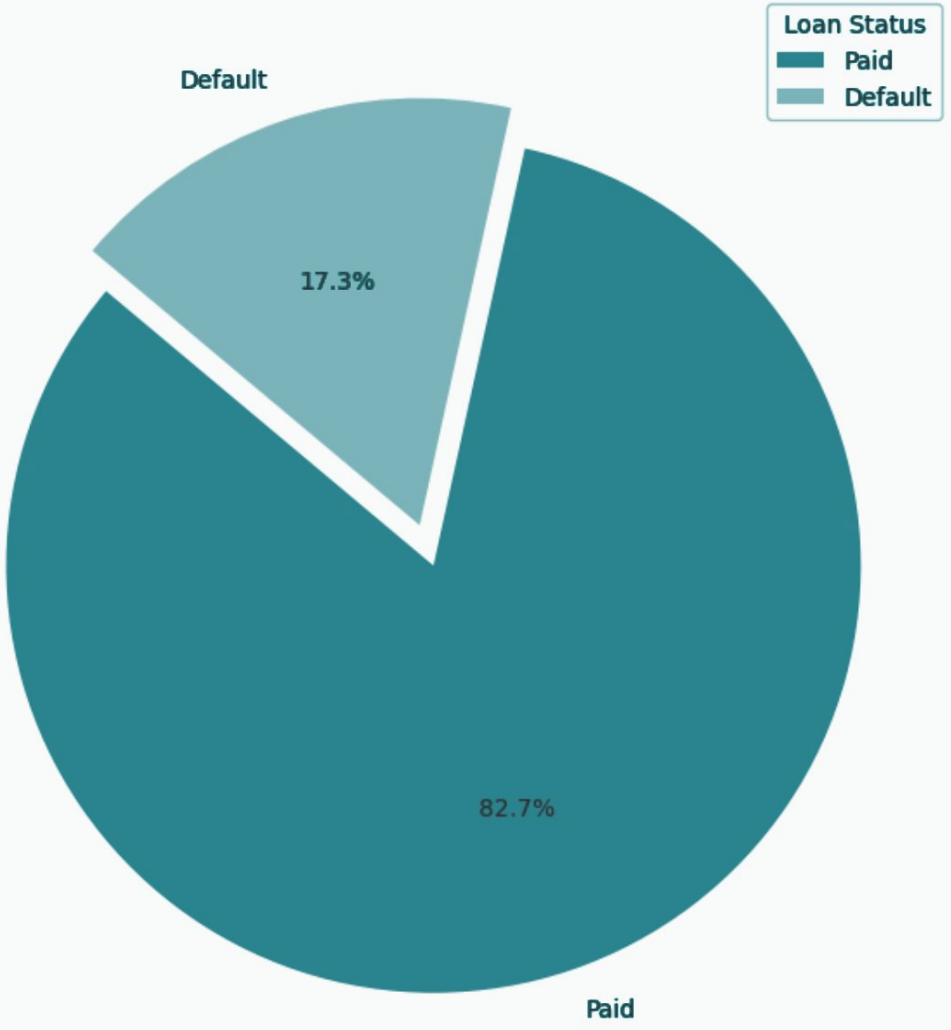
Data Characteristics

Dataset Overview: The dataset offers in-depth information on loan applications, encompassing various aspects of the loan process.

Details Covered: It includes borrower information, loan terms, and other crucial factors pivotal for predicting loan default risks.

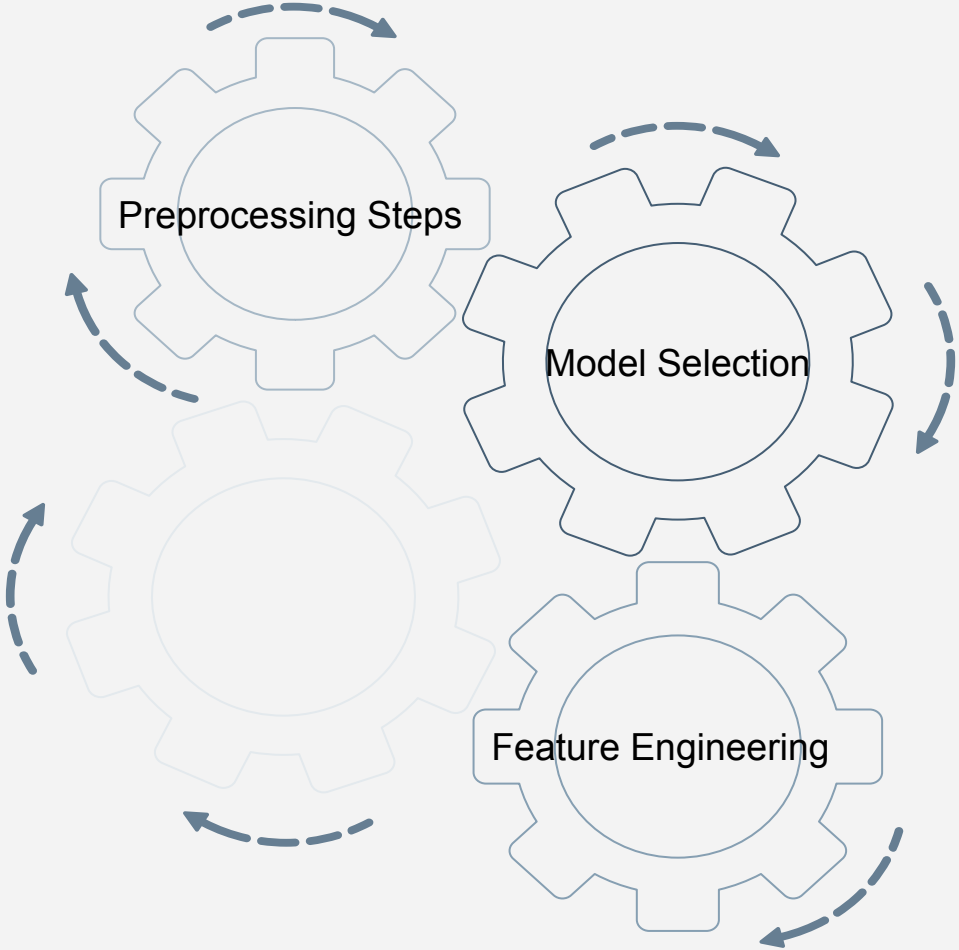
Key Variables: As outlined in the data dictionary, key variables include borrower financials, loan status, repayment terms, business characteristics, and credit history.

Percentage of Different Loan Status



03 Methodology

Methodology



Key Insights

04

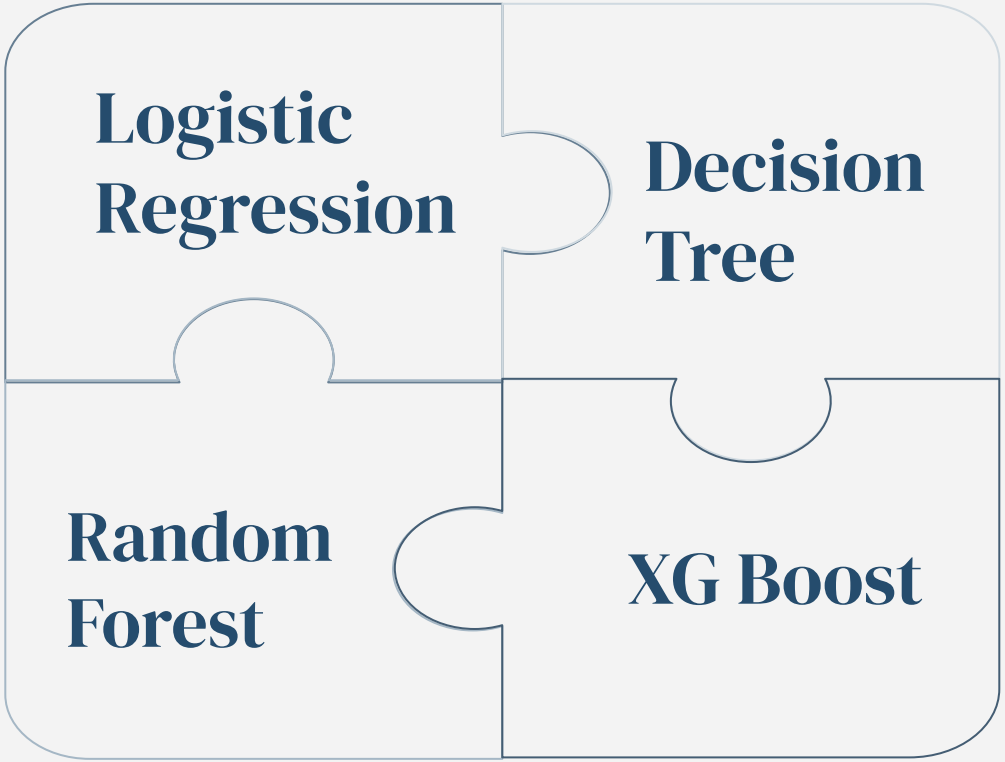
Key Insights

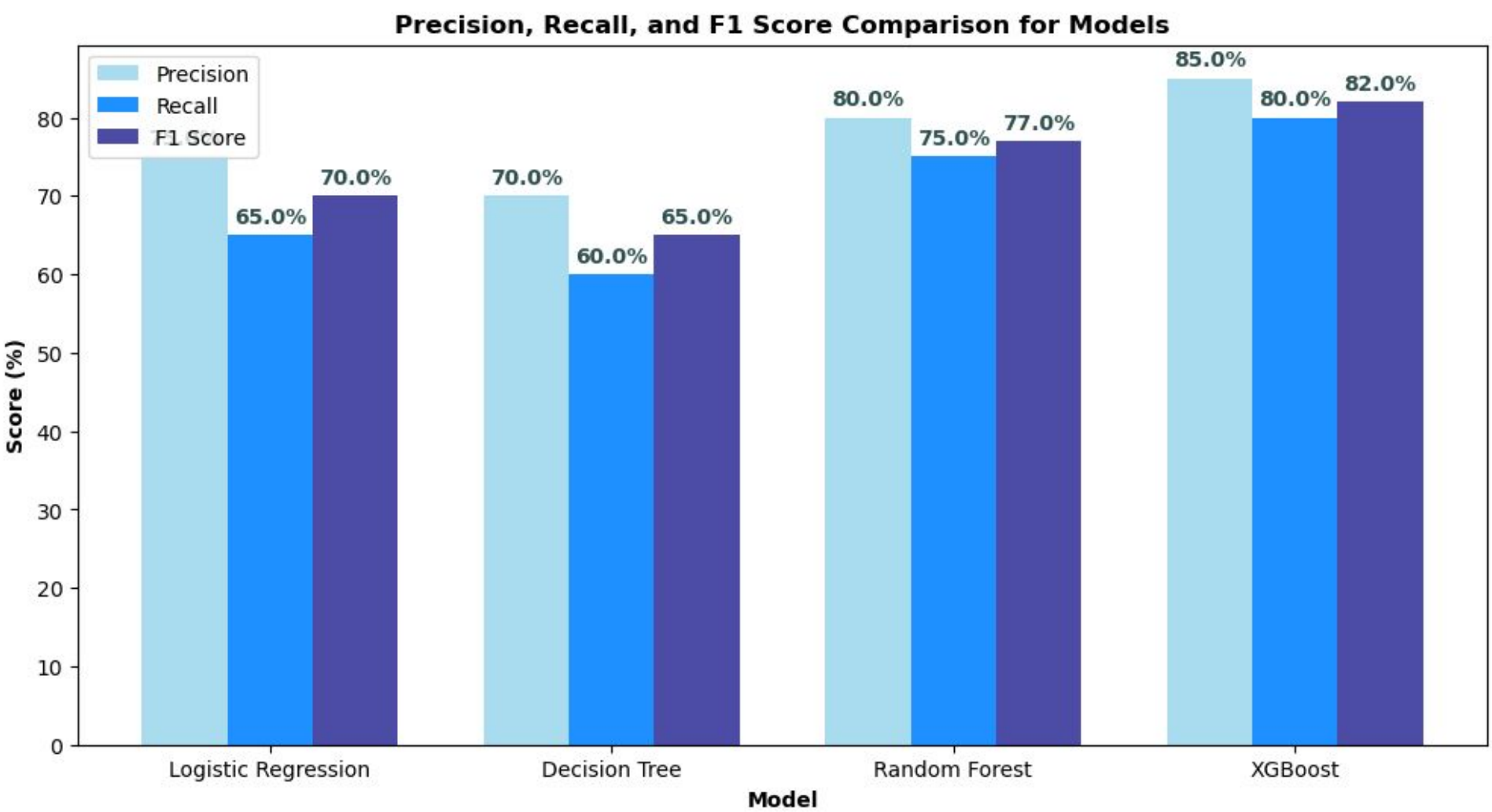
EDA revealed significant insights: Industry type strongly correlates with default rates, and larger loan amounts are more prone to default.

Data Imbalance Between Paid and Default Status: There was a high imbalance in the dataset between loans that were paid off and those that defaulted. This imbalance presents a challenge for predictive modeling as it can skew the results towards the majority class (paid status).

05 Model Analysis

Machine Learning Models




Precision, Recall, and F1 Score Comparison for Models
Precision
Recall
F1 Score
Score (%)
0
10
20
30
40
50
60
70
80
65.0%
70.0%
70.0%
60.0%
65.0%
80.0%
75.0%
77.0%
85.0%
80.0%
82.0%
Logistic Regression
Decision Tree
Random Forest
XGBoost
Model

06

Conclusion

Conclusion

Project Overview: Developed a predictive model for assessing the risk of loan defaults in small businesses using data from the Small Business Administration.

Key Findings: Identified critical factors influencing loan defaults, including loan amount, business age, industry trends, and borrower's credit score.

Model Performance: Successfully implemented and compared multiple models, with showing the most promise in terms of precision, recall, and F1 scores.