Loan interest rate predictor

Alex Umansky|Dominic Kang|Leo Liu|Viren Parmar
Team 5

Agenda

Business Overview	**01**
Data understanding	**02**
Data modeling	**03**
Potential impact	**04**

Business understanding

LendingClub is America's largest lending marketplace and first public US neobank. It leverages a full spectrum fintech platform that offers consumer and commercial loan products to every credit bracket.

LendingClub focuses on offering a transparent, efficient, and streamlined process. This translates into lower administrative costs, thus affording consumers lower APR than the average credit card.

LendingClub

Business case

According to the borrower risk profile, LendingClub provides loans with different interest rates. Incorporating data mining into the parameters that impact the interest rate offered by LendingClub can help borrowers change their parameters, leading them to save time and money and more quickly reach their goals.

Data understanding and preparation

- The full dataset contains all loans approved by LendingClub from 2007 to 2018.
  - 1.3 million entries, 150+ features

- Many features describe what happened after the loan is approved
  - Hardships, postponed payments, etc

- Many more describe the outcome of the issued loans
  - Defaults, debt settlements, etc
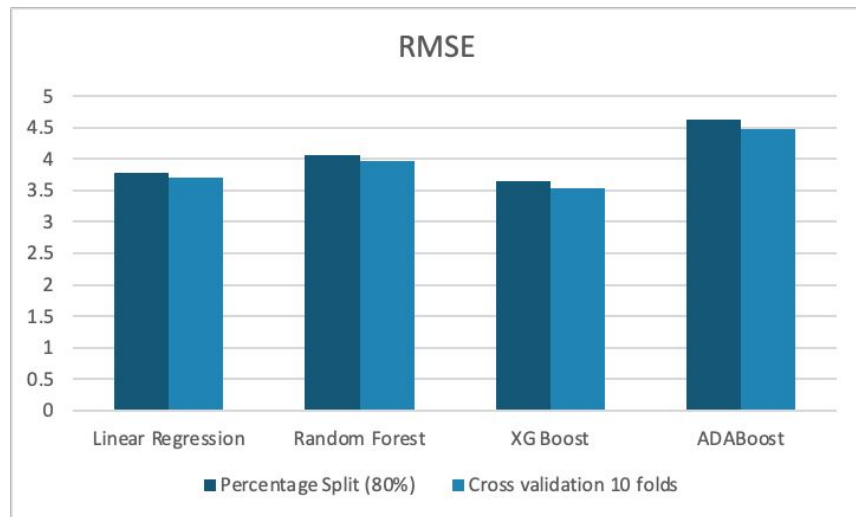
Data modeling and processing

- Not all features were helpful
  - Some are mostly null values/empty
  - Some are known only after loans are approved
- We dropped those features and checked the remaining features for data leakage
- For our use case, interest rate was the target feature
  - Loan grade was a candidate feature we considered
- We're focusing on a smaller subset of 10 thousand entries from 2018.
  - 84 total features, 164 dimensions
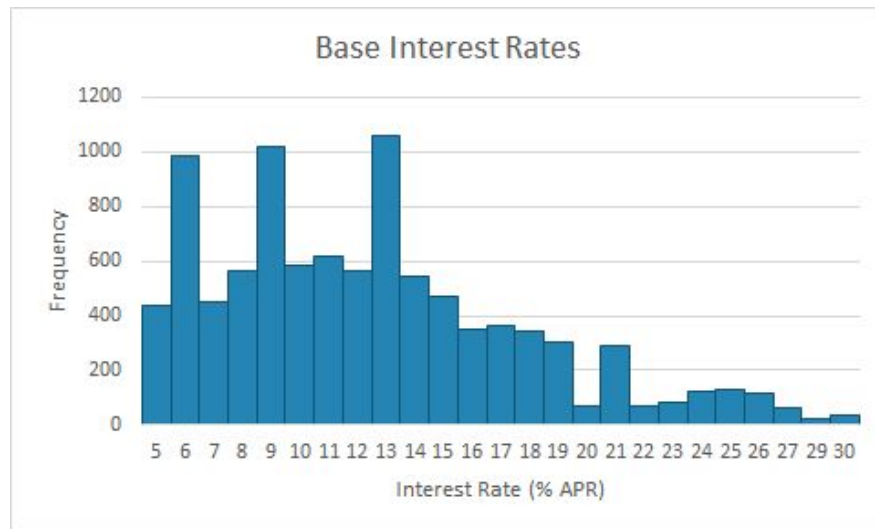
Model selection

- Interest rate is a numerical feature, hence we need to use regression models.
- For our dataset we ran 4 different types of regression models and found that XGBoost works the best.
- Our accuracy measure was Root Mean Squared Error.

RMSE

	Linear Regression	Random Forest	XG Boost	ADABoost

■ Percentage Split (80%) ■ Cross validation 10 folds

Question – which of the following loan purposes will result in the worst interest rate?

A.   Home Improvements

B.   Debt Consolidation

C.   Wedding

D.   Renewable Energy

Model Statistics

Purpose of the Loan

| Feature	Coefficients
purpose_wedding	0.092303360
purpose_vacation	0.017586830
purpose_medical	0.017370809
purpose_small_business	0.015892623
purpose_other	0.014245786

Term of the Loan

Feature	Coefficients
term_60months	0.047846943
term_36months	0.000273548

Evaluation – Impact of term on interest cost

$1,303 savings for each $10k loan

60 months, 5% interest rate:

Tax: $0.00
HOA & Insurance: $0.00
Interest: $1,322.74
Principal: $10,000.00
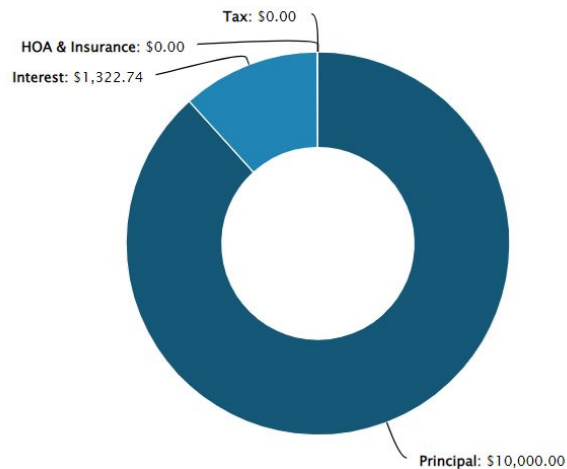
36 months, 0.247% interest rate:

Tax: $0.00
HOA & Insurance: $0.00
Interest: $19.59
Principal: $10,000.00
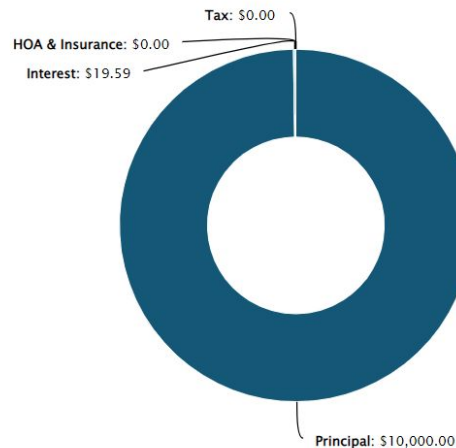
Potential impact?

- Incorporating this strategy of data mining could allow for lower interest rates for clientele
  - The impacts of this are noticeable/impactful in that people who are most in need of these loans could potentially get lowered interest rates
- For LendingClub, there could potentially be an impact on future interest rate calculations

Lessons learned

- We should have started earlier.
- Finding the data took the most amount of time and caused us to pivot our business model multiple times.
- Data cleaning was a challenge. Especially due to finding the balance of the right amount of instances and features.
- Checking for leakage is difficult.
- Using WEKA was challenging and we had to pivot to Python.

Thank you!

Q&A

Resources

- https://drive.google.com/file/d/16VmycE0fboCTmq3P2LCYKZGvtlUIr2rZ/view?usp=sharing
- https://www.kaggle.com/datasets/wordsforthewise/lending-club