Machine Learning Loan Processor

By: Sean Santiago

Loan Processor

- WAU Bank

- Loan Processing Costs

- Historical Loan Data

- Machine Learning Classification Model

- Label Loan Applications Good or Bad

- Prioritize Minimizing Approving Bad Loans

Data

- Historical loan data from 2007-2018

- Variety of loan types

- Application available data

- Loan status data

Class Labels

- Good Loans = 1
  - Positive Class
  - Loans that are 30 days late or less
  - Loans in good standing
  - Loans paid-in-full

- Bad Loans = 0
  - Negative Class
  - Loans that are more than 30 days late
  - Defaulted loans
  - Charged-off loans

Evaluation Metrics Main Model

- Baseline = .872

- Loans predicted good (380,003, 88%)

- Accuracy Score: 0.929

- Recall Score: .963

- Specificity Score: .696

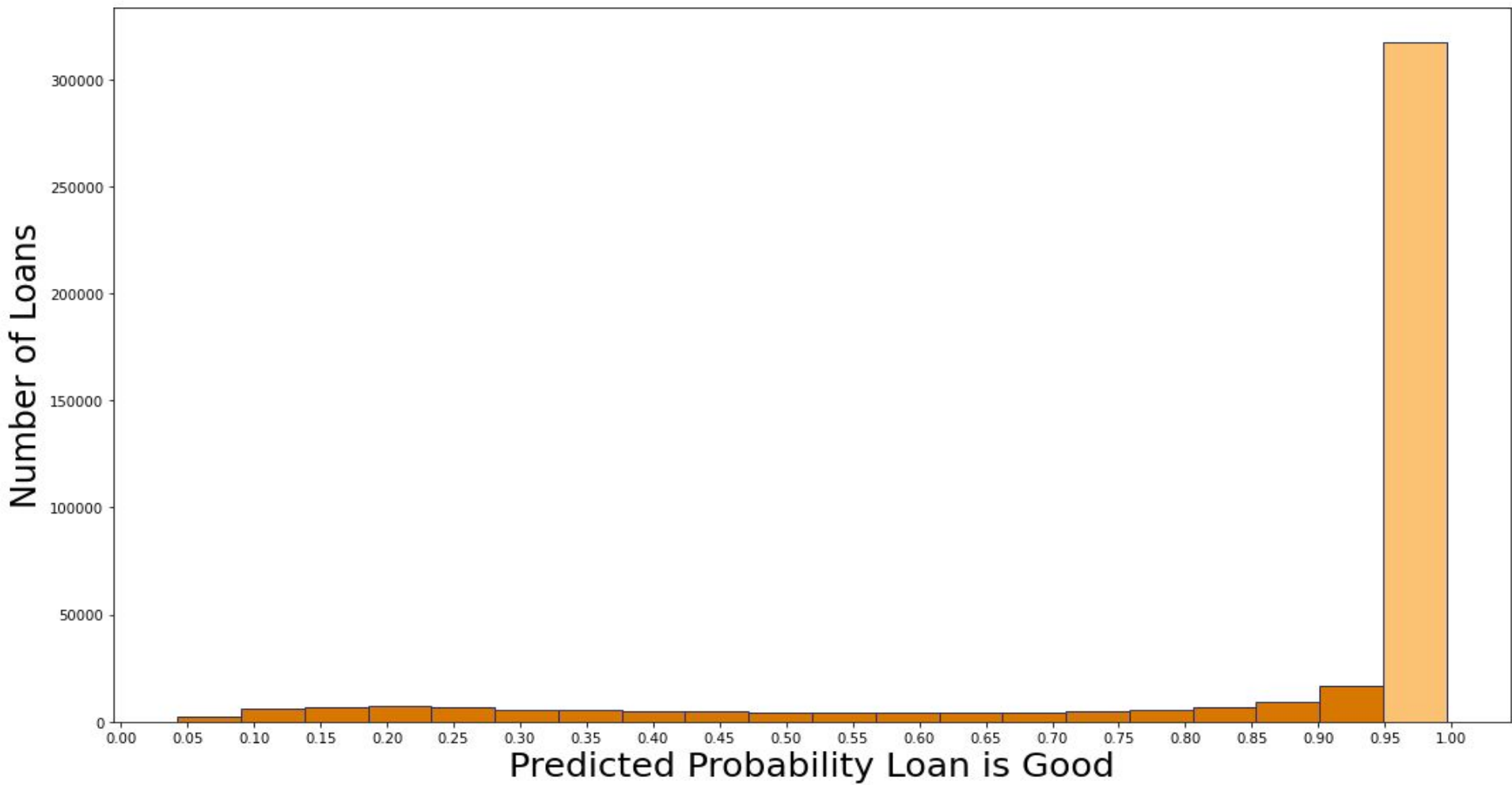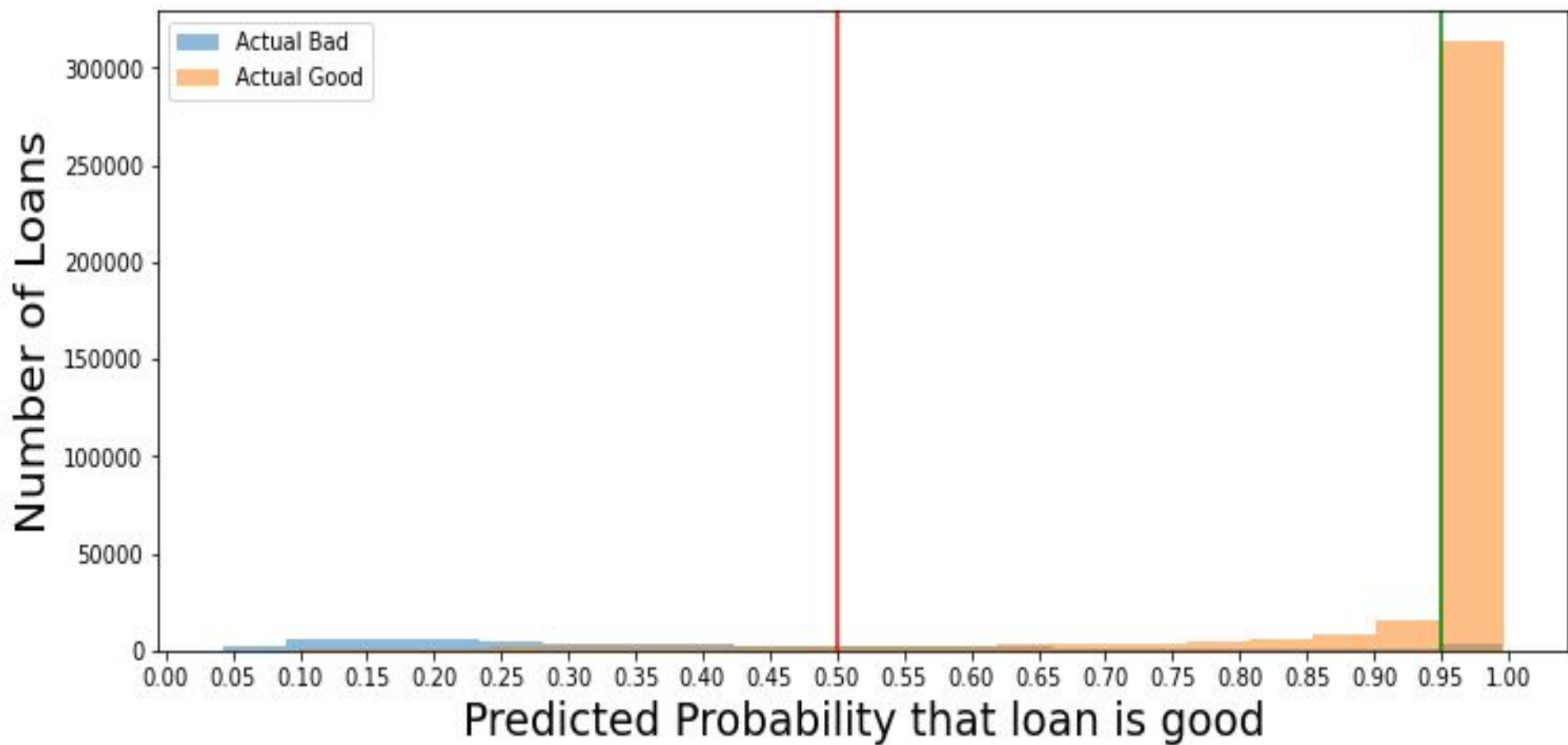- Precision Score: .956
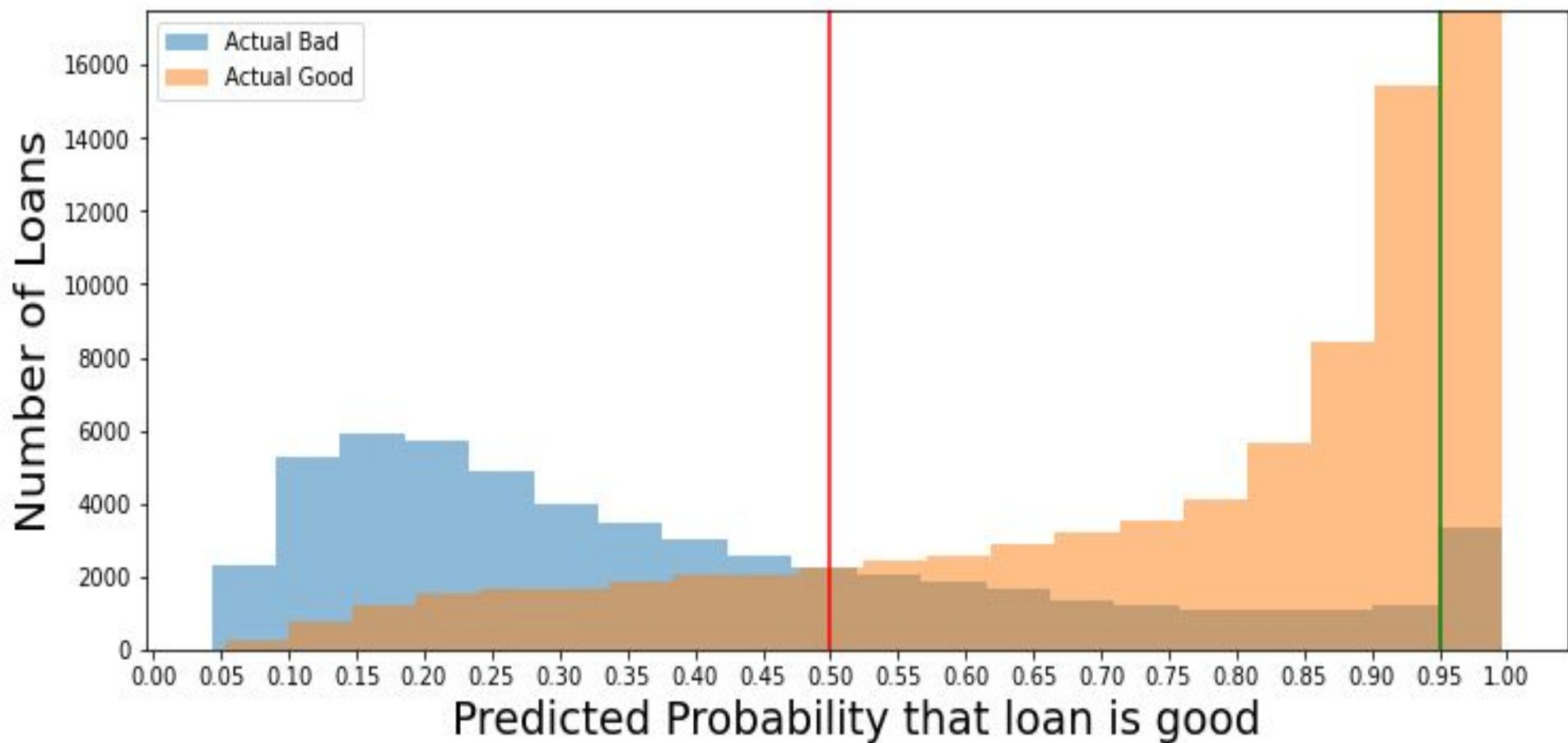
- F1 Score: .959

ROC Curve
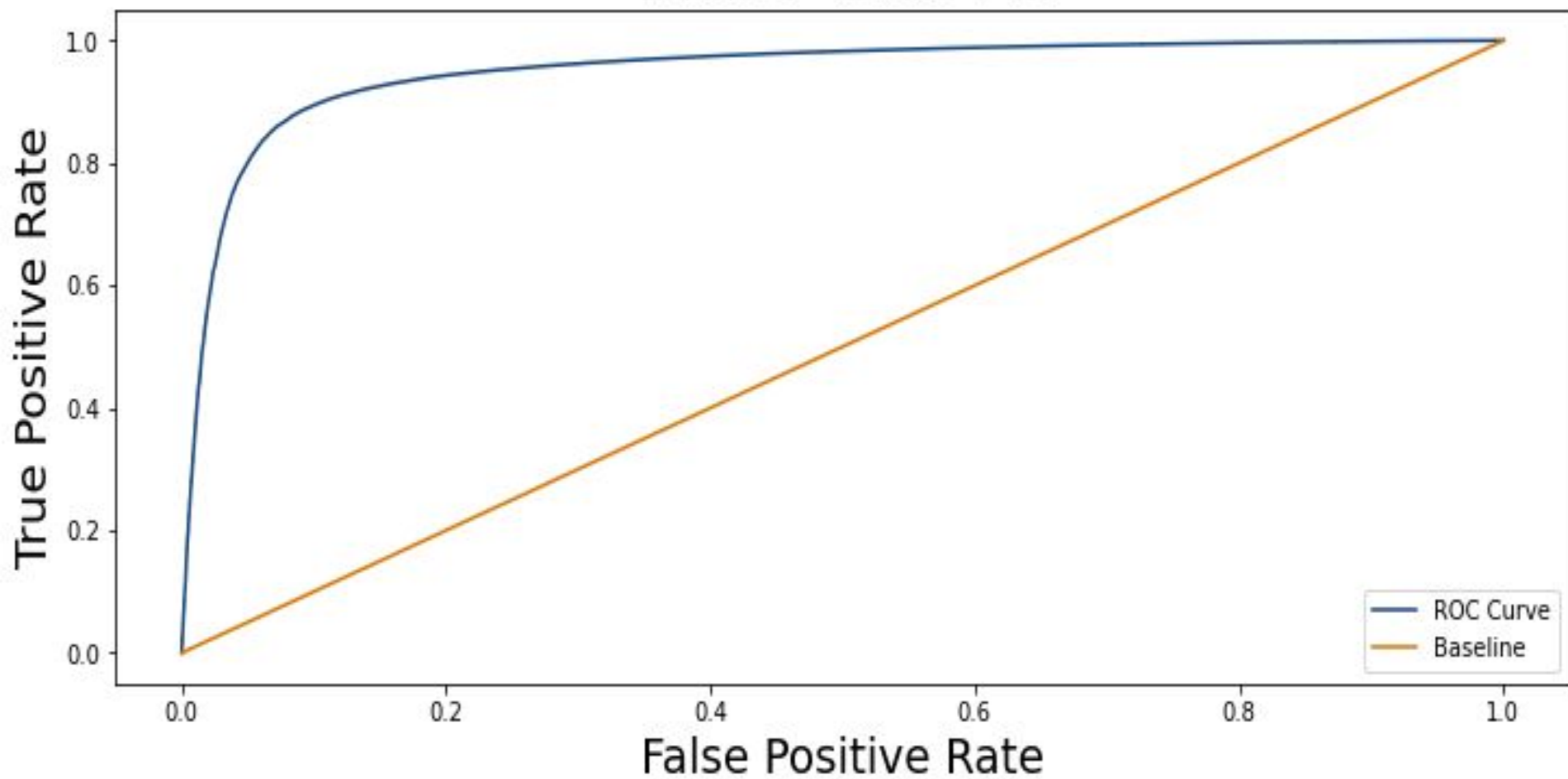
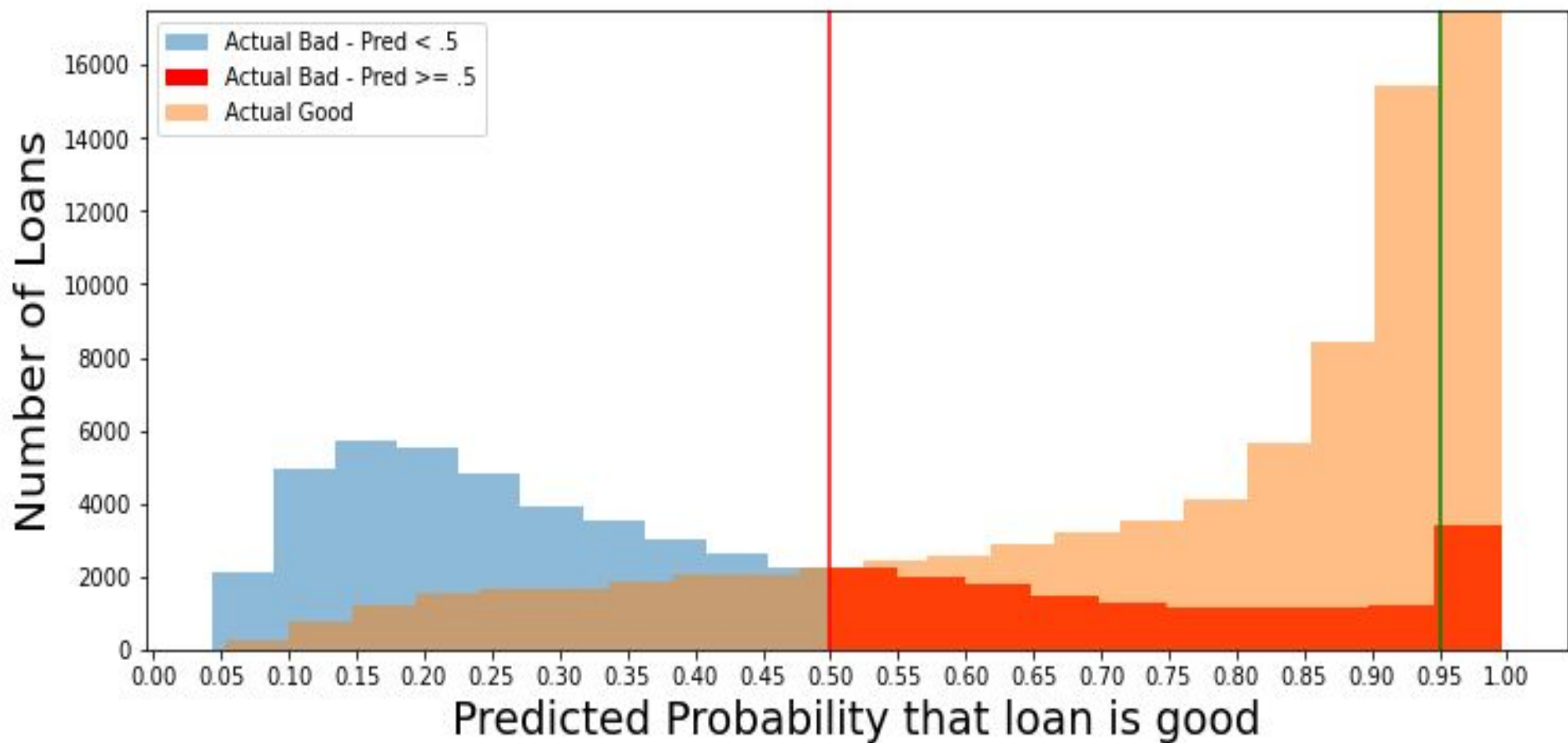# Loan Confusion Matrix

Distribution of Prediction Probabilities

Distribution of Prediction Probabilities

Zoomed in on Probability < .95

Zoomed in on Probability < .95

Histogram with legend:
- Actual Bad - Pred < .5
- Actual Bad - Pred >= .5
- Actual Good

Y-axis: Number of Loans

X-axis: Predicted Probability that loan is good

Loan Confusion Matrix

Loan Confusion Matrix

Loan Confusion Matrix

	Bad	Good
Bad	51950	3286
Good	63686	313538

Actual (rows) / Predictions (columns)
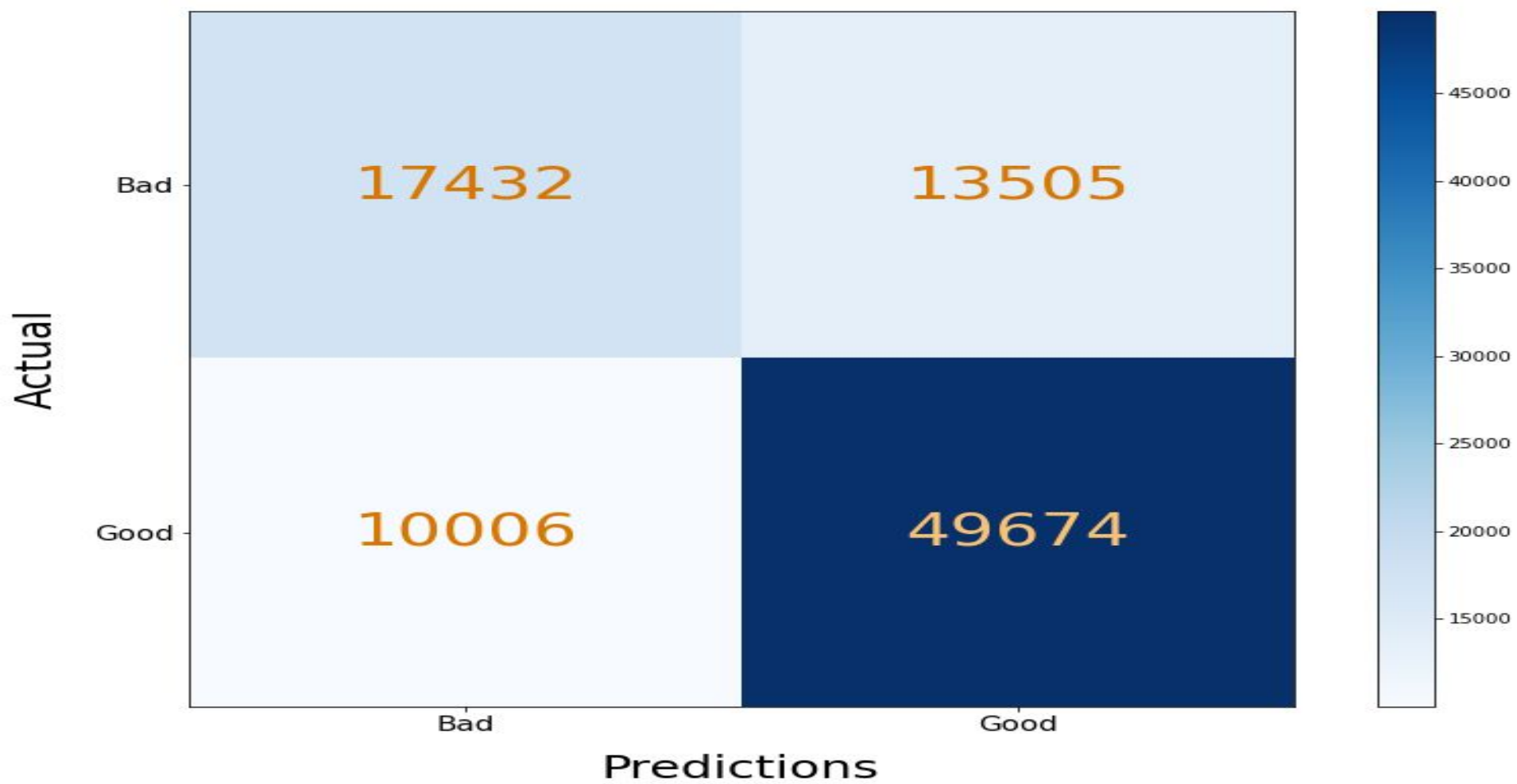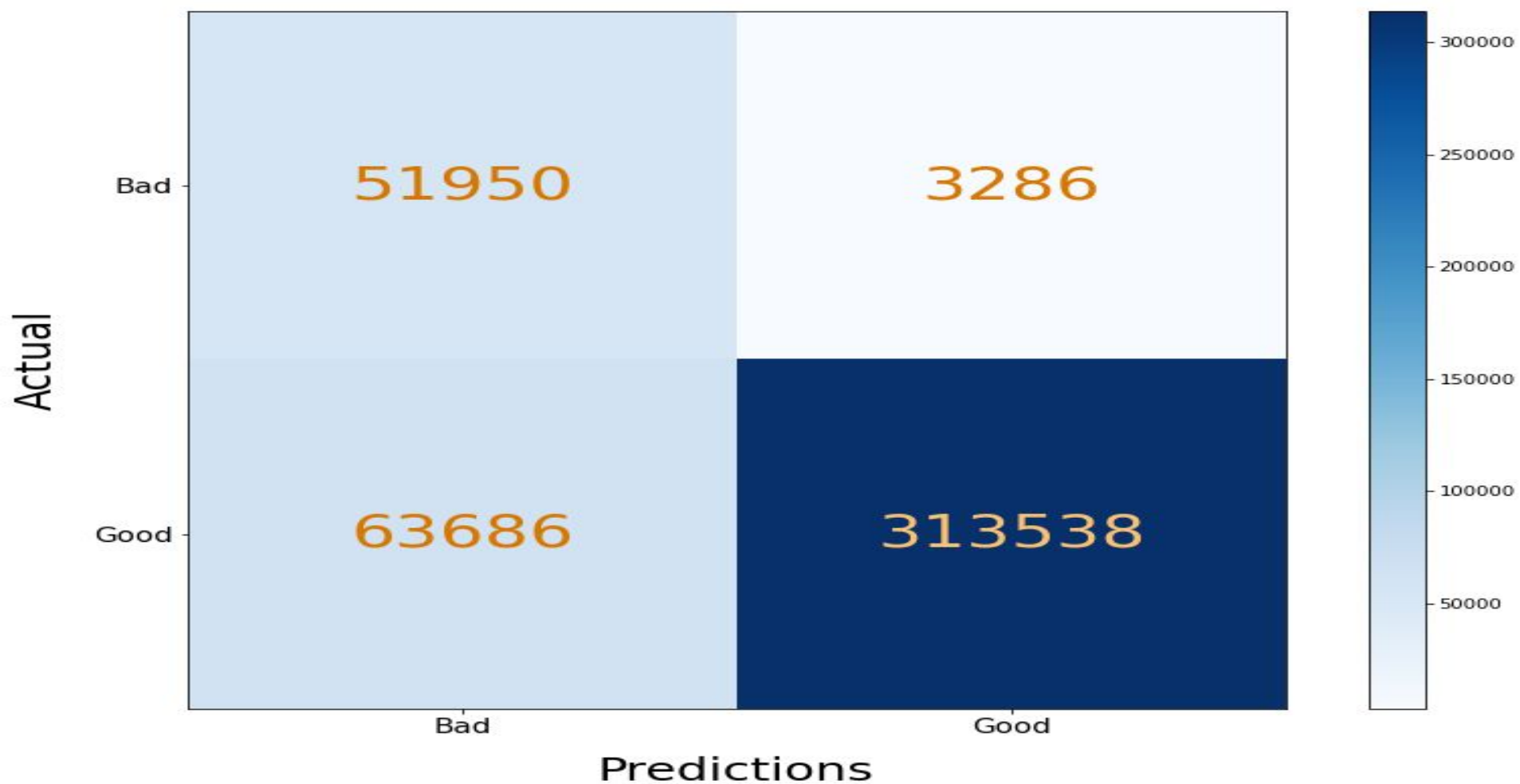
Prediction Probabilities > .95 Model

- 316,824 - 73% of loans had prediction probability of .95 or higher

- Predicted good, actually good = 313,538

- Predicted good, actually bad = 3286

- Accuracy Score: .845

- Precision Score: .990

- F1 Score: .904

Conclusion and Recommendations

- Models can significantly cut down on manual loan processing costs

- Using .95 Prediction Probability as threshold increases Precision

- Other scoring metrics decrease with .95 model

- Thresholds can be adjusted to accommodate business goals