

Caleb Harris

Yeonjae Zhang

Loan Default Predictor

Avoiding risky loans: using classifying models to save banks money.

Cash Value Provided by Model

Banks set aside billions of dollars in anticipation for defaults.

How defaulting affects a bank:

- ❖ Loss of Revenue
- ❖ Loss of Profit

The Goal:

- ❖ Predict if an applicant is high risk to default on a loan.
- ❖ Less risk = more money



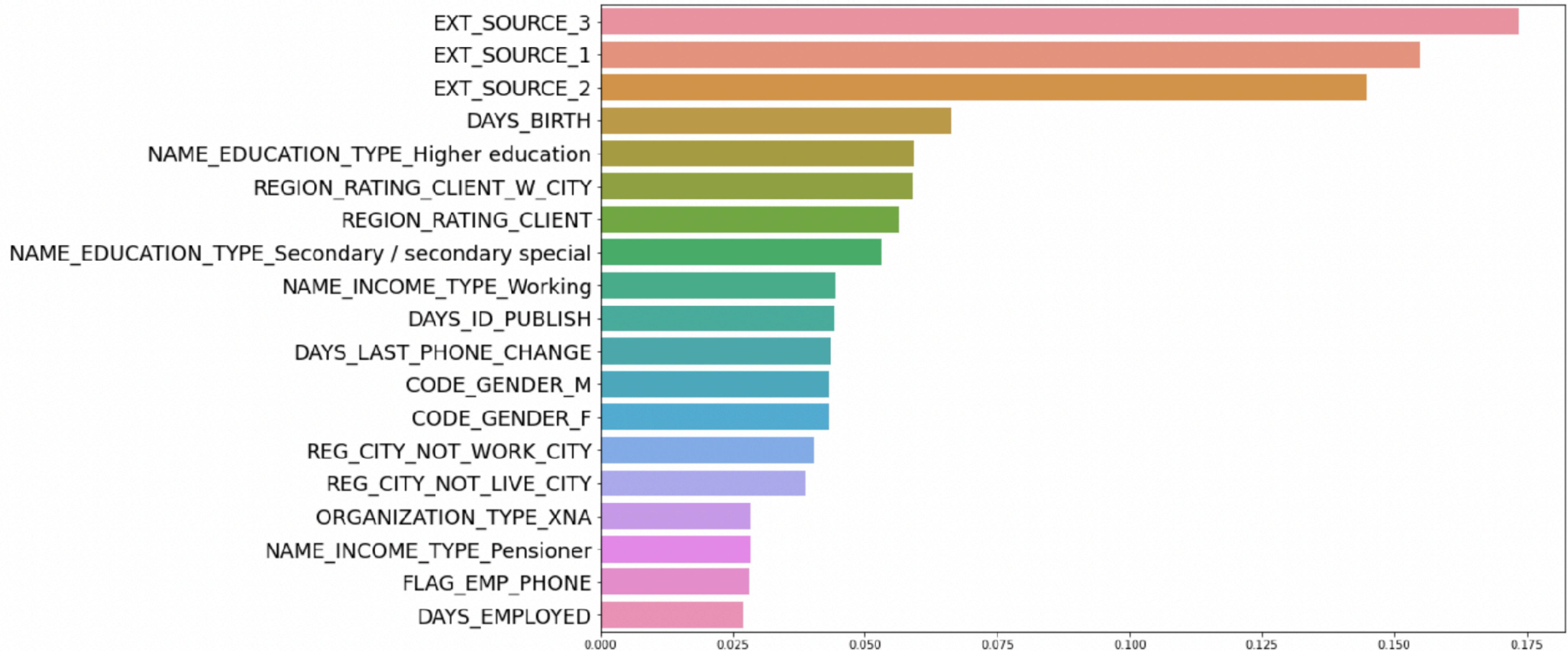
Data

The Dataset:

- ❖ Banking and finance dataset
- ❖ Contains personal and financial data of loan applicants



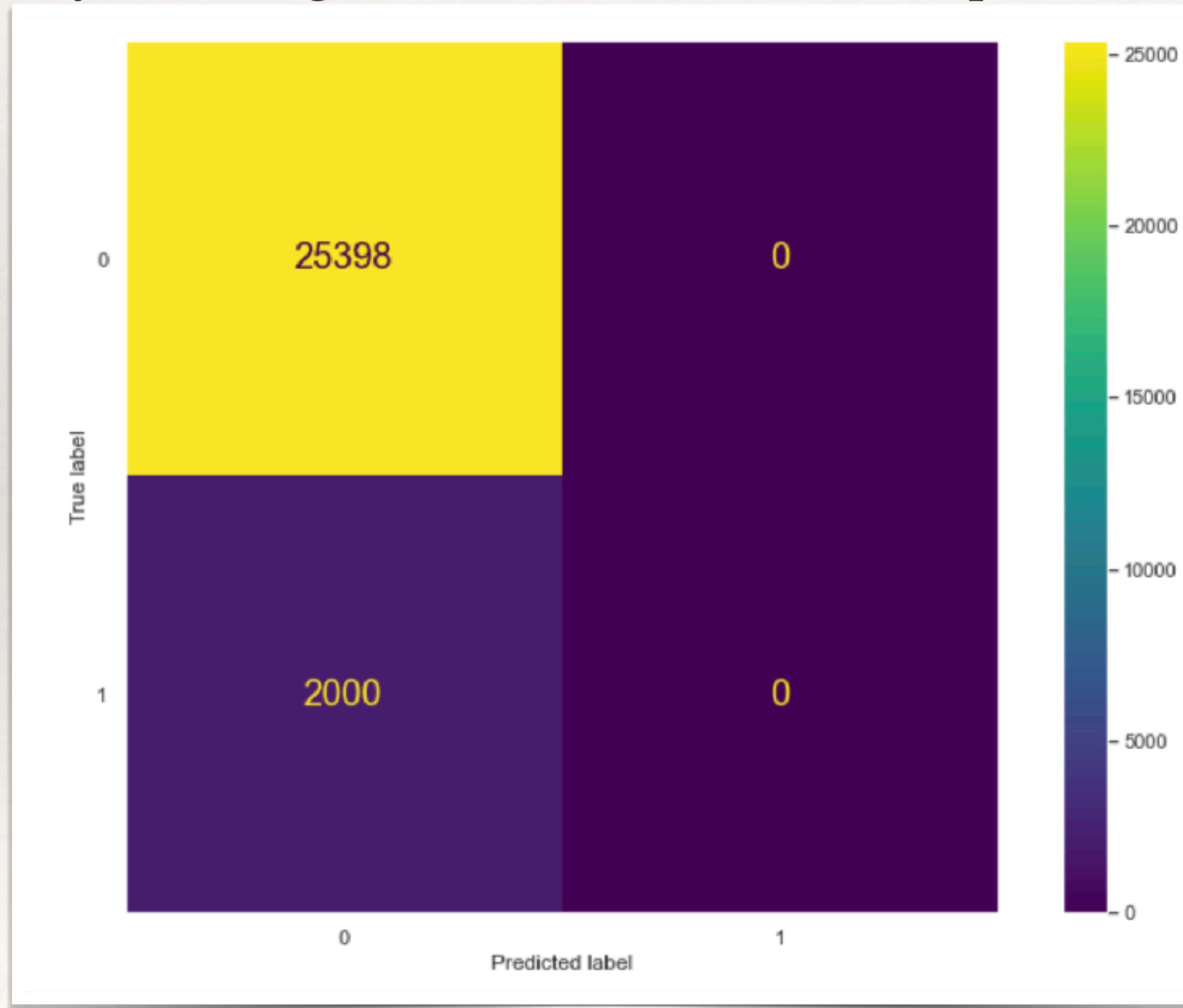
Most Correlated Features with Target Variable



Model Performance

BAD MODEL MATRIX:

This matrix shows our model is predicting way too many false negatives (2,000) and no true positives (0)



When Looking at our goal and potential outcomes, what do we not care about?

- ❖ FALSE POSITIVES - we don't really care about them

When Looking at our goal and potential outcomes, what DO we care about?

- ❖ FALSE NEGATIVES - we do really care about them

False Negatives and False Positives

- ❖ We care about serving our client so they will pay for our model
- ❖ False negative - Bank will be mad at us
- ❖ False positive - Borrower will be mad at us
- ❖ If the bank is not pleased with us because our model isn't improving the number of borrowers defaulting on loans - nobody will pay for our model.



Better Performance

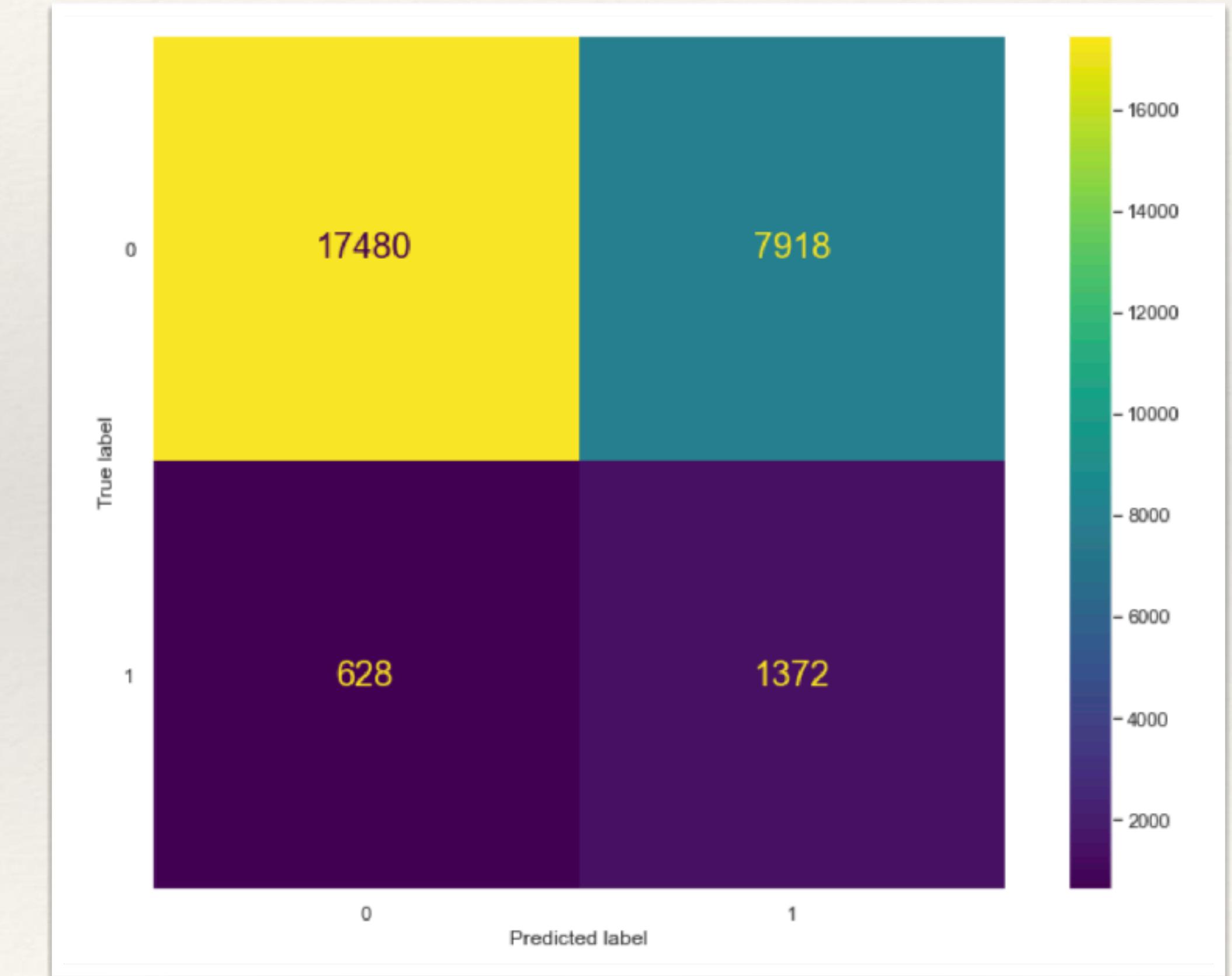
To improve our model we focused on

minimizing false negatives.

- ❖ Accuracy Score: 69.3%
- ❖ F-1 Score: 24.4%
- ❖ Recall Score: 68.6%

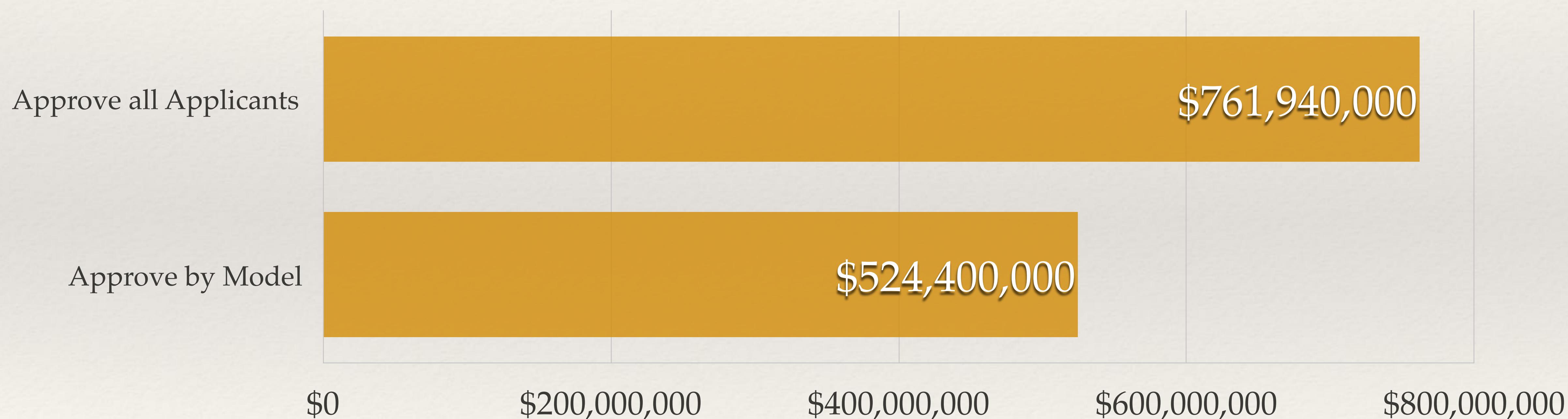
Better Model Matrix:

This model predicts 628 false negatives and 1372 true positives



Potential Earnings Ignoring Defaults

Earning Estimates From 27,398 future clients



Potential Earnings Including Defaults

Earning Estimates From 27,398 future clients

Approve All Loans

\$162,000,000

Approve Loans by Model

\$336,000,000

\$0

\$85,000,000

\$170,000,000

\$255,000,000

\$340,000,000

How to improve:

- ❖ Talk with industry insiders.
- ❖ Collect more data!!!!

