The background is a dark blue gradient. On the left, there are two overlapping geometric shapes: a blue parallelogram and a light green parallelogram. In the bottom left corner, there is a circular inset showing a close-up of a circuit board. In the top right corner, there is a faint, stylized pattern of interconnected lines and nodes, resembling a neural network or data flow diagram.

Machine Learning for Banking: Loan Approval Use Case

Agne Subelyte

BSc Economics

MSc Mathematical Economics and Econometric Methods

Bank for International Settlements

Data source: LendingClub.com

Largest fintech loan provider in the United States

2 million approved and 28 million rejected applications since 2007

4 borrower features:

- Fico score
- Loan amount
- Debt-to-income ratio
- Employment length



Challenge

Processing and fitting machine learning models on 30 million loan applications

Solution

Amazon's Elastic Compute Cloud (EC2)

Spark on Amazon's Elastic MapReduce (EMR)

10% training sample, 90% testing sample



Research scope:

1. Modeling interest rate:

- User age not available
- Predictions not accurate enough

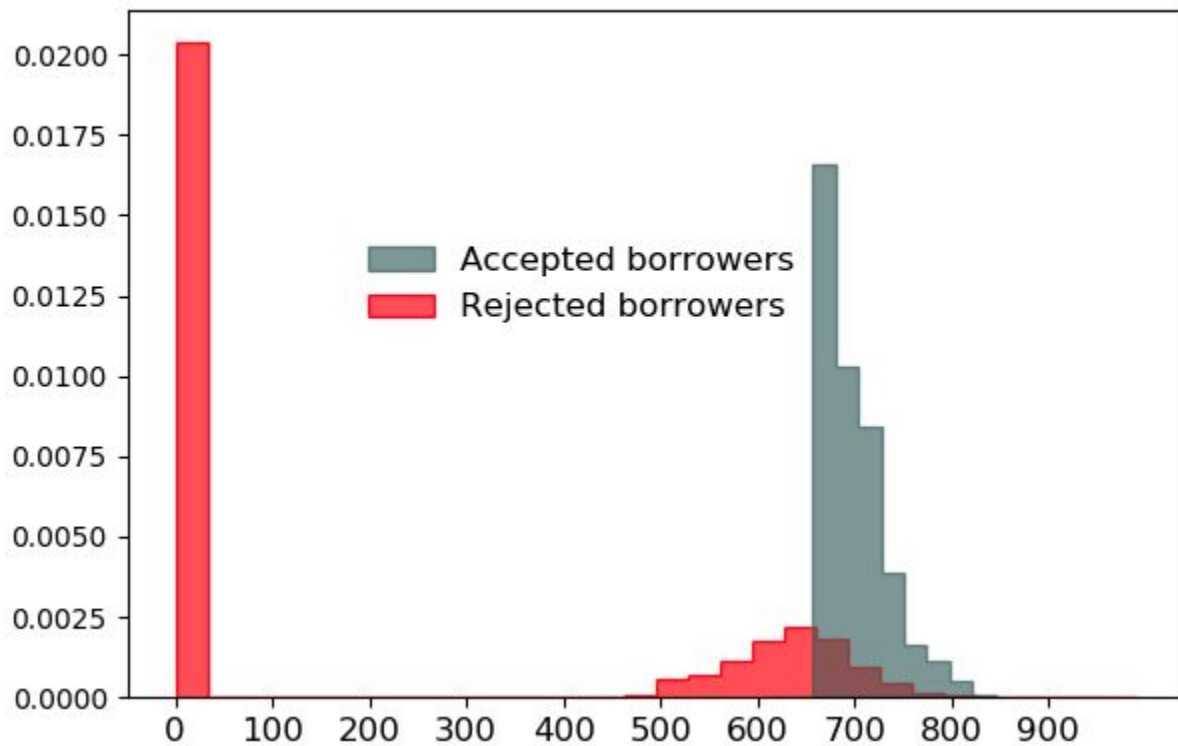
2. Modeling the likelihood of default:

- Requires a lot of user input (30+ features), low R^2
- Not suitable for an actual app

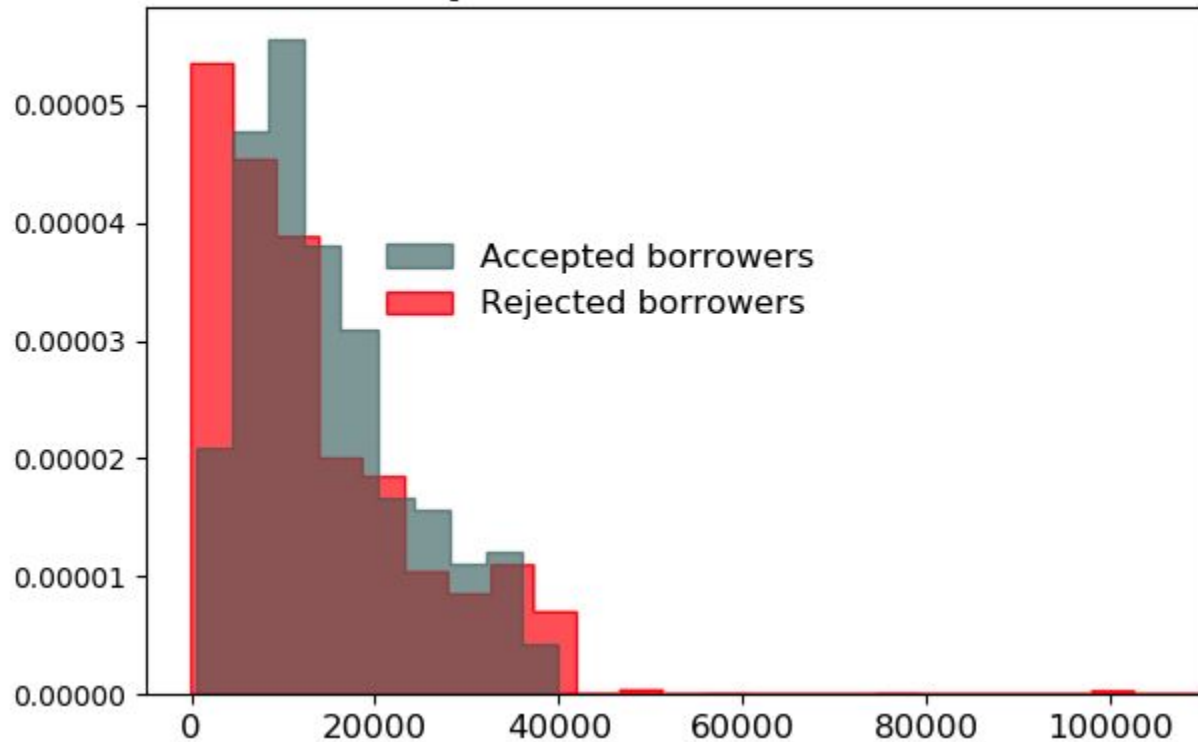
3. Modeling loan approval:

- High accuracy score with just 4 features
- Suitable for an easy, fun and accurate application

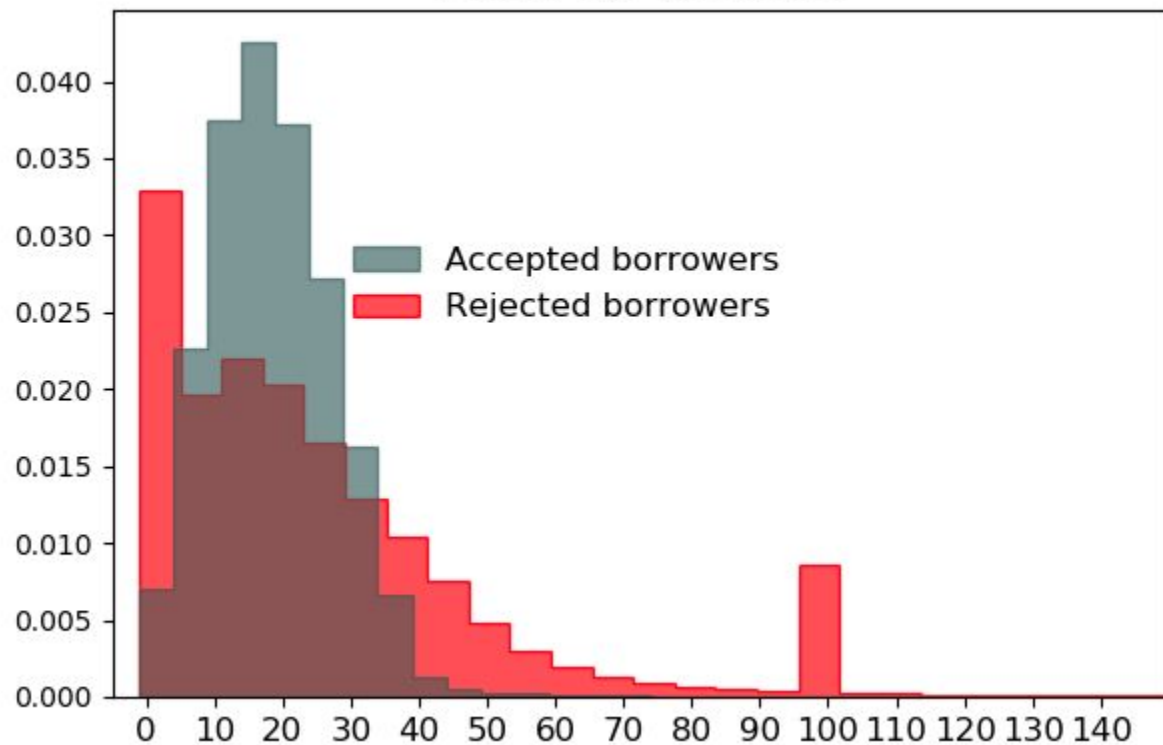
Fico Score



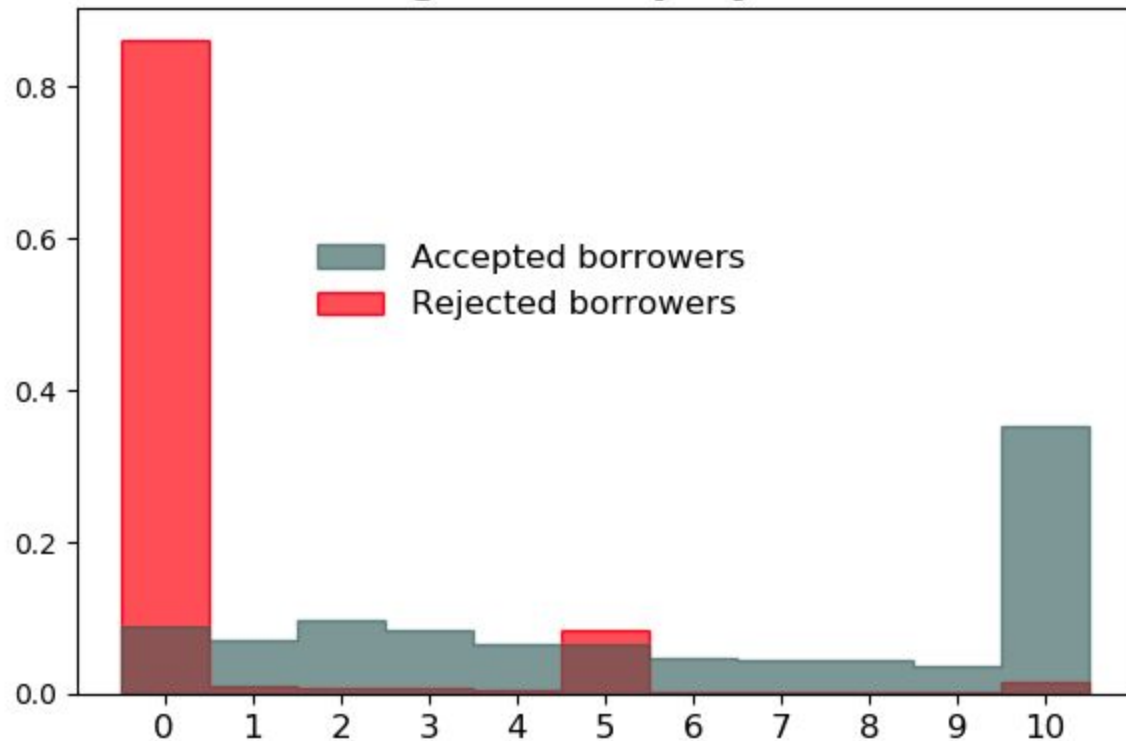
Requested Loan Amount



Debt-To-Income

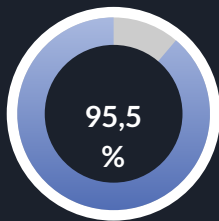


Length of Employment



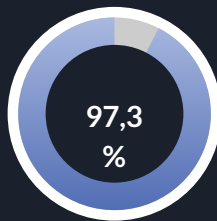
Fitting machine learning models

Logistic
Classifier



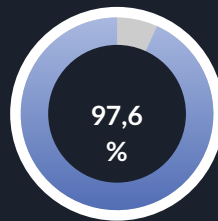
Accuracy
score

Random
Forest
Classifier



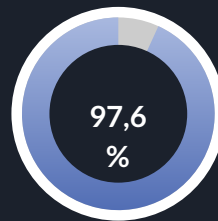
Accuracy
score

Gradient
Boosting
Classifier



Accuracy
score

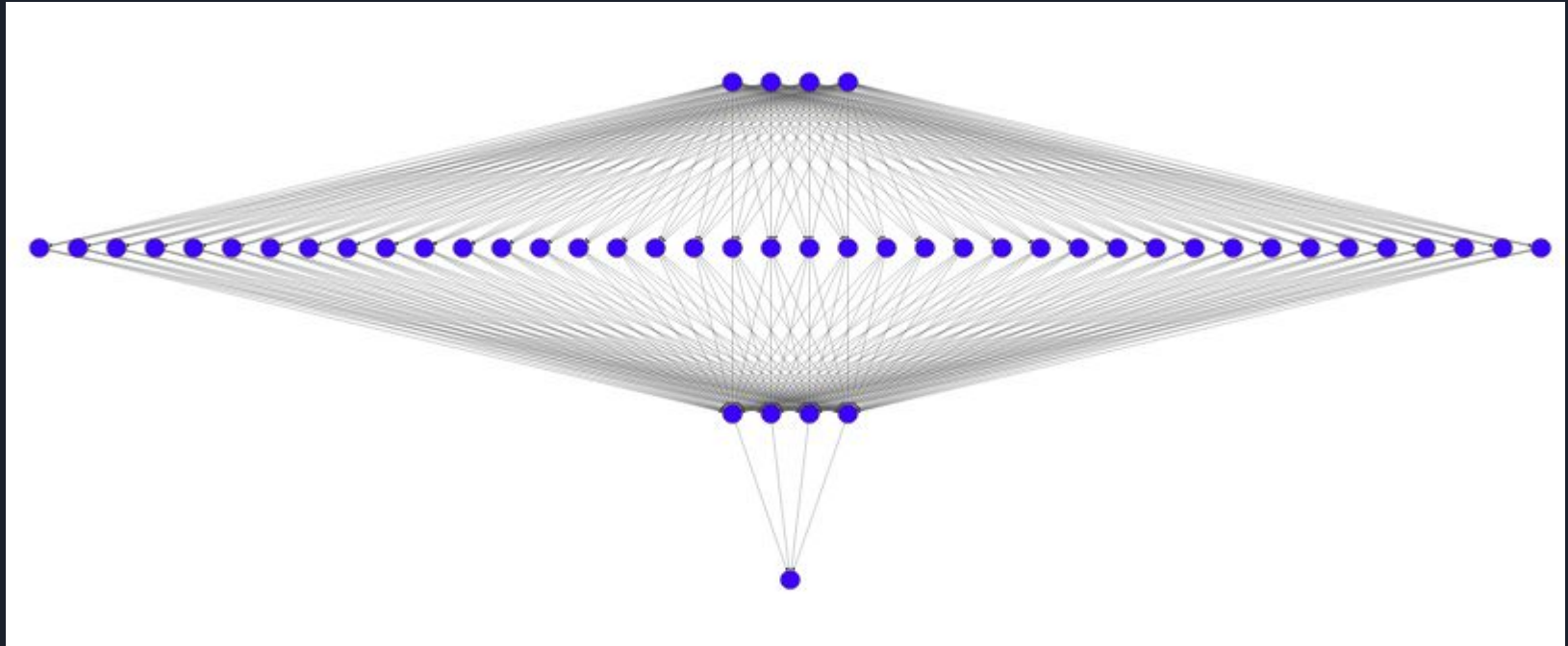
Neural
Network



Accuracy
score

Choice: Fully connected neural network

- Two hidden rectified linear unit (ReLU) layers
- Sigmoid activation function





Best performance in terms of:

- Speed of fitting
- Accuracy score 97.6 %
- False positive rate $FP/(FP+TN) = 1.8\%$
- False negative rate $FN/(FN+TP) = 10\%$
- Sensibility of predictions



Why is it relevant:

- Lenders: understanding their approval process better
- Allows potential borrowers to check their approximate approval likelihood without affecting their credit score
- Allows rejected applicants understand why they've been rejected and what they could do differently



Future research:

- Use these data to fit machine learning models across different years
- Understand how the applicants and application process have evolved since 2007
- Continue working on interest rate and default predictability models