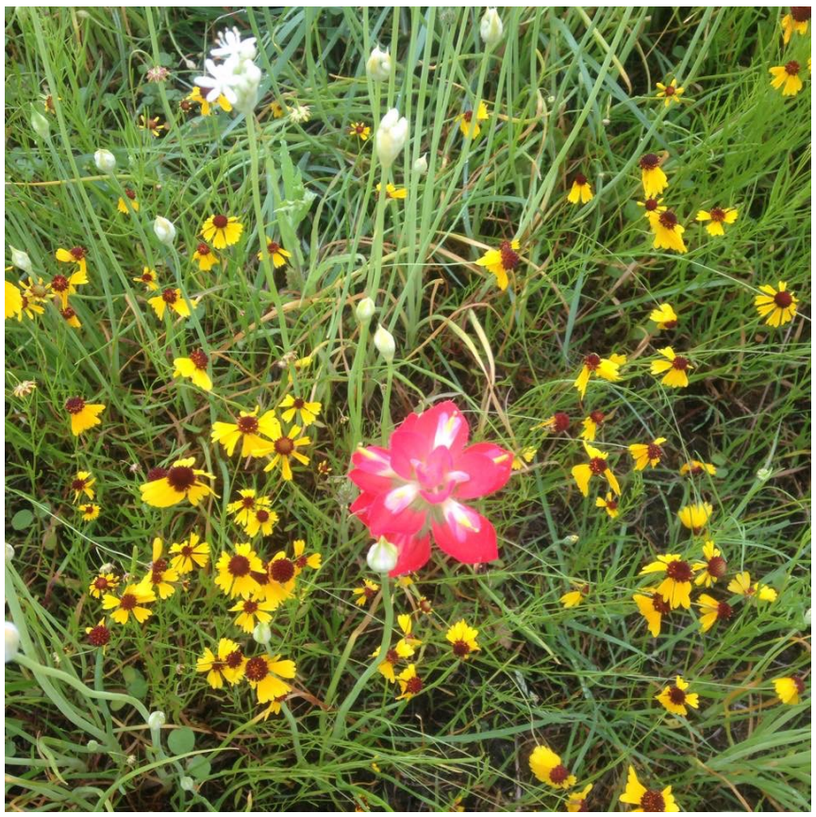
Loan Default Risk Prediction



Online Lending

- Establishing a lending platform is easier than ever
- Data for risk prediction is plentiful
- Need for methods that can accurately predict default risk

The Lending Club

- Peer-to-peer lending service
- Loan status and features available
- Data on nearly one million loans

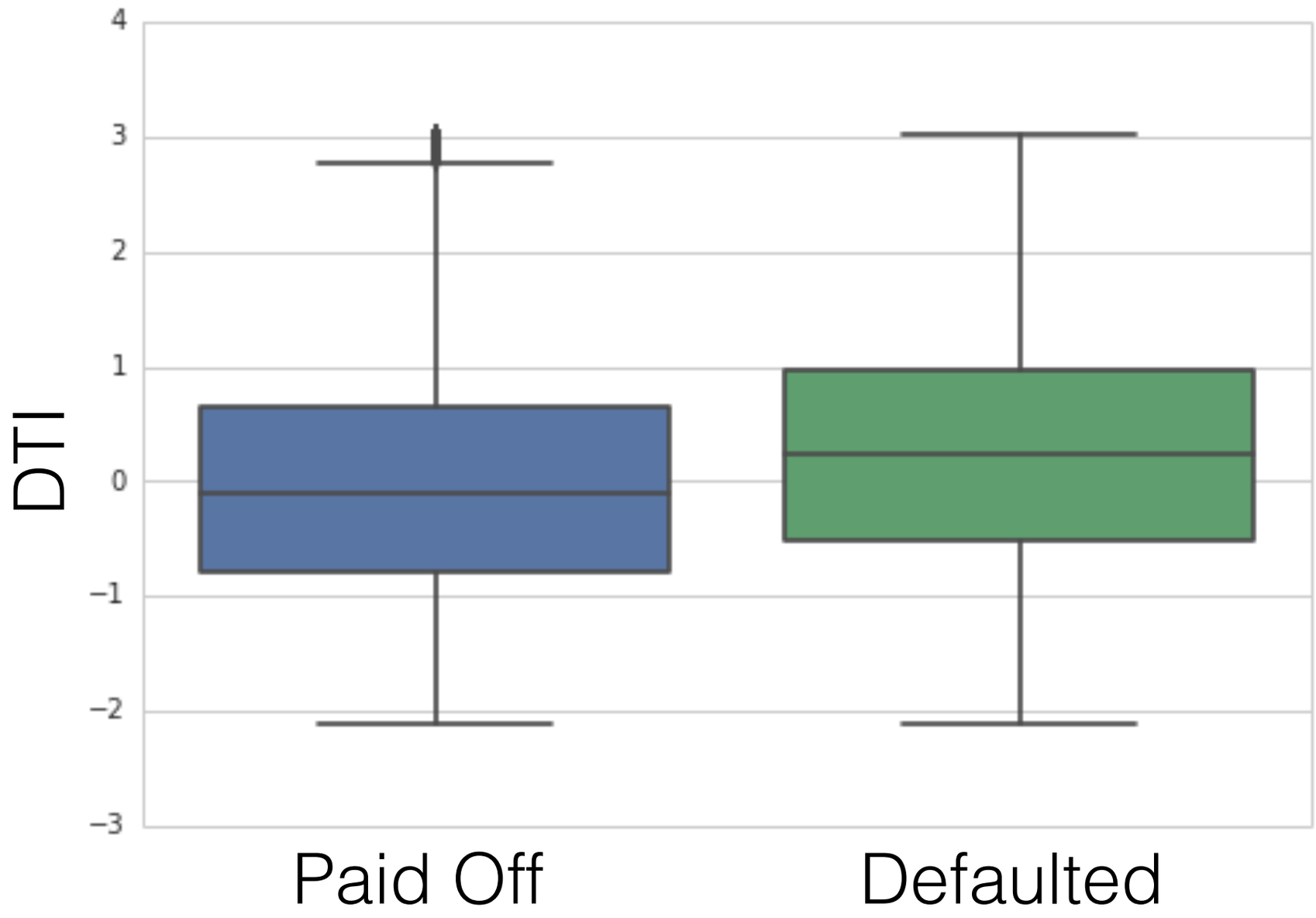
Goals

- Develop risk model
- Go from borrower features to default probability
- Assess resulting loan profitability

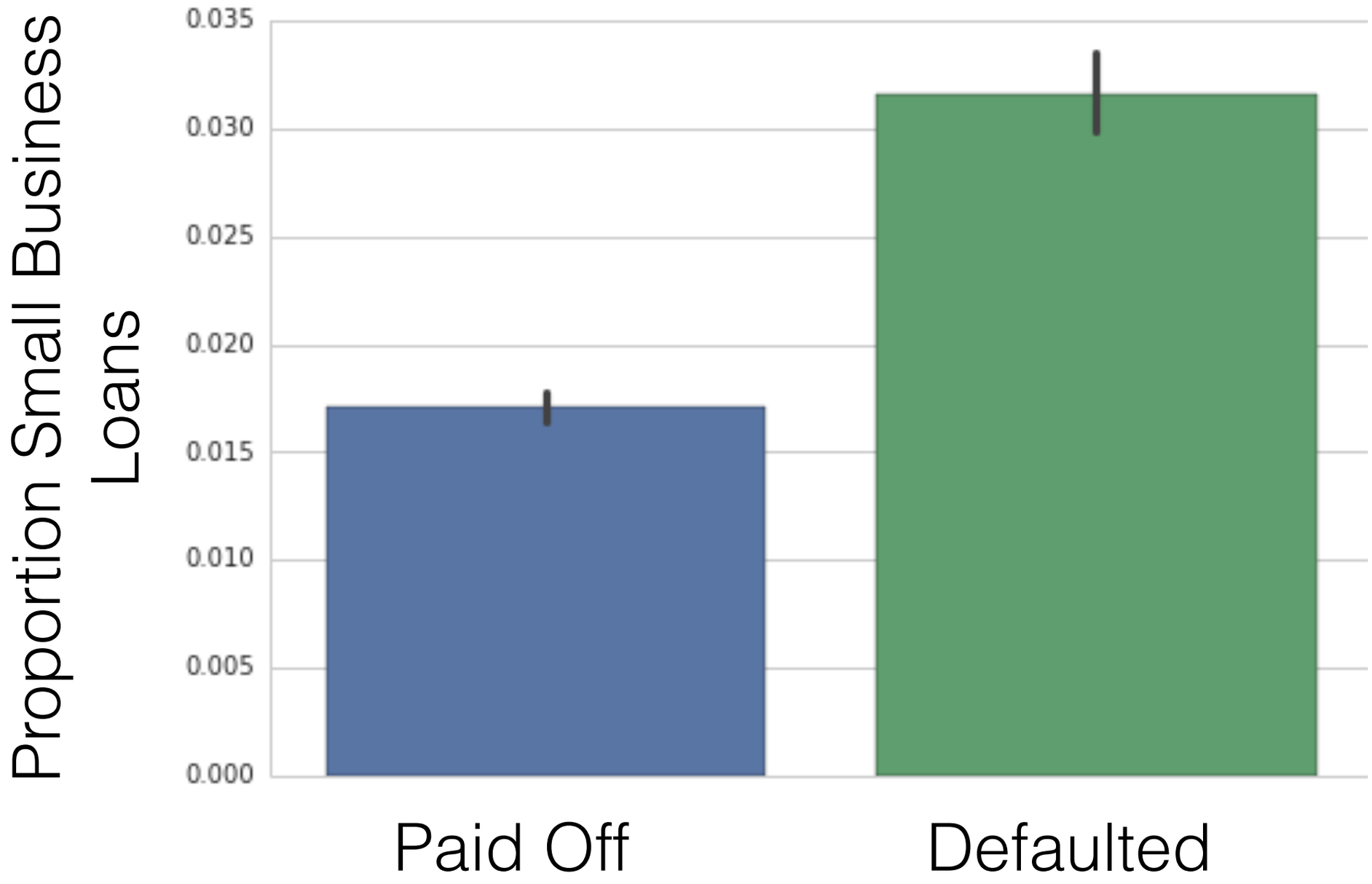
What Matters?

- Borrower financial status and habits
- Borrower's purpose
- Borrower's living situation

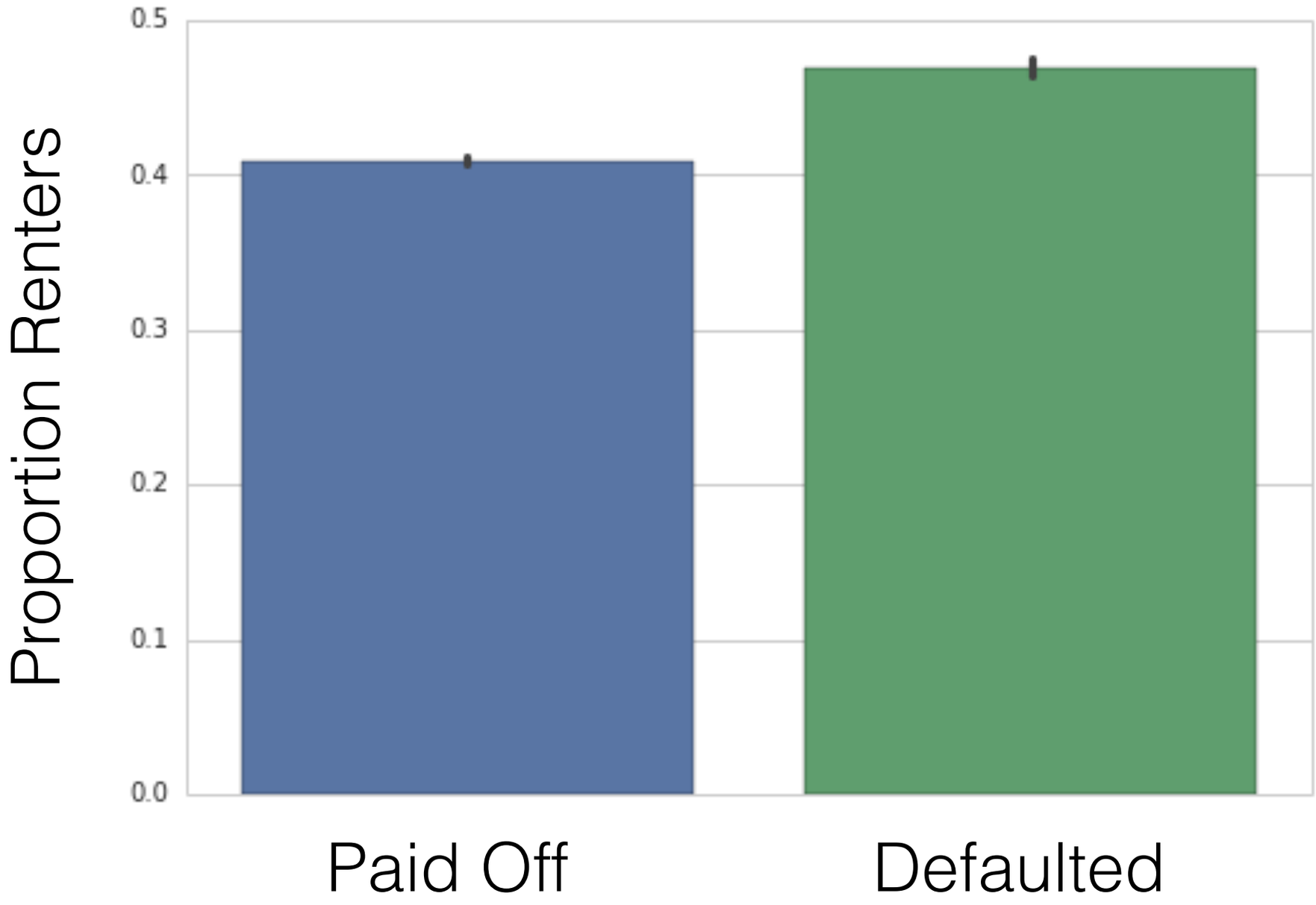
Example: Debt-to-Income



Example: Business Loan



Example: Renters



Examples

- Higher debt-to-income ratio, more likely to default
- Business loan, more likely to default
- Renter, more likely to default

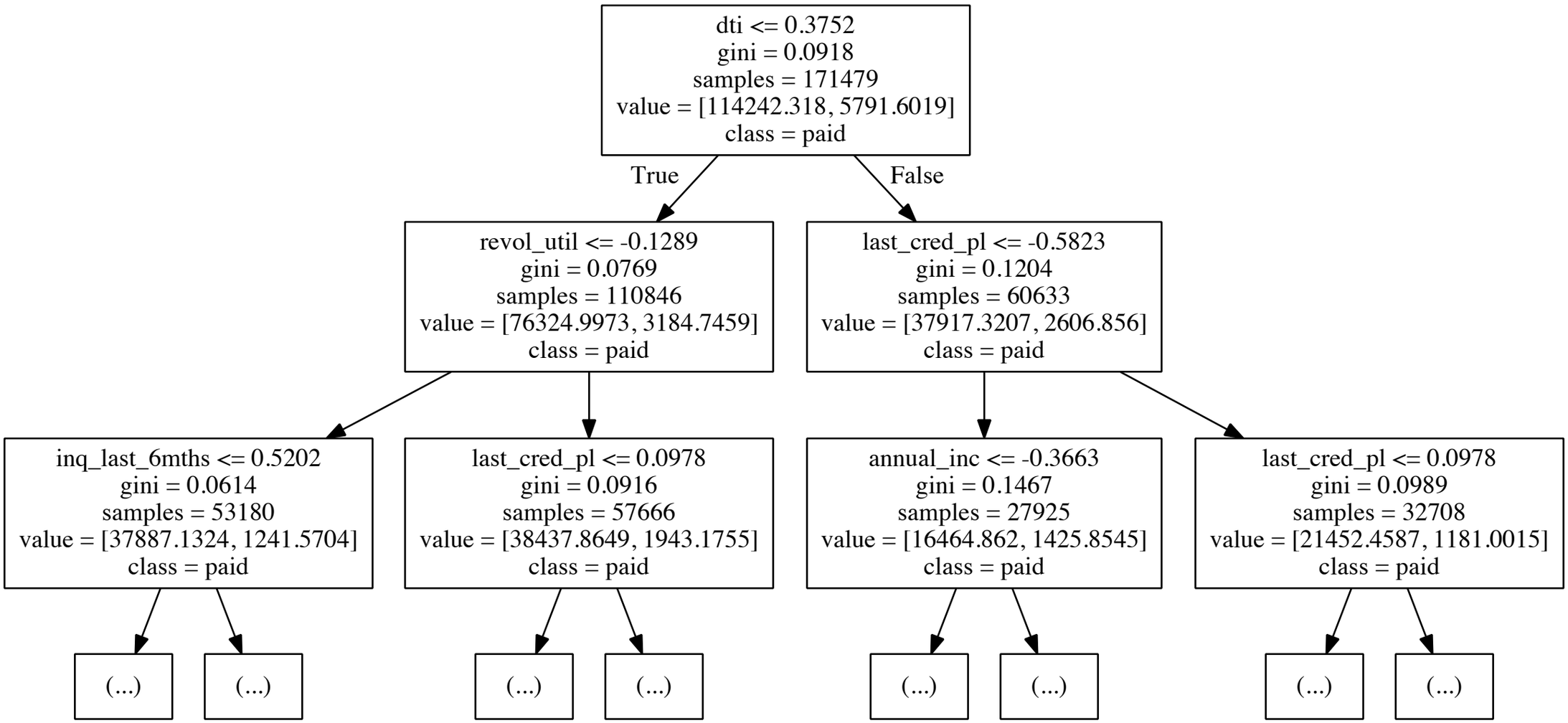
Where to go from here?

- Many features weakly predict loan default

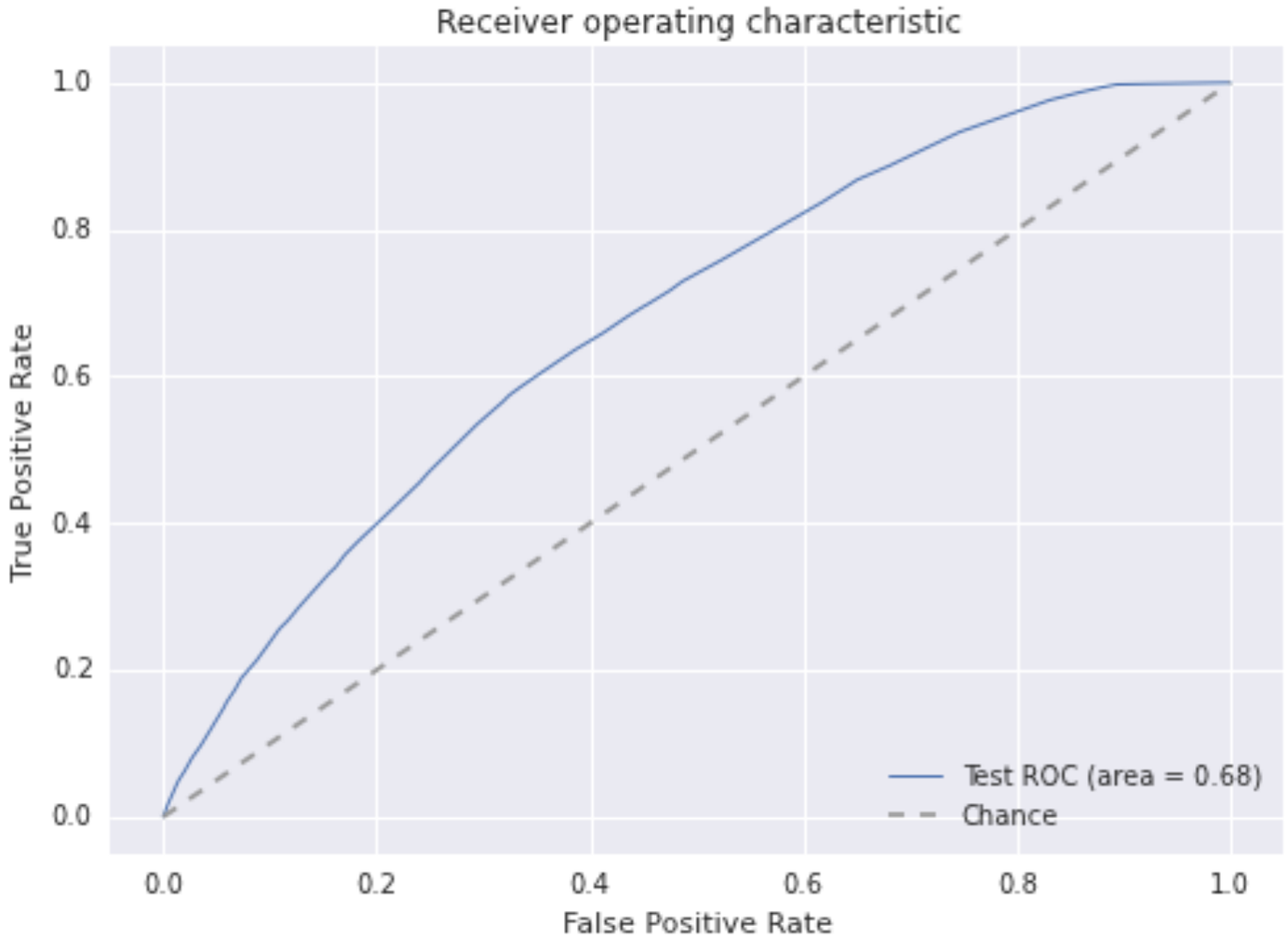
- Some relationships are simple (i.e., linear)

- Some aren't (i.e., nonlinear)

Tree-based Methods



Tree Performance



- ROC curve shows relationship between predicting real defaults and predicting default for loans that are paid off

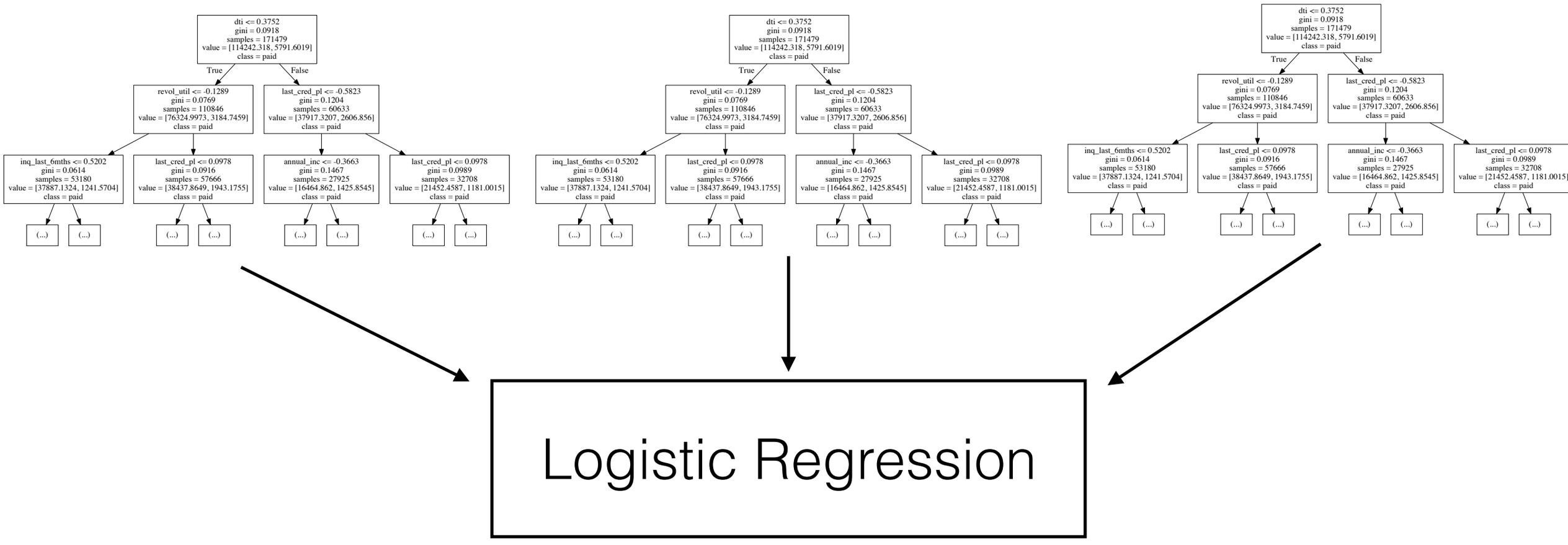
- Area under the curve shows that the tree can accurately predict default

What's it worth?

If the 15% of riskiest loans are rejected, ROI = $1052.04

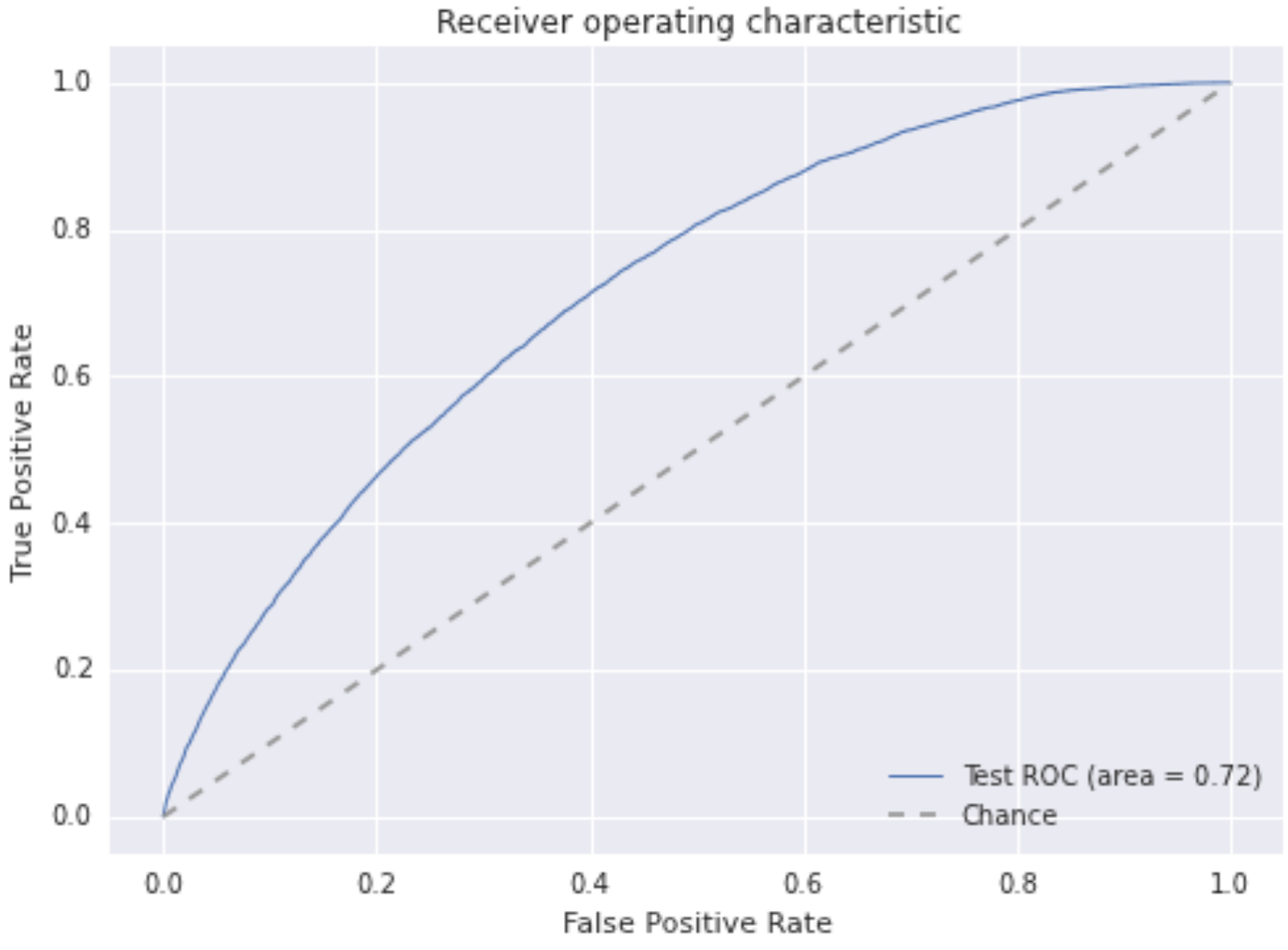
Can we do better?

Ensemble Methods



Combine estimates from multiple tree-based classifiers in order to improve performance

Ensemble Performance



- Area under ROC improves from 0.68 to 0.72 with ensemble
- Ensemble classifier is stronger than individual tree-based classifiers

What's it worth?

If the 15% of riskiest loans are rejected, ROI = $1059.95

Results

- For a $10,000 investment, tree-based ROI = $1052.04, ensemble ROI = $1059.95
- Tree-based ROC area = 0.68, ensemble ROC area = 0.72
- The ensemble is the stronger and more profitable classifier

Recommendations

- Use ensemble model as initial risk assessment model for setting interest rates and loan amounts

- Improve ensemble performance by adding credit scores and other financial info

- Use ROI as a benchmark for evaluating how the ensemble performs

Thanks!