

Presented by: Whitney Bullock, Caleb Call, Thomas Beck

HOME CREDIT ANALYSIS: PREDICTING MINORITY DEFAULTING CLASS

Logistic Regression, Decision Tree, Random Forest, Gradient Boost



Research Team



Whitney Bullock

Whitney is an MSBA candidate with over three years of experience in B2C sales and account management. She is focusing on using analytics to optimize customer acquisition and enhance client relationship management.



Caleb Call

Caleb is an MSBA candidate with experience in the financial industry, specializing in client relations and data reporting. His key proficiencies include analytical problem-solving, clear communication, data analysis, and a focus on achieving measurable goals.



Thomas Beck

Thomas is an MSBA candidate with five years of experience in finance, M&A, and operations analysis across healthcare and technology verticals. He specializes in financial analysis and process optimization, aiming to bridge finance and data science through strategic forecasting.





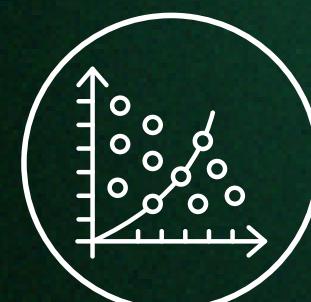
Table of Contents



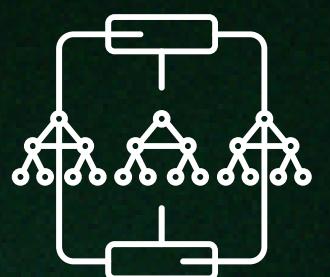
Business Problem



Exploratory Data Analysis



Logistic Model



Random Forest Model



Gradient Boosted Model



Conclusion



Business Problem

HOME CREDIT FINANCE

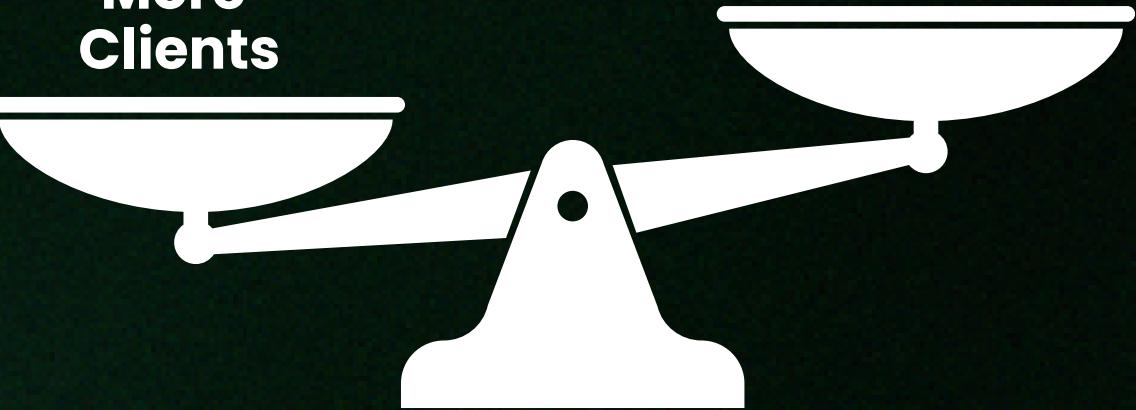
<https://www.homecredit.ph/>

Estimate your loan in seconds,
get fast loans with easy
installments



Approve
More
Clients

Reduce
Default
Risk



The Company: Home Credit is a **global lender**. Their mission is to serve "**thin-file**" applicants—clients who are *invisible* to traditional banks because they lack a credit history

The Kaggle Problem: The company provided a real-world dataset of alternative data (instead of FICO scores), which we must use to build a model that accurately predicts default.

Business Problem: The Business Value of a Solution

A successful model directly solves the "two-sided risk" by providing a reliable risk score for every applicant.



Benefit of a Solution

- Safely approve more qualified, "thin-file" applicants.
- This expands market share and fulfills the core mission of financial inclusion.
- Accurately identify high-risk applicants before approval.
- This reduces default rates, minimizes financial losses, and creates a more stable, profitable business.

Objective

Accurately classify
target customers

Measure of Success

Targeting Model's AUC
above 0.75



Data Provided



Application data



Credit history

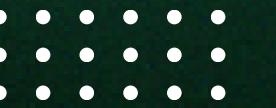


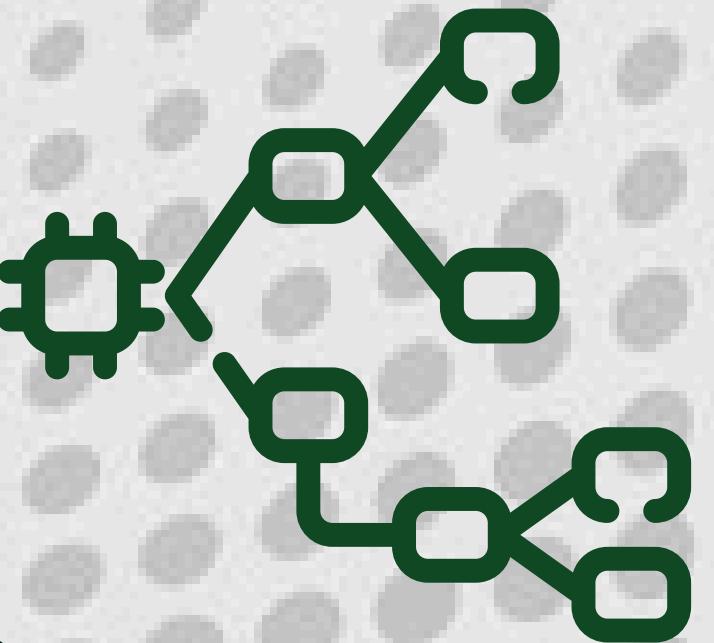
Alternative data sources

Exploratory Data Analysis

Strong predictors of repayment

- ✓ External Credit Score
- ✓ Age
- ✓ Employment duration
- ✗ Gender
- ✗ Total income





Analysis Models Performed

Logistic
Regression

Decision
Tree

Random
Forest

Gradient
Boost

Logistic Regression

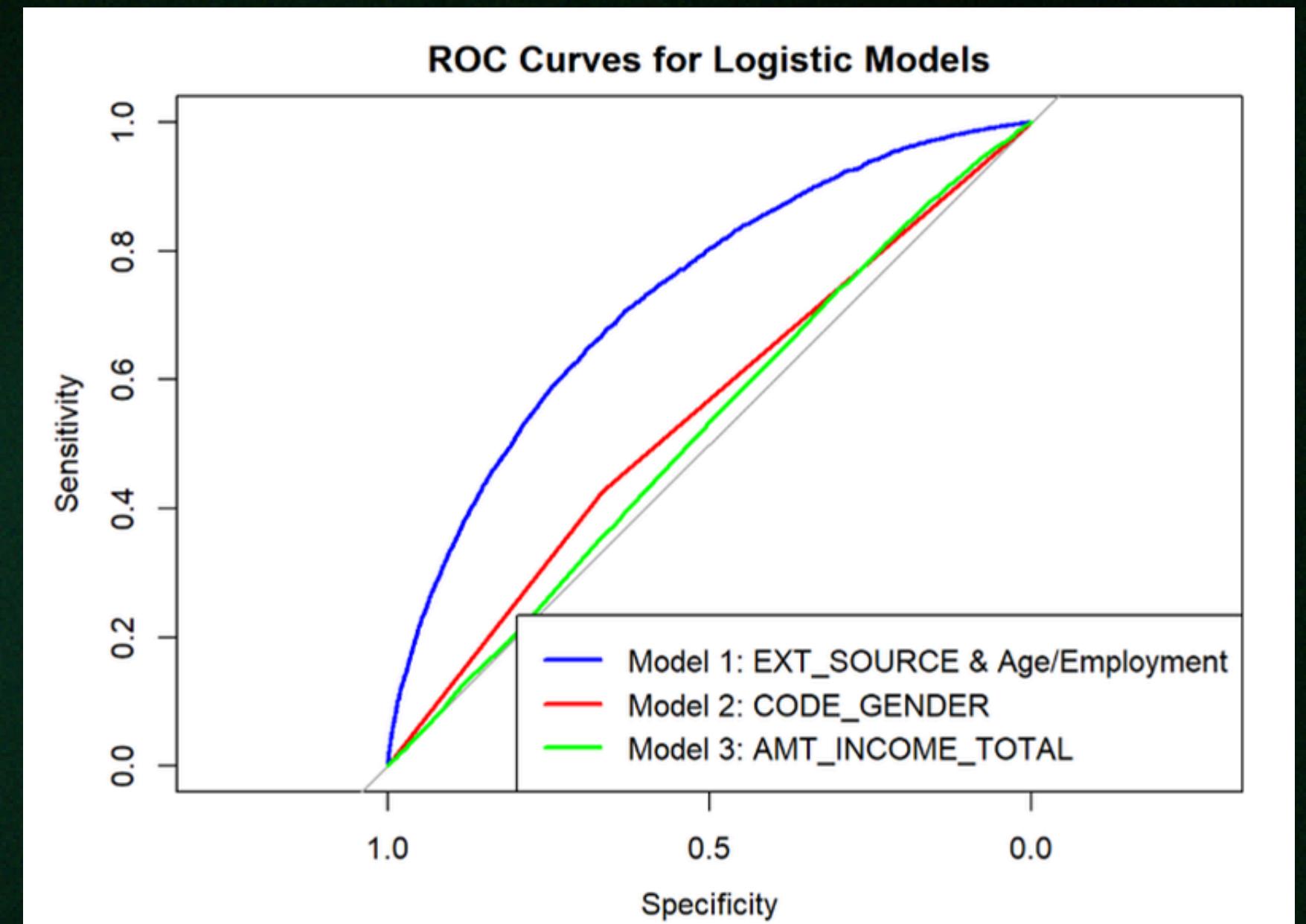
AUC = .729

Combining credit score, age,
and employment length

Gender alone = .54

Income = .52

barely better than 50/50 guess

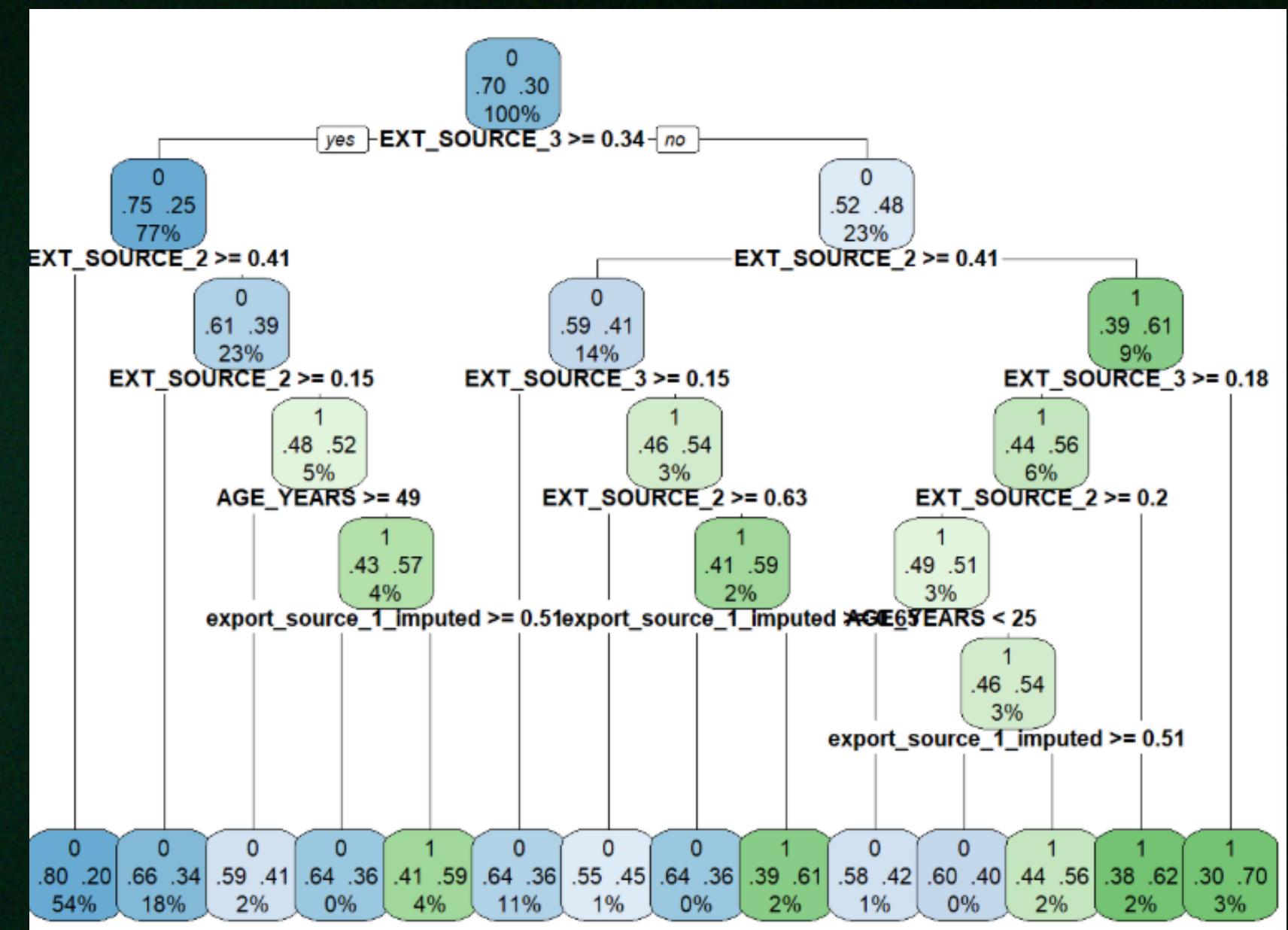


Weighted Decision Tree

AUC = .685

Key characteristics to identify high-risk segments

Automated rejections to speed rejection/approval process

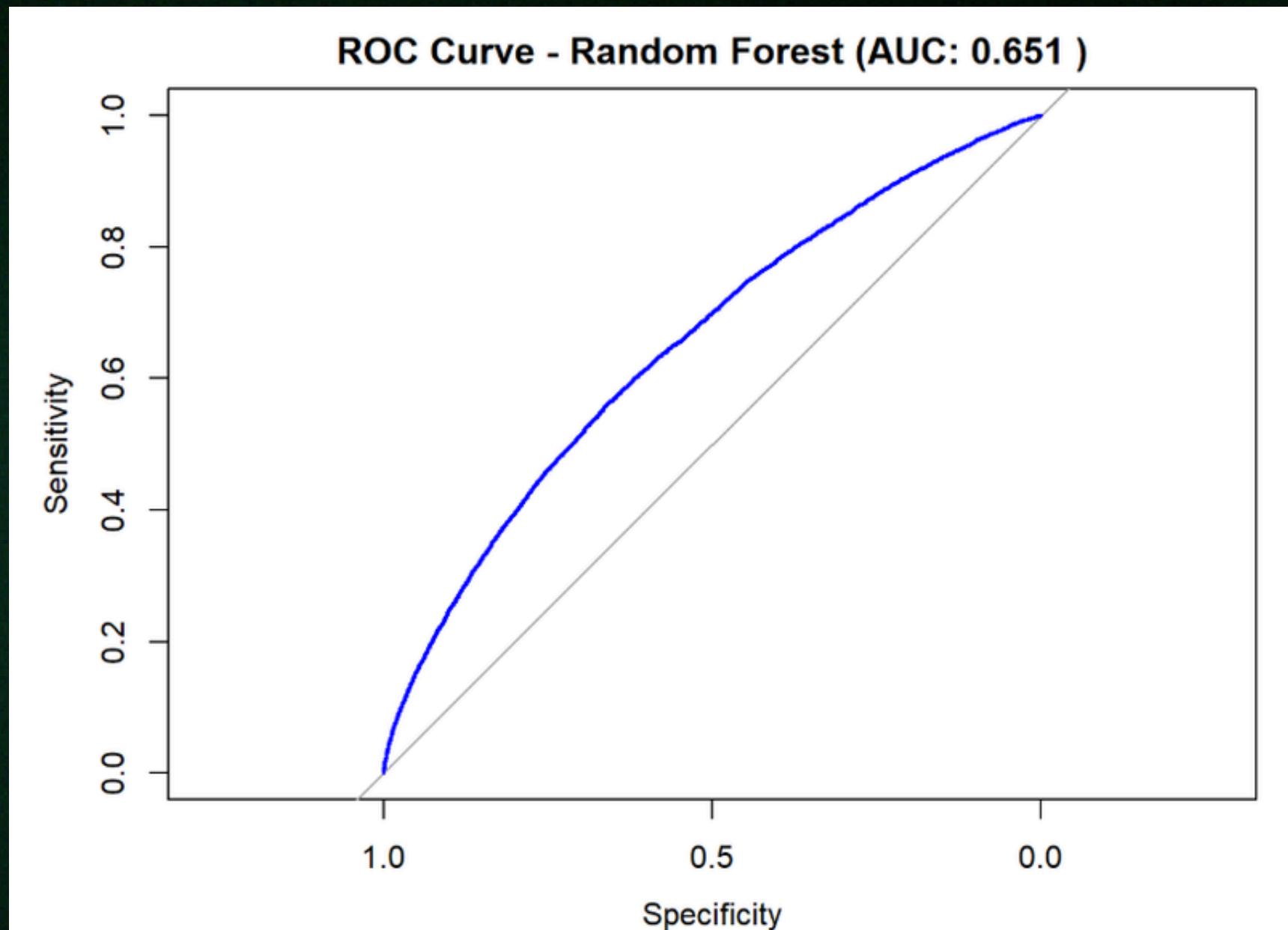
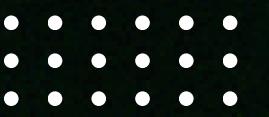


Random Forest Model

AUC = .651

Decreased AUC from Random Tree

Difficult to interpret because tree predictions are averaged from 500 trees

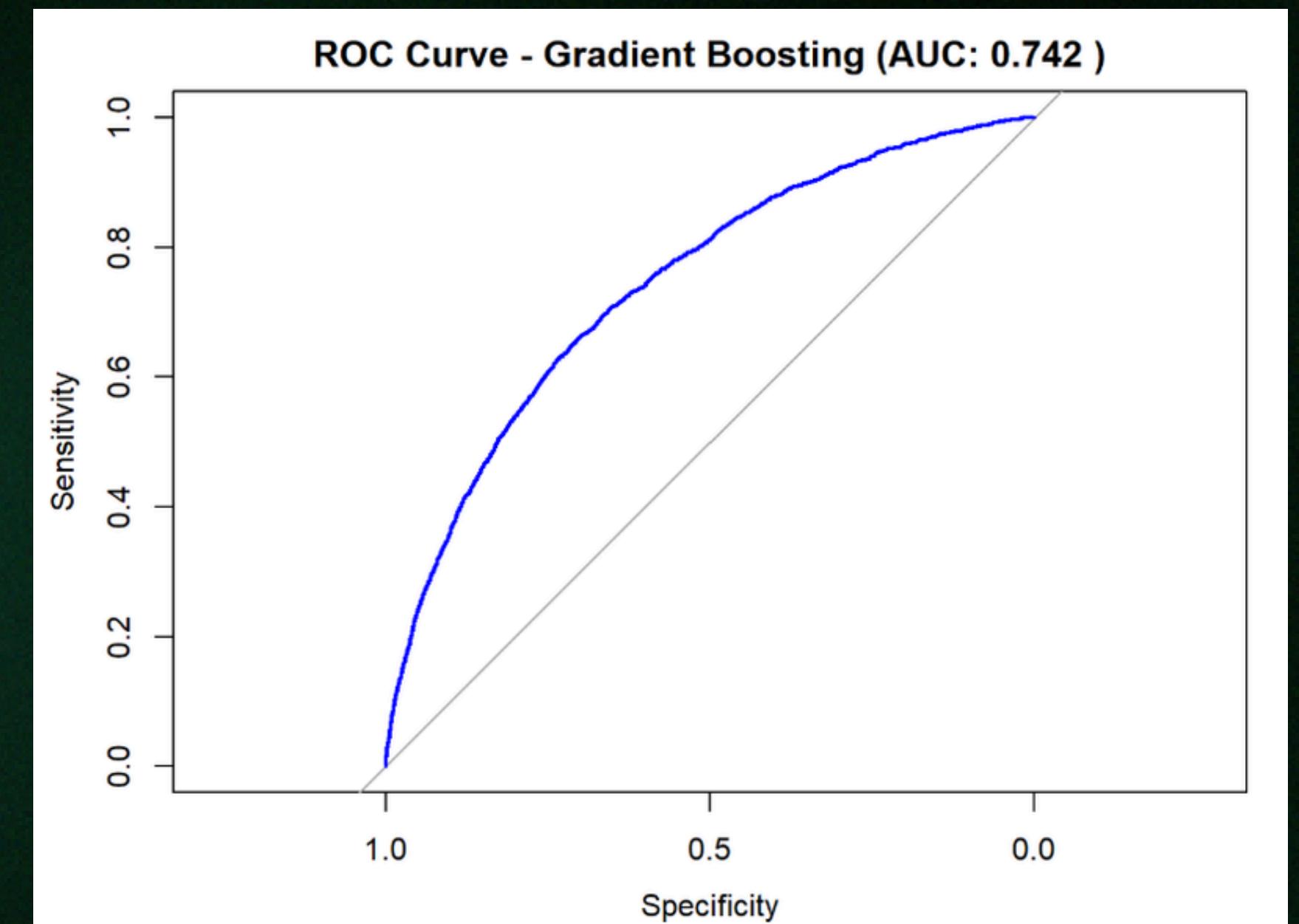
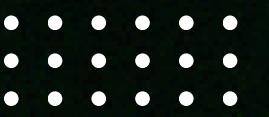


Gradient Boosted

AUC = .742

Large improvement from
other models

Model corrects for errors
made as it runs unlike random
forrest



Comparision



Gradient Boosted Model
performs the best

Improved upon Random
Forrest accounting for errors



Model	Test AUC	Notes
Gradient Boosting (XGBoost)	0.742	Early stopping on AUC selected best iteration; ROC shown in submission.
Logistic Regression (EXT_SOURCEs + Age + Employment)	0.73	Best logistic of the three tested; single-feature gender and income models were near-chance.
Decision Tree (weighted)	0.685	Weighted for class imbalance; ROC reported.
Random Forest	0.651	500 trees; underperformed, likely due to lack of class weighting in that run.

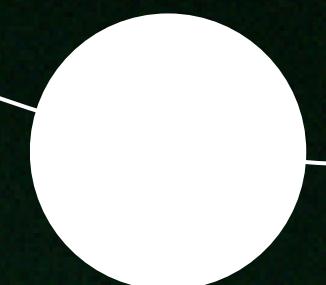
JANE DEFAULT PREDICTION

Married Woman with 2 children

\$315,000 Income

\$1,575,000 credit amount of the loan

Predicted .387 probability of default



Business Impact



- Median credit amount is \$450,000
- Default rate is about 8%
- For every person we expect defaults to cost
 $\$450K * .08 = \$36K$
- Assume 100,000 applicants a year = \$3.6 Billion in losses



Implementation and Monitoring

- Using the model will help to reduce giving loans to bad clients and reduce millions of dollars of default money
- Compare default rate and total revenue monthly to see if the company is outperforming past years



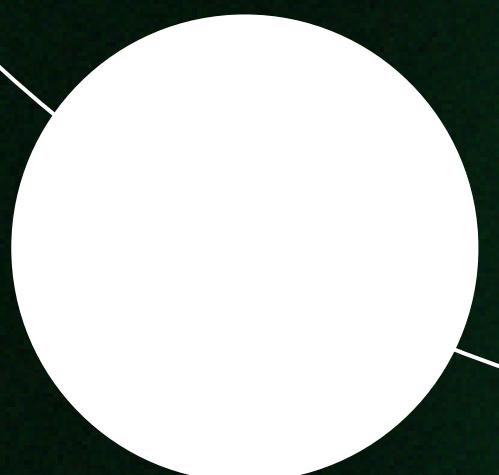
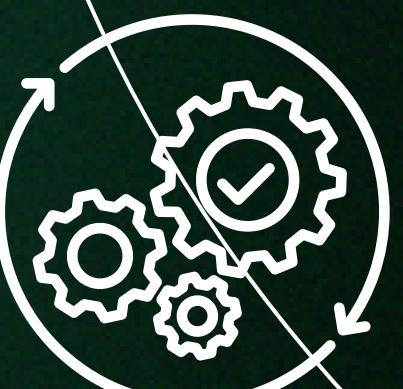
Reduce Loan Defaulters



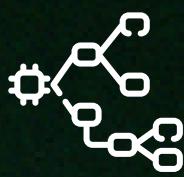
Improves Profitability



Operational Efficiencies



Business Impact



Use decision tree models help to identify segmenting characteristics

The gradient model provided meaningful predictive power for the minority class



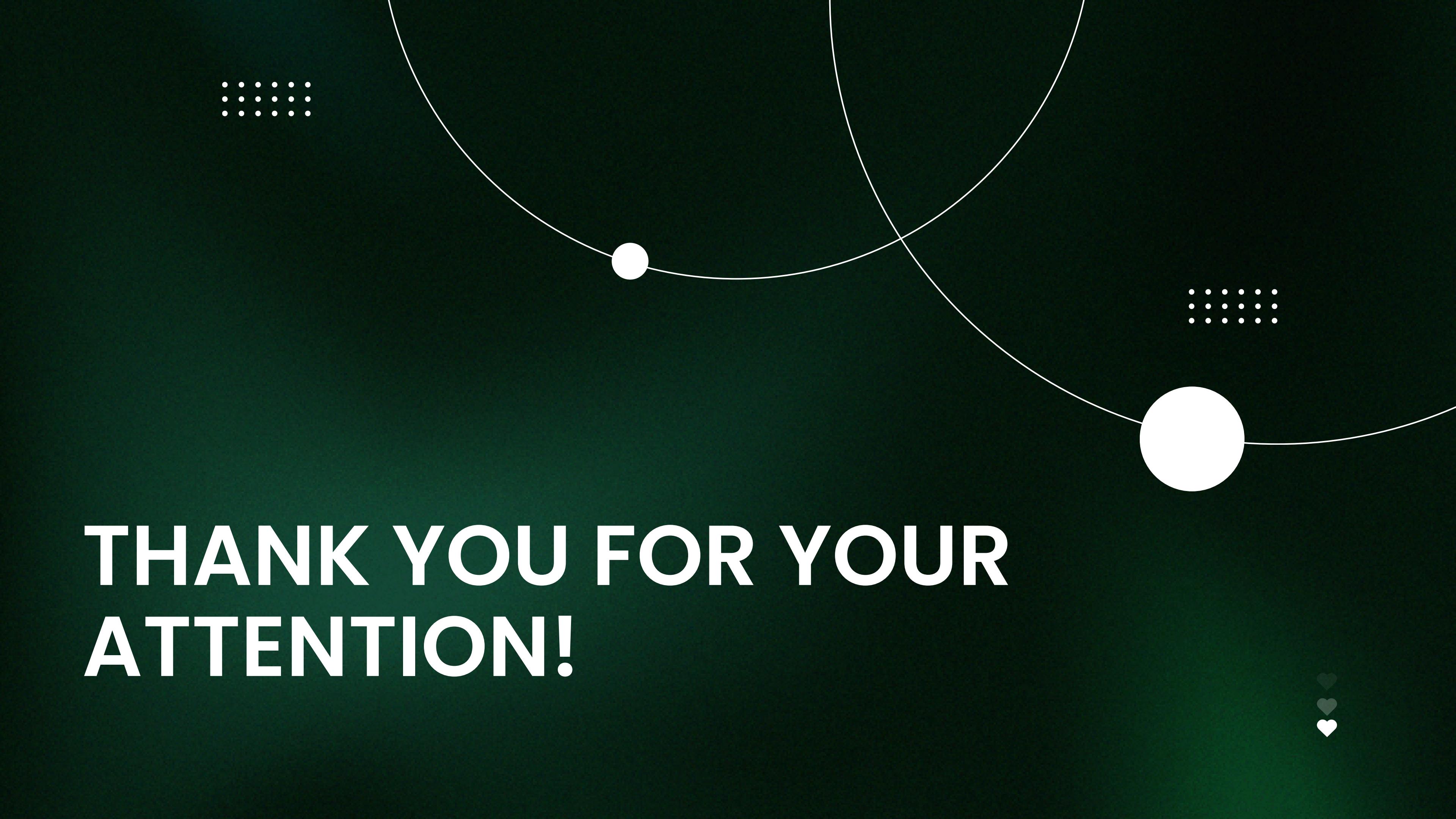
More is not always better, don't need to collect 100+ variables of data



AUC from Kaggle = .70



Conclusion



A dark green background featuring abstract white geometric elements. At the top center is a large circle containing a smaller circle. A horizontal line segment connects the centers of these two circles. Above this line is a row of five small white dots. Below it is a row of seven small white dots. To the right of the bottom row is a single large white circle. To the left of the top row is a small white circle. In the bottom right corner, there is a vertical cluster of three small white heart-shaped dots.

THANK YOU FOR YOUR
ATTENTION!