CLASSIFYING LOAN APPLICATIONS AND OPTIMIZING INTEREST RATES

Valentin Pinon and Jan Philipp Girgott

December 13th, 2023 | Machine Learning Under a Modern Optimization Lens

**MIT MANAGEMENT** SLOAN SCHOOL

LOAN DEFAULTS LEAD TO CREDIT DAMAGE, LENDER LOSSES, AND ECONOMIC TURBULENCE

PROBLEMS

1 **Loss for the lender:** lower profitability and stability, higher interest rates

2 **Credit damage for borrower:** lower credit score affects ability to get housing/job

3 **Economic impact:** less access to capital, slower economic growth, financial crises

DELINQUENCY RATE: US COMMERICAL BANKS

Legend:
— Commerical Loans
— Mortages

Y-axis: Delinquency Rate (%)
X-axis: Quarter (Q1 2012 – Q1 2023)
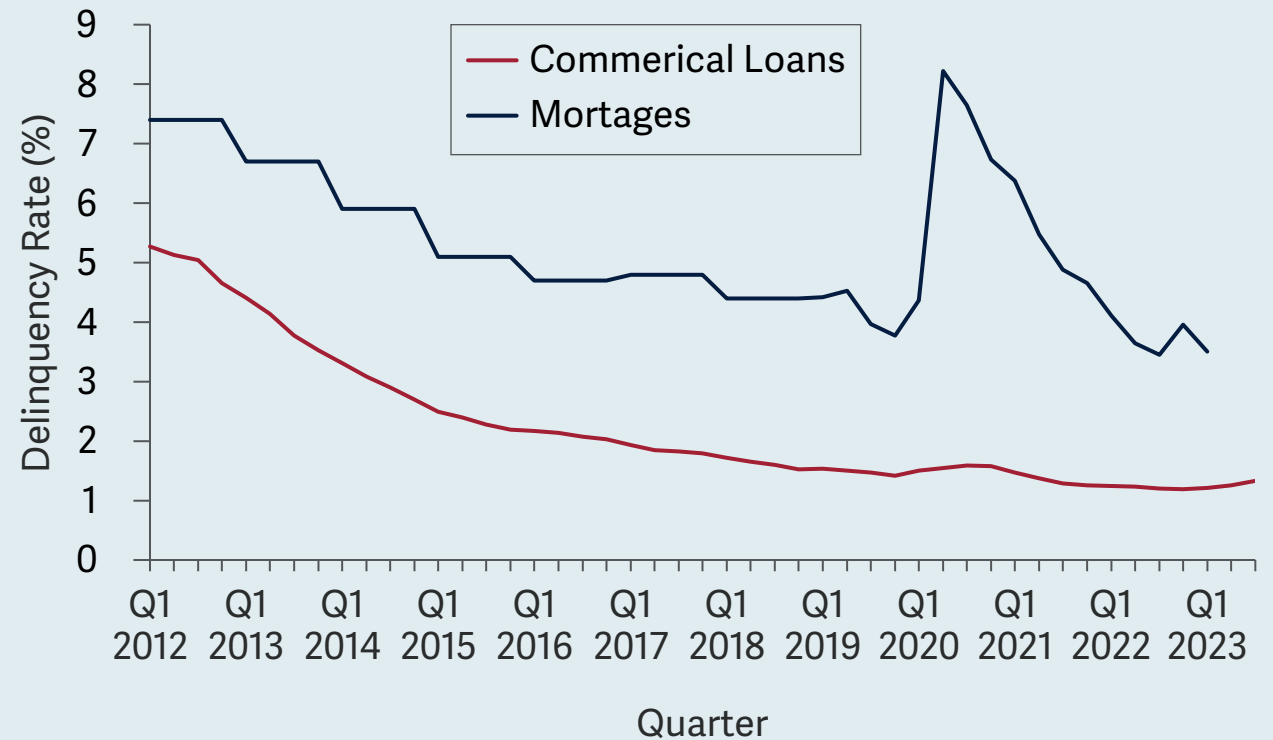
WE USED TWO DATASETS ON CURRENT AND PREVIOUS LOAN APPLICATIONS WITH > 100 FEATURES

PREPROCESSING

Two datasets (some client overlap):
- Current applications
- Previous applications

Preprocessing steps:
1. Treat **missing values** (drop columns, treat NMAR values)
2. Prepare datasets for merge (drop **non-common columns**, **enrich** data where **reasonable**)
3. **Merge** datasets
4. Avoid **leakage** in train-test split

Personal
Gender, education, family, age, contact data

Financials
Income, employment, car ownership, realty

Housing
Location (incl. rating), data on the property

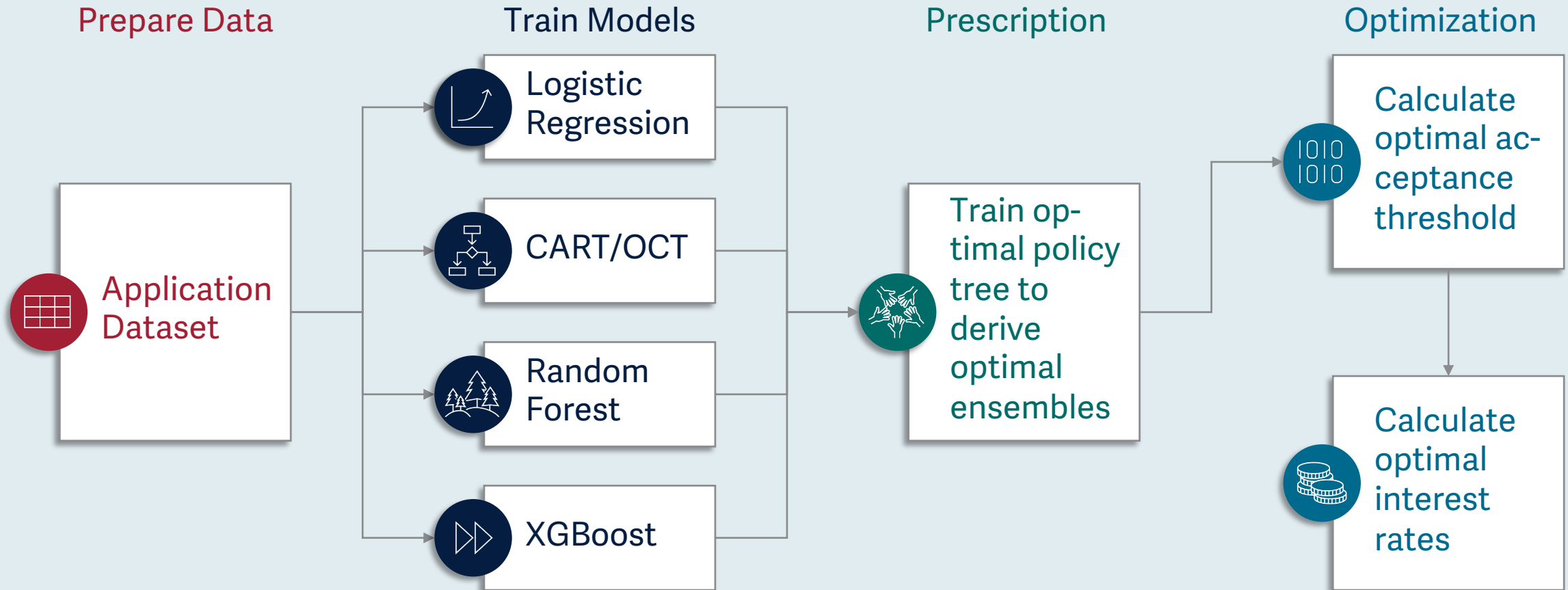**Ratings**
External ratings, credit score

Loan
Amount, purpose, contract type, provided documents

MIT MANAGEMENT SLOAN SCHOOL

WE ARE COMBINING CLASSIFICATION, PRESCRIPTION AND OPTIMIZATION IN OUR APPROACH

Prepare Data

Application Dataset

Train Models

Logistic Regression

CART/OCT

Random Forest

XGBoost

Prescription

Train optimal policy tree to derive optimal ensembles

Optimization

Calculate optimal acceptance threshold

Calculate optimal interest rates

MACHINE LEARNING MODEL PERFORMANCES
XGBOOST ACHIEVES THE HIGHEST AUC AND ALMOST PERFECT SPECIFICITY

	Out-of-Sample Values	AUC	Sensitivity (TPR)	Specificity (TNR)
Individual Models	CART	0.918	85.1%	98.5%
	OCT	0.909	78.1%	99.8%
	Rand. Forest	0.93	82.8%	87.5%
	XGBoost	0.959	90.3%	99.9%
	Log. Reg. (l2)	0.95	90.0%	100.0%
OPT	**Ensemble**	**0.964**	**91.8%**	**94.3%**

MIT
MANAGEMENT
SLOAN SCHOOL

THE OPTIMIZATION MODEL CHOOSES AN ACCEPTANCE THRESHOLD OF 25.9%

INSIGHTS FROM CLASSIFICATION

By analyzing the trees and feature impor-tance values, we derived these top 5 features:

- 1. Income type
- 2. Education type
- 3. Car and realty ownership
- 4. Gender
- 5. Family status

INSIGHTS FROM OPTIMIZATION

25.9% **Acceptance threshold** for the default probability

84.2% **Loans coherently rejected** compared to dataset decisions

82.7% **Loans coherently accepted** compared to dataset decisions

- **Higher interest** rates for loan applications with **higher default risk**
- Ceteris paribus, **higher loans** receive **higher interest rates** from the model

MIT MANAGEMENT SLOAN SCHOOL

FINE-TUNING OF OPTIMIZATION PARAMETERS AND MORE DATA CAN HELP TO IMPROVE

Try more models for the ensembling

Support Vector Machines
Neural Networks

Include data on money habits

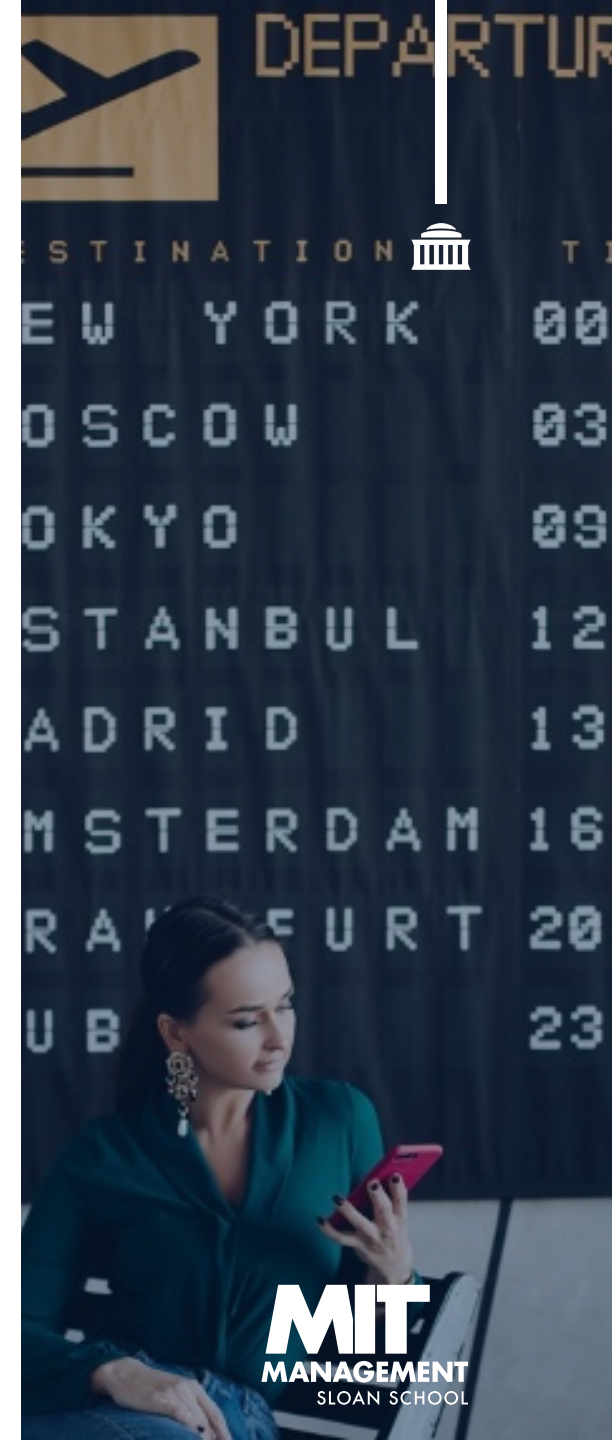Data from checking account
Past granted loans

Finetune optimization parameters

Acceptance probability
Different stages of defaults
Interest rate ranges

Test with interest rate dataset

Get a dataset with interest rates
Run predictions and model
Compare outcomes

THE MODEL MAXIMIZES TOTAL EXPECTED NET PROFIT OF THE BANK

Decision Variables

1. Acceptance of application
Binary variable

2. Assigned interest rate
Continuous variable

> **Threshold of acceptance**
> Binary variable

Objective Function

Maximize   **Total expected interest revenue** — **Total expected losses due to defaults**   (Total expected net profit)

Product of (1) interest rate, (2) amount, (3) acceptance decision, (4) assigned interest rate, and (5) 1–client acceptance probability

Product of (1) interest rate, (2) amount, (3) acceptance decision, and (4) 1–client acceptance probability