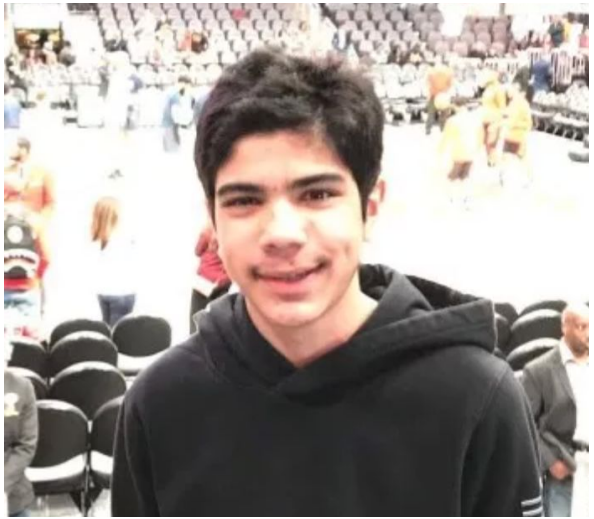
Credit Risk Prediction With Python

Vidur Modgil
Advisor: Dr. Priestley

Introduction

- Rising Junior at Johns Creek High School
- Robotics
- Academic Bowl
- H^3 Charity Work
- Basketball



Overview

1. Credit Scoring Code - STAT 8330 Binary Classification
2. Impacts of Data Science on Banking

Credit Scoring Project

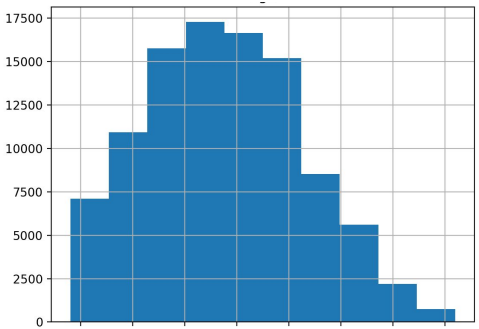
- Goal: predict whether a given person would default
- Uses predictors to come up with a 0/1 outcome
- Binning of predictors to increase accuracy
- Logistic Regression

The Data

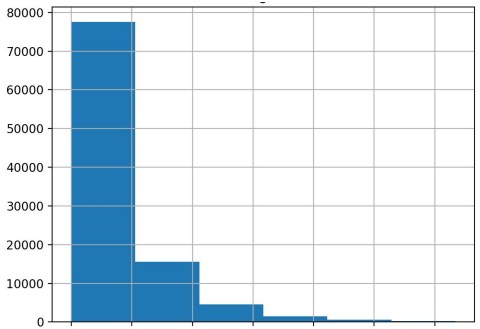
- Data Was 100,000 Rows By 8 Columns
- Predict 0/1 goodbad variable
- 4 Predictors:
  - Age
  - RBAL
  - BRPCTSAT
  - Trades

The Distribution

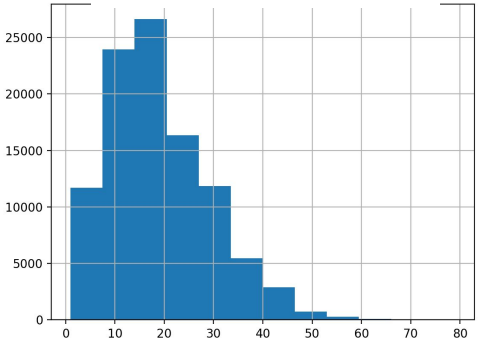
AGE



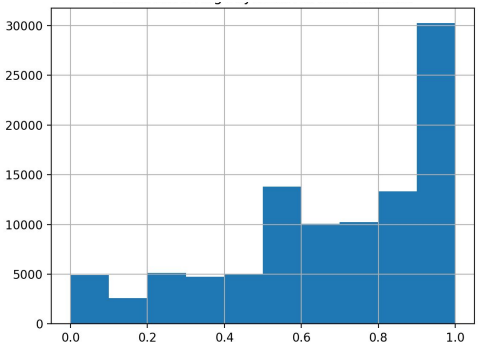
Revolving Balance



Trades



Bank Revolving Percent Satisfied



Binning of Data

- Final form of data were binned versions of predictors
- Example of bin ranges
  - RBAL: 0-5000; 5000-10,000; 10,000-15,000; 15,000-20,000; 20,000-40,000; 40,000-Max
  - 0-5000 Bin has 17% chance of Default
  - 40,000-Max has 19% chance of Default

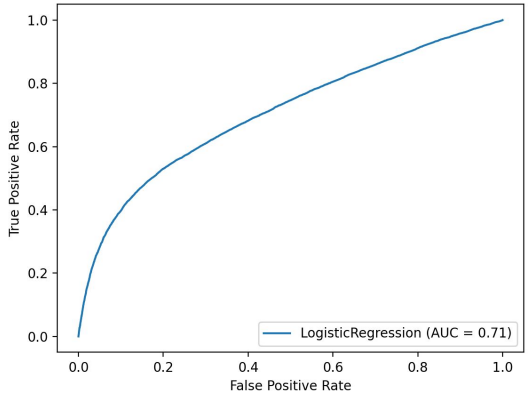
The Code

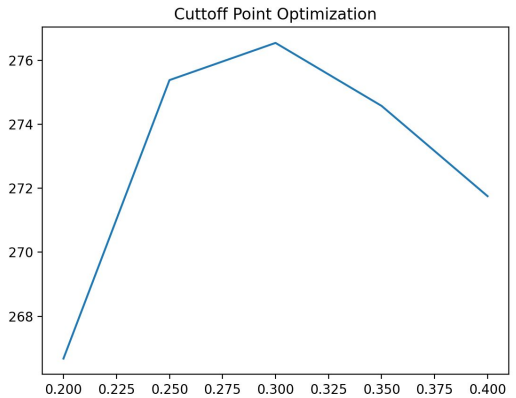
```
# Fitting of the data
logisticRegOrd = LogisticRegression()
logisticRegOrd.fit(usableModelDataX, usableModelDataY)

# Prediction of p(default)
ordPrecitionProbability = list(logisticRegOrd.predict_proba(usableModelDataX)
ordModelCoeff = list(logisticRegOrd.coef_)
ordModelInt = list(logisticRegOrd.intercept_)
ordModel = list(ordModelCoeff[0])

ordModel.append(ordModelInt)

# Output of Model Scoring using AUC(Expected value is 0.71)
modelData['Prediction Prob_ORD'] = ordPrecitionProbability
metrics.plot_roc_curve(logisticRegOrd, usableModelDataX, usableModelDataY)
plt.show()
```





The Math

- Equation

$$P(default) = 1 - \frac{e^{0.7736 - 0.1770(OrdTrades) - 0.0617(OrdAge) + 0.1759(OrdBRPCTSAT) - 0.2929(OrdRBAL)}}{1 + e^{0.7736 - 0.1770(OrdTrades) - 0.0617(OrdAge) + 0.1759(OrdBRPCTSAT) - 0.2929(OrdRBAL)}}$$

- Example
  - Hypothetical Customer John Smith: age 84, 10 tradelines, 0.6667 BRPCTSAT, and 1635 RBAL
  - Johns Smith P(default) = ~4.3%, and according to the cutoff point, is safe to extend credit

Questions

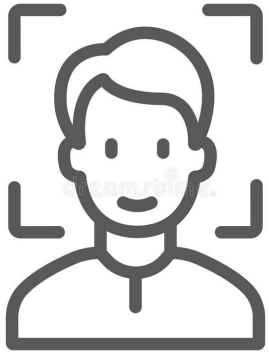
Data Science Innovations: Impacts on Banking

- NLP Chatbots
- Image Recognition with CV
- Biometric Data Authentication
- Mobile App Integration
- Unintended Consequences

NLP Chatbots

- Use Natural Language Processing algorithms (NLPs) to maximize effectiveness
- Decrease Operating Costs of banks
- Increase Customer Satisfaction when compared to performance of non-NLP powered chatbots

Computer Vision



- Integration of Computer Vision allows for more remote banking
- Able to eliminate some back-end processing costs and increase accuracy

Biometric

- Biometric data used for fraud prevention
  - Eyeprint
  - Voice Recognition

Mobile App Integration

- Easiest method of deployment for the above technologies
- Also allows for increased data collection to improve personalization of experience

Unintended Consequences

- Ethics of Collecting Data
- Replacement of Human Workers with AI
- Bias in algorithms

Questions

Three Things I learned

1. How to prepare data before building a model
2. Building a Logistic Regression model in Python
3. Research methodologies for current innovations

Works Cited

Arthur, Alison, and Bethany Frank. “Five Examples of Biometrics in Banking.” *Alacriti*, 8 May 2019, www.alacriti.com/biometrics-in-banking.

Joshi, Naveen. “How Machine Vision Can Transform Financial Services.” *Forbes*, Forbes Magazine, 9 Sept. 2019, www.forbes.com/sites/cognitiveworld/2019/09/09/how-machine-vision-can-transform-financial-services/?sh=3c320be61e88.

Ostapchenya, Denis. “The Role of Big Data in Banking : How Do Modern Banks Use Big Data?” *Finextra Research*, Finextra, 11 June 2021, www.finextra.com/blogposting/20446/the-role-of-big-data-in-banking--how-do-modern-banks-use-big-data.

Rijmenam, Mark. “How AI Is Changing Retail Banking.” *Open Data Science - Your News Source for AI, Machine Learning & More*, 6 Apr. 2018, opendatascience.com/how-ai-is-changing-retail-banking/.

Shroff, Raj. “Natural Language Processing in Banking: Current Uses.” *Medium*, Towards Data Science, 27 Dec. 2019, towardsdatascience.com/natural-language-processing-in-banking-current-uses-7fbbaee837de.