From Data to Dollars:
Investment Strategies for P2P Lending

488 Section 001 - Team Lending E

Leonard Yang, Isaac Sparrow, Theo Zhang, Jose Camacho, Nicholas Wong, Christopher Arraya

Introduction

Problem: Nearly **2 million** records with **151 features** from LendingClub.com

Solution:

Had 151 direct and indirect features in the approved dataset	Trimmed 135 irrelevant features	Kept 16 features directly related to loan and borrower's reliability

Data Features:

Quantitative Features
- loan_amnt, funded_amnt: Basic loan amount information
- term, int_rate, installment: Loan repayment structure

Indicator Features
- home_ownership: Demographics indicating financial stability
- grade, emp_length: Indicators of borrower's reliability

Credit Risk Features
- DTI, fico_range_low, fico_range_high, delinq_2yrs

Contextual Features
- Purpose: Provides the context for the loan

Data Trimming

Process: Trimmed dataset directly in Excel for computational efficiency

Data Features Dropped:

High Correlation Features

Location Features

Missing Values

Detailed Sub-Features
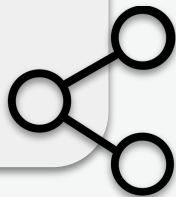Some features are sub-columns of the chosen 16 features

Improved:
*Efficiency*
Effectiveness
Clarity

Stacking

Detail: The two datasets have 4 features in common

Shared Features
- DTI (Debt-to-Income)
- purpose
- loan_amnt
- emp_length

- DTI (Debt-to-Income) = a ratio comparing debt to income
- purpose = debt consolidation, credit cards, or other
- loan_amnt = amount of the loan
- emp_length = range from less than 1 year, between 1-10 years, and more than 10 years
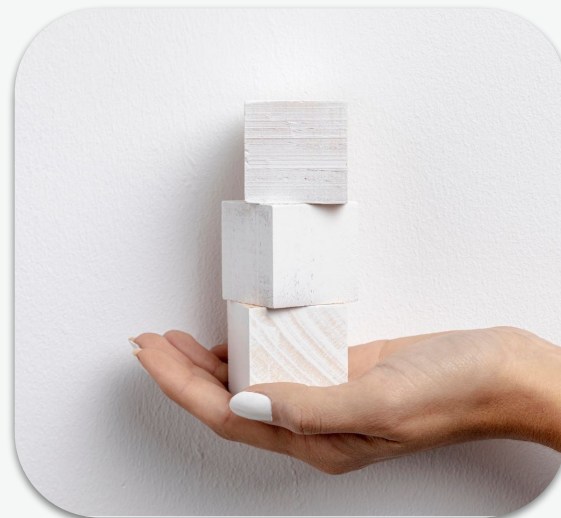
EDA - Accepted/Denied Loans

Analysis: Strong positive corr. status ↔ DTI

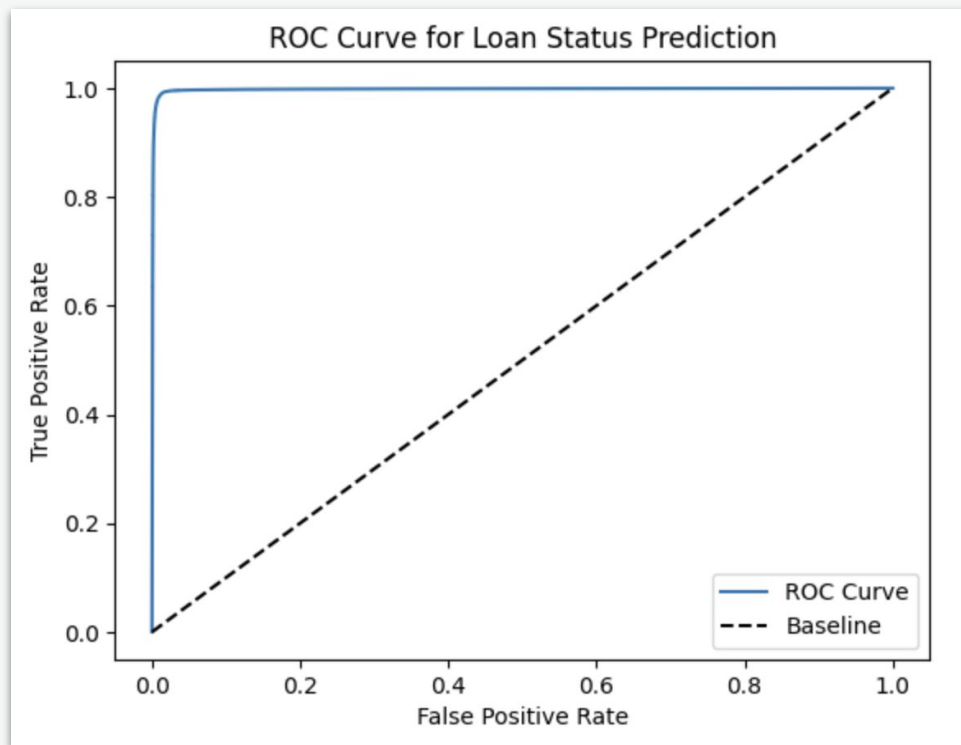**Conclusion**: DTI ratio might be a critical indicator in loan approval
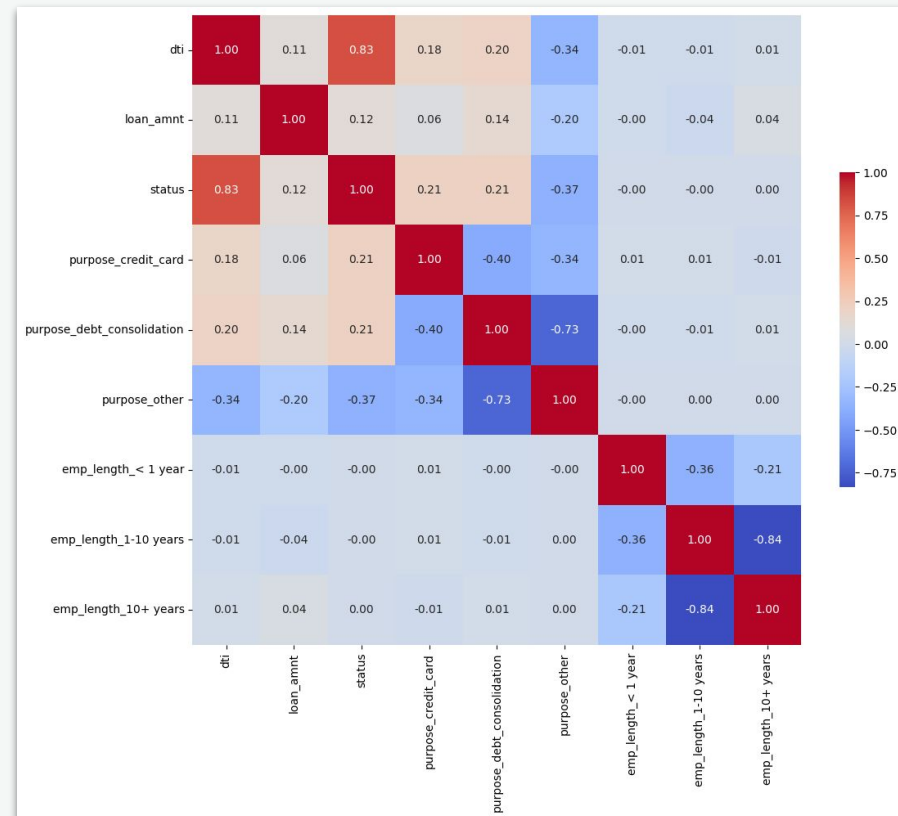
Prediction - Accepted/Denied Loans

Analysis: We ran a logistic regression to find the accuracy of our model

Result
98%
Accuracy

Explanation: The graph to the right shows the ROC curve for loan status prediction

ROC Curve for Loan Status Prediction

DTI - Most Important Feature

Analysis: We analyzed the coefficients of each feature in an descending order

Most Important Feature: Debt-to-Income Ratio (DTI)

Explanation: The Debt-to-Income Ratio (DTI) has the biggest coefficient of 15.2

	Feature	Coefficient
0	dti	15.200904
1	purpose_credit_card	0.292738
2	emp_length_< 1 year	0.161685
3	loan_amnt	0.083459
4	emp_length_1-10 years	0.002915

DTI - Adjust DTI Threshold using Stratified Strategy

Distribution of DTI by Loan Status
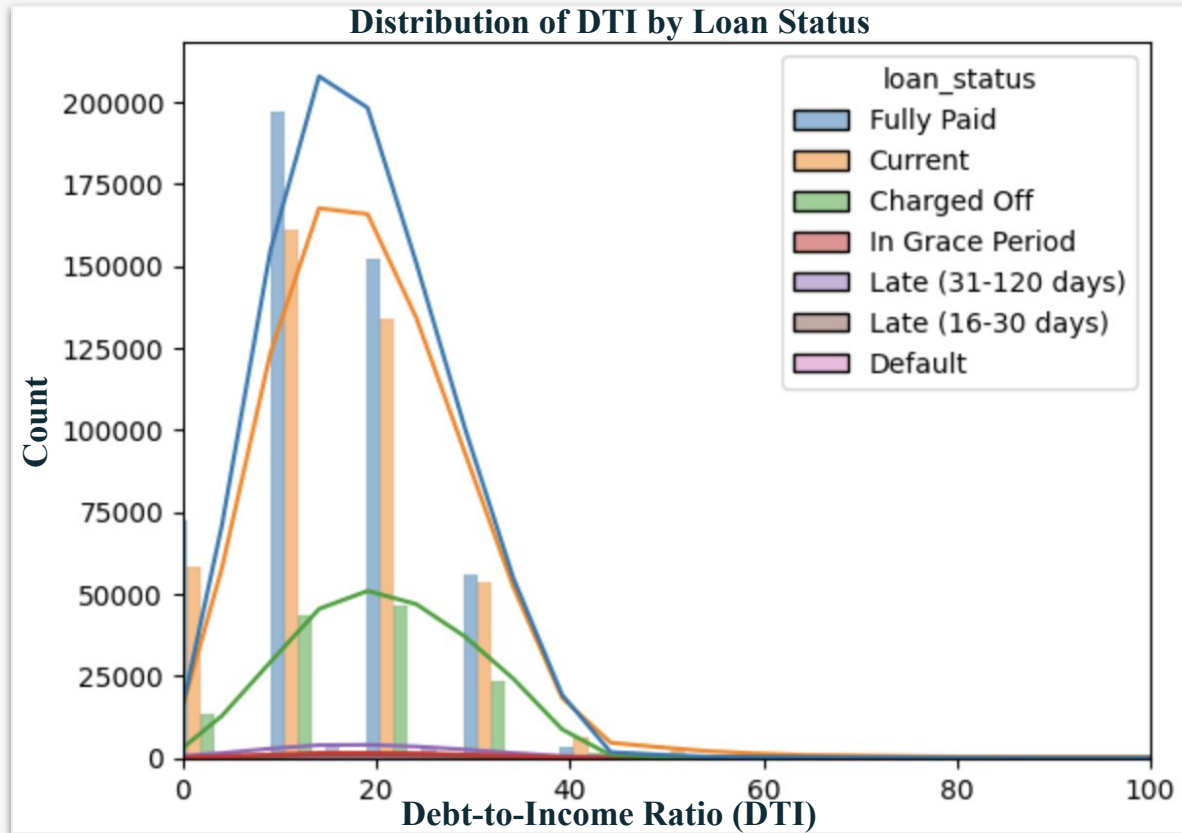
Profitability and Financial Impact

Problem:

High accuracy (**too high**) in comparing denied vs accepted data sets.

Debt-To-Income ratio was highly relied on.

Solution:

Looked at only the accepted data set

Feature Engineered ratio:

Profit Ratio = Total Payment/Funded Amount

If >=1, then the loan is desirable for investors

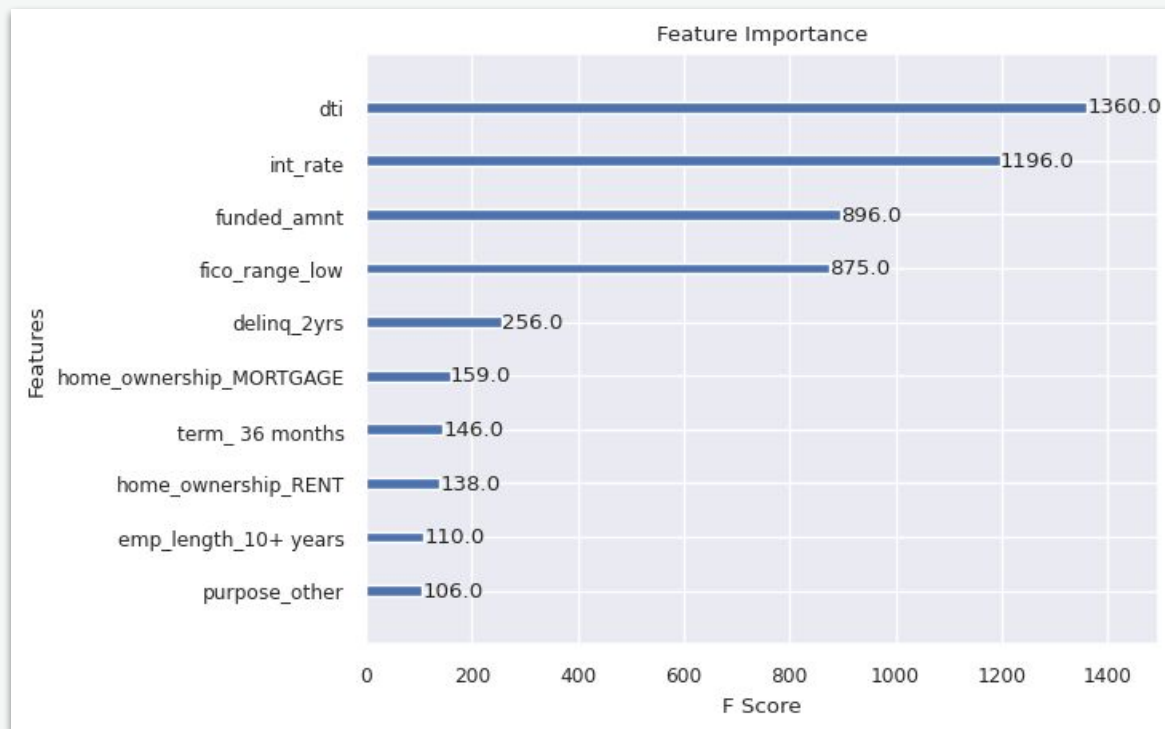If <1, then the loan lost money for investors

Comparison of Tree-based Methods for Predicting Profitability

XGBoost performed the best against:

- AdaBoost
- Random Forest
- Decision Tree

Most Important Features

1. Debt-to-Income ratio
2. Interest rate
3. FICO score

Loan Profitability Feature Selection using LASSO

Top features for increasing loan profitability:
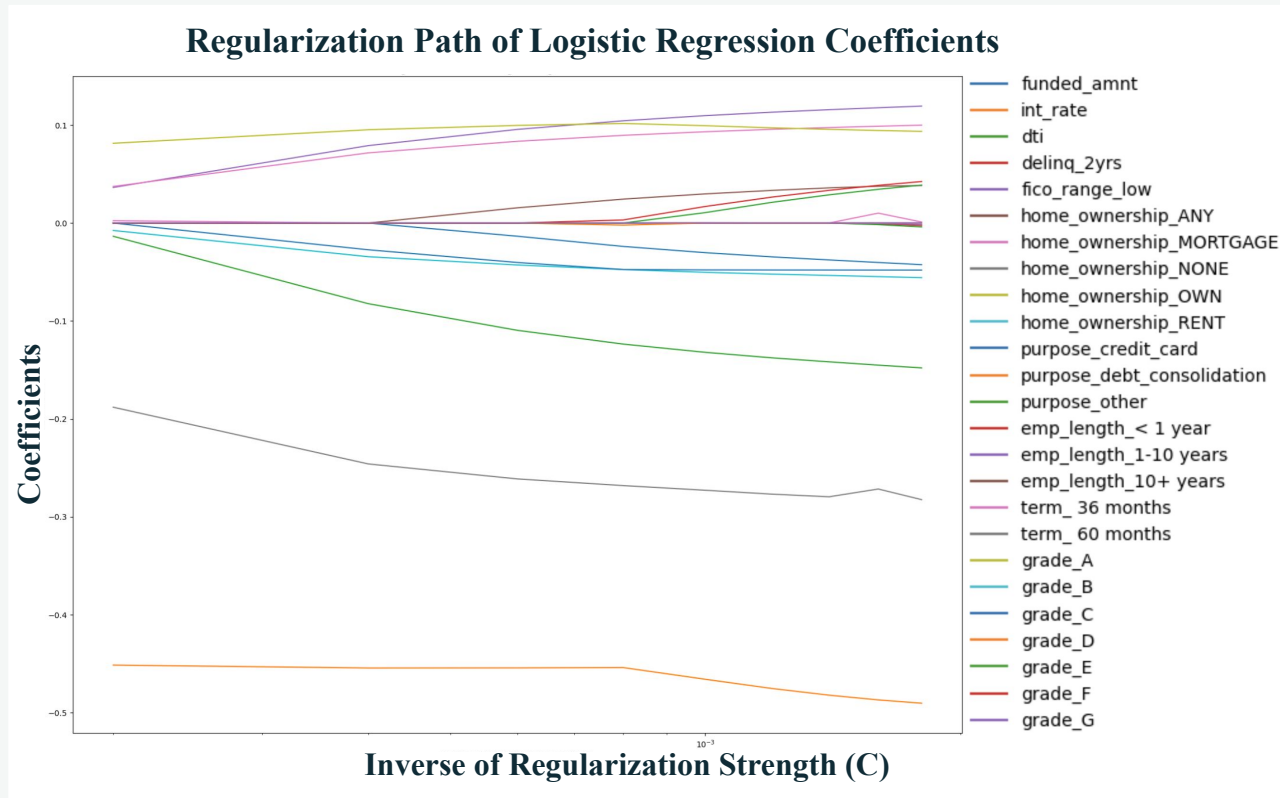*Grade_A*
Term_36months
Emp_length_1-10years

Top features reducing loan profitability:
Grade_D
Term_60months
DTI

Regularization Path of Logistic Regression Coefficients

Suggestion

Based on our analysis, we suggest Dr. D. use the **Debt-to-Income Ratio (DTI)** as the primary metric in assessing investment profitability and:

Invest in Loans that are:
Short-term (<36 months)
Grade A
Mid-career Professionals (emp_length 1-10 years)

Avoid Loans that are:
Long-term (60 months)
Grade D

THANK YOU!

Q&A