Machine Learning Project
Lending Club

Anna Soloveva, Hayley Perry, Volodymyr Medin

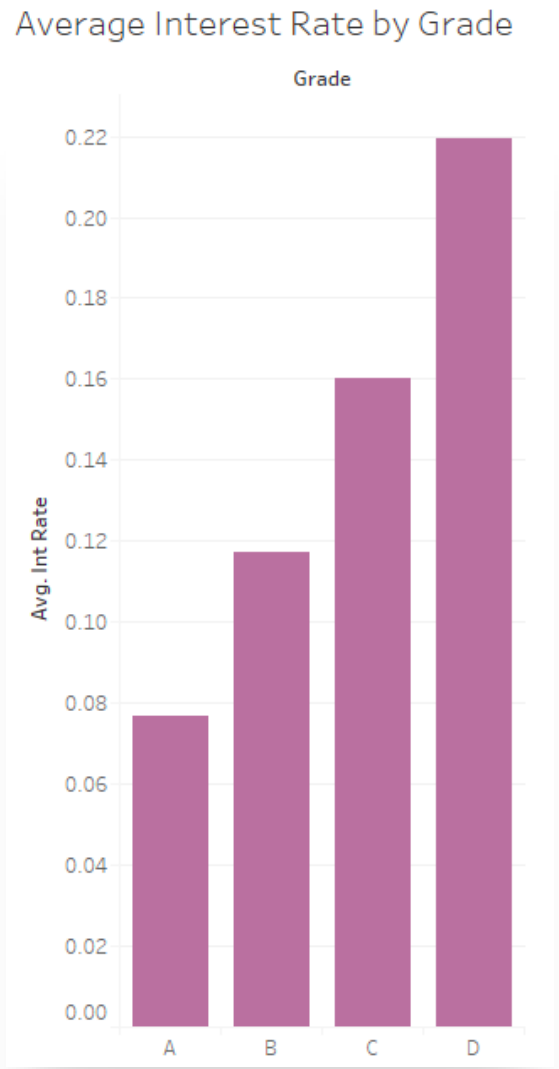
May 5th, 2020

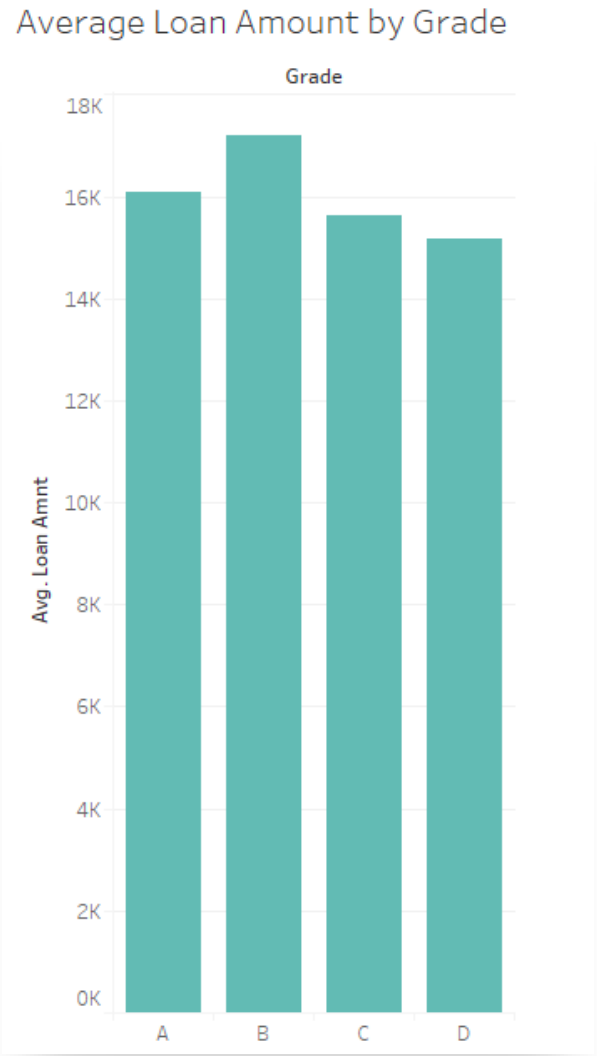
Agenda

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Model Evaluation
- Conclusion
- References

Business Understanding

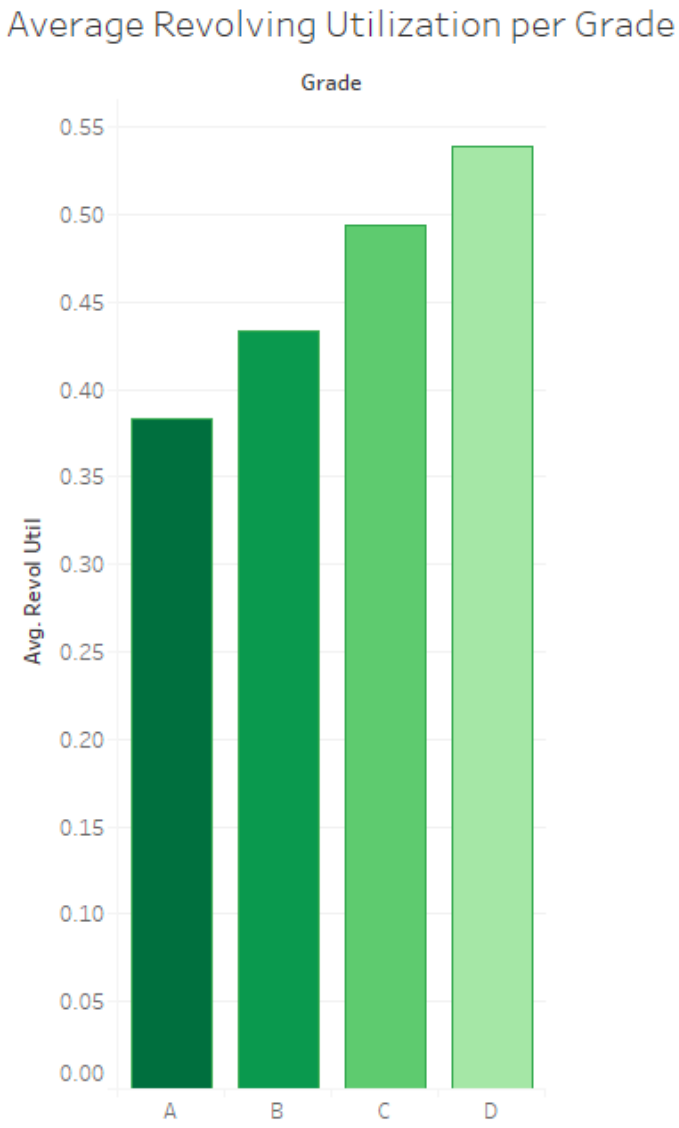
- Peer-to-peer lending platform
- Offers various loan products to interested parties through their proprietary technology platform
- The platform automates main aspects of the borrowing process such as data and application processing, decision generation, loan funding and compliance with regulations.

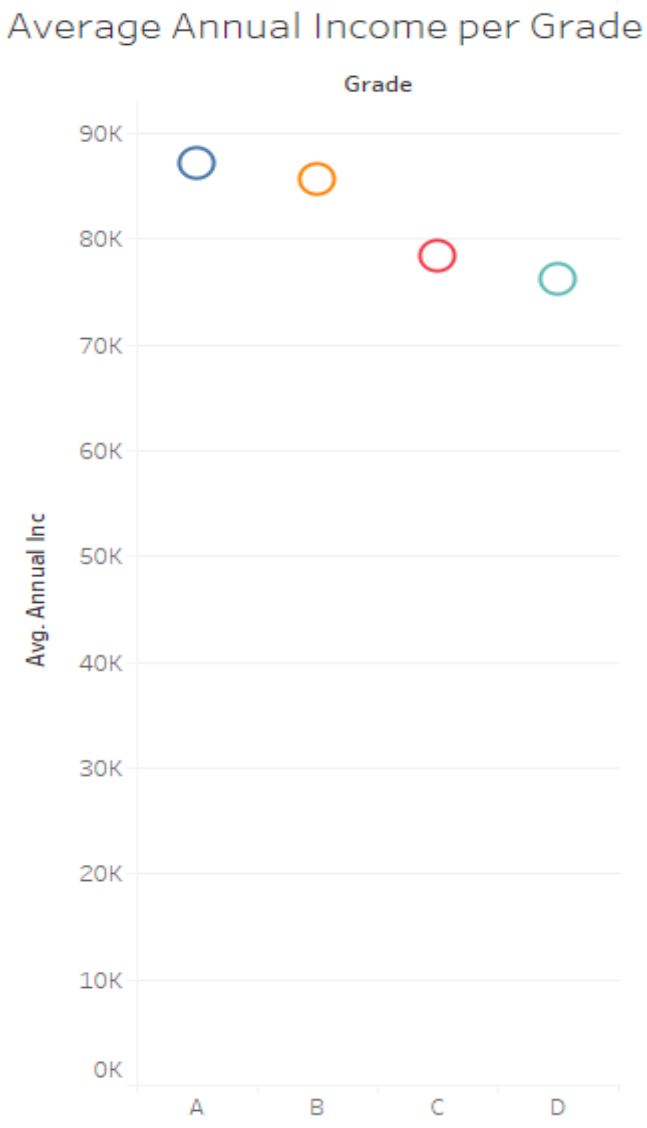




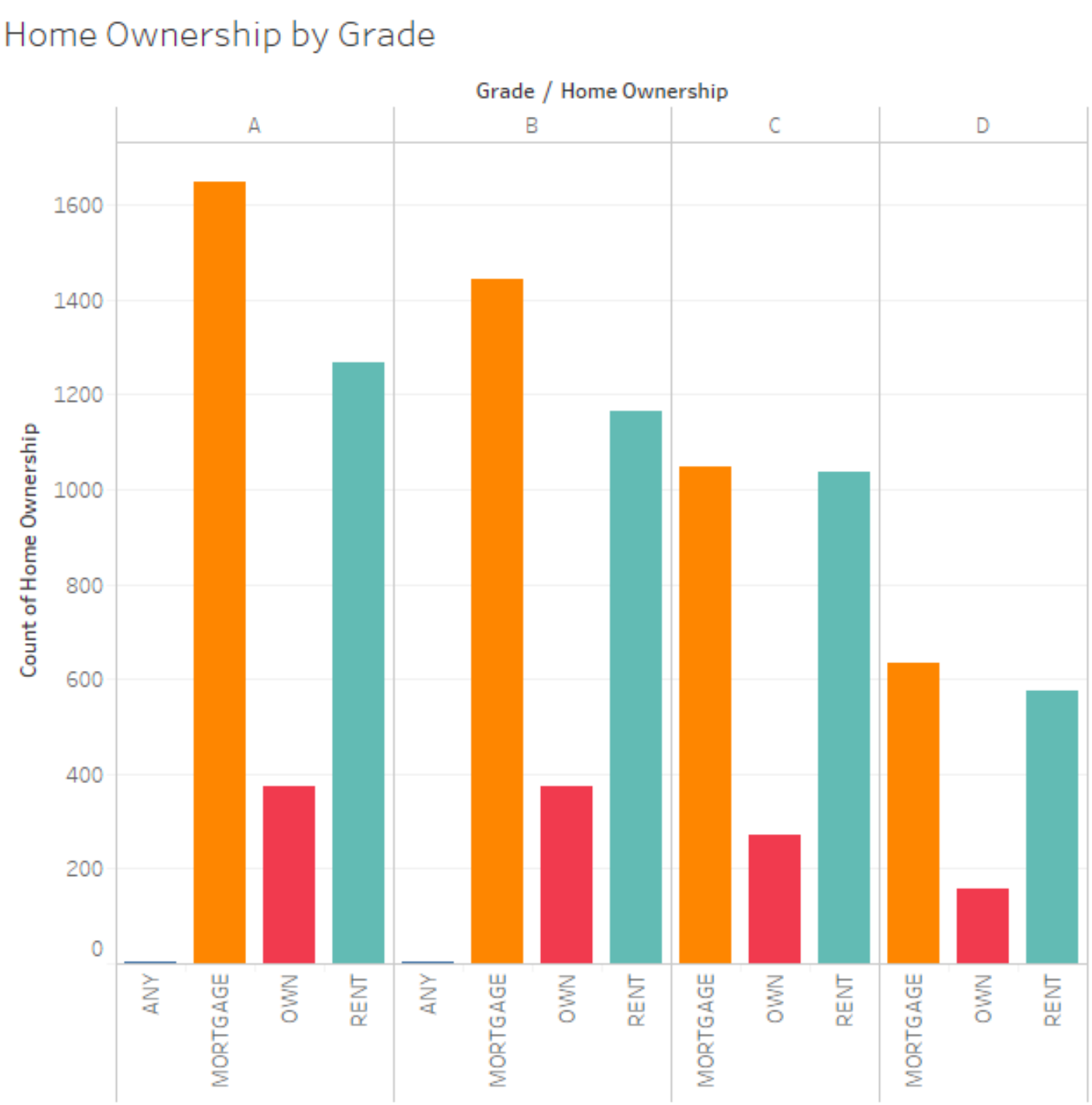
Data Understanding

- The LendingClub has loan data available publicly to investors on their website.
- The data ranges from first quarter of 2019 to third quarter of 2019.
- Original data has 664,031 observations and 150 variables
- Used 10,000 loan applications
- Some of the explanatory variables are Loan Amount, Funded Amount, Term, Interest Rate, Installment, Grade, Sub Grade, Employment Status, etc.

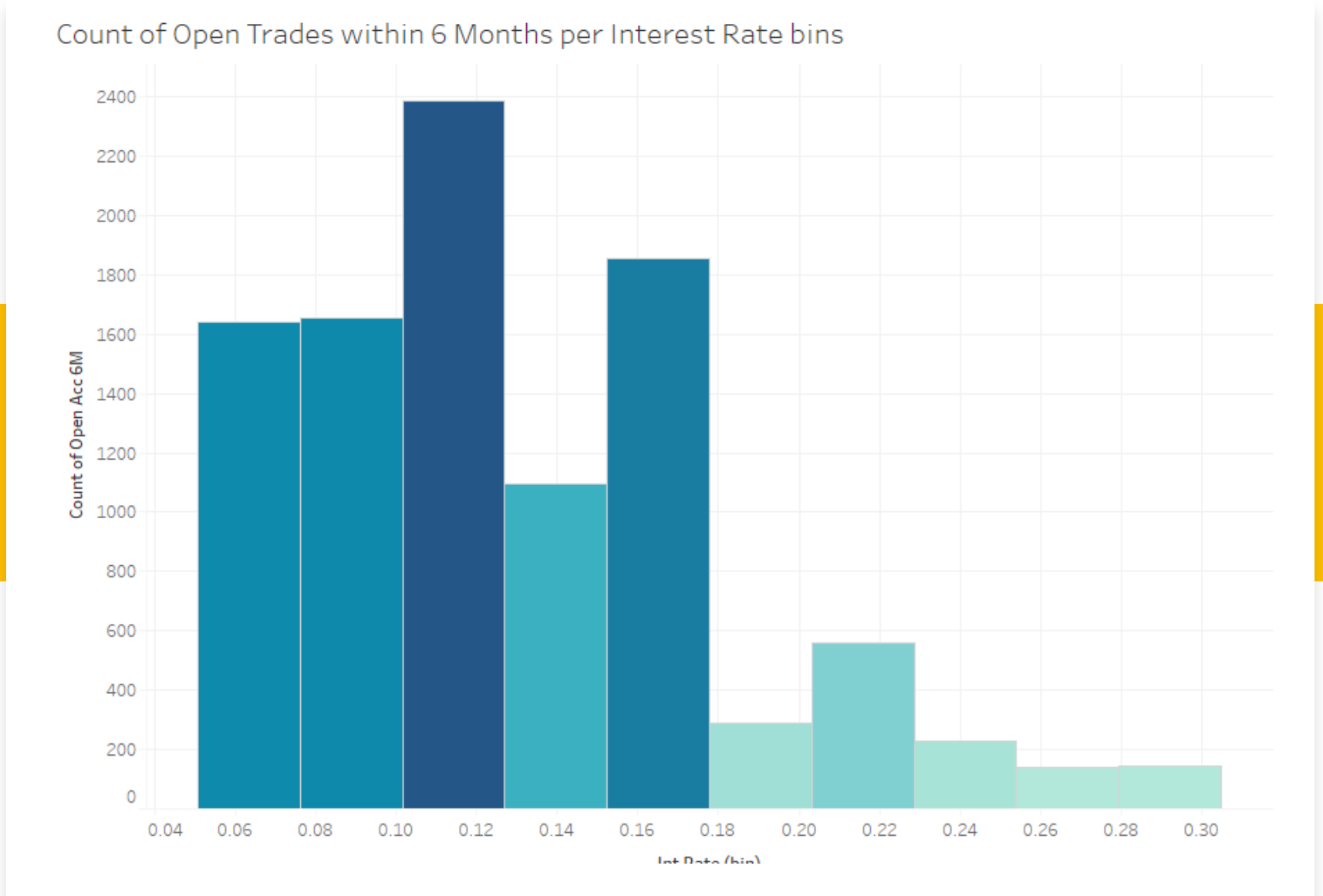




Data Understanding



Data Understanding



Data Understanding

Data Preparation:

- 9 Appended the three quarters for 2019
- Filtered the data to get the first 10,000 observations
- Removed the original data set to empty space in the R environment
- Changed the regression variables to the appropriate type
- For KNN and Neural Networks we removed observations with NA values

Modeling

Linear Regression

Target Variable – Interest Rate

Model 1:

- Revolving Credit Utilization
- Number of Open Trades in the last 6 months

Model 2:

- Revolving Credit Utilization
- Number of Open Trades in the last 6 months
- Last FICO Range Low

Model 3:

- Revolving Credit Utilization
- Number of Open Trades in the last 6 months
- Last FICO Range Low
- Annual Income

K-nearest neighbors (KNN)

Target Variable – Grade

Model 1: (K = 5)

- Revolving Credit Utilization
- Number of Open Trades in the last 6 months

Model 2: (K =10)

Model 3: (K = 5)

- Revolving Credit Utilization
- Number of Open Trades in the last 6 months
- Last FICO Range Low
- Annual Income

Model 4: (K = 10)

Artificial Neural Network (ANN)

Target Variable – Interest Rate

Model 1: (1 hidden layer of 2 nodes)

- Revolving Credit Utilization
- Annual Income

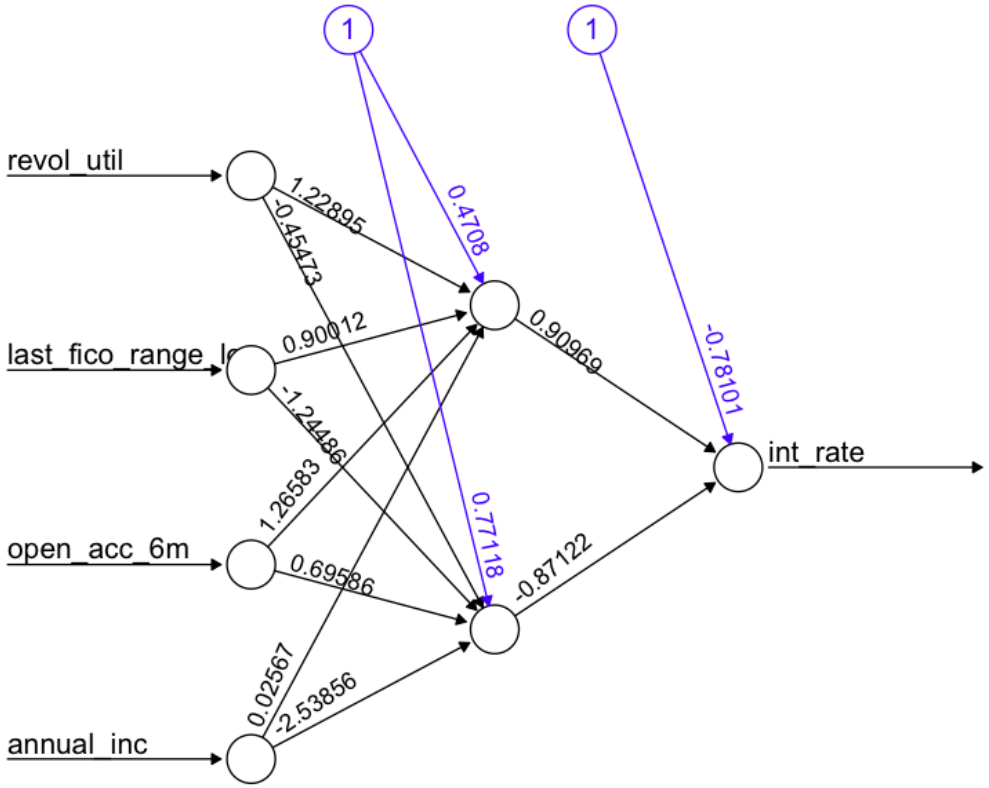
Model 2: (2 hidden layers of 5 and 2 nodes)

Model 3: (1 hidden layer of 2 nodes)

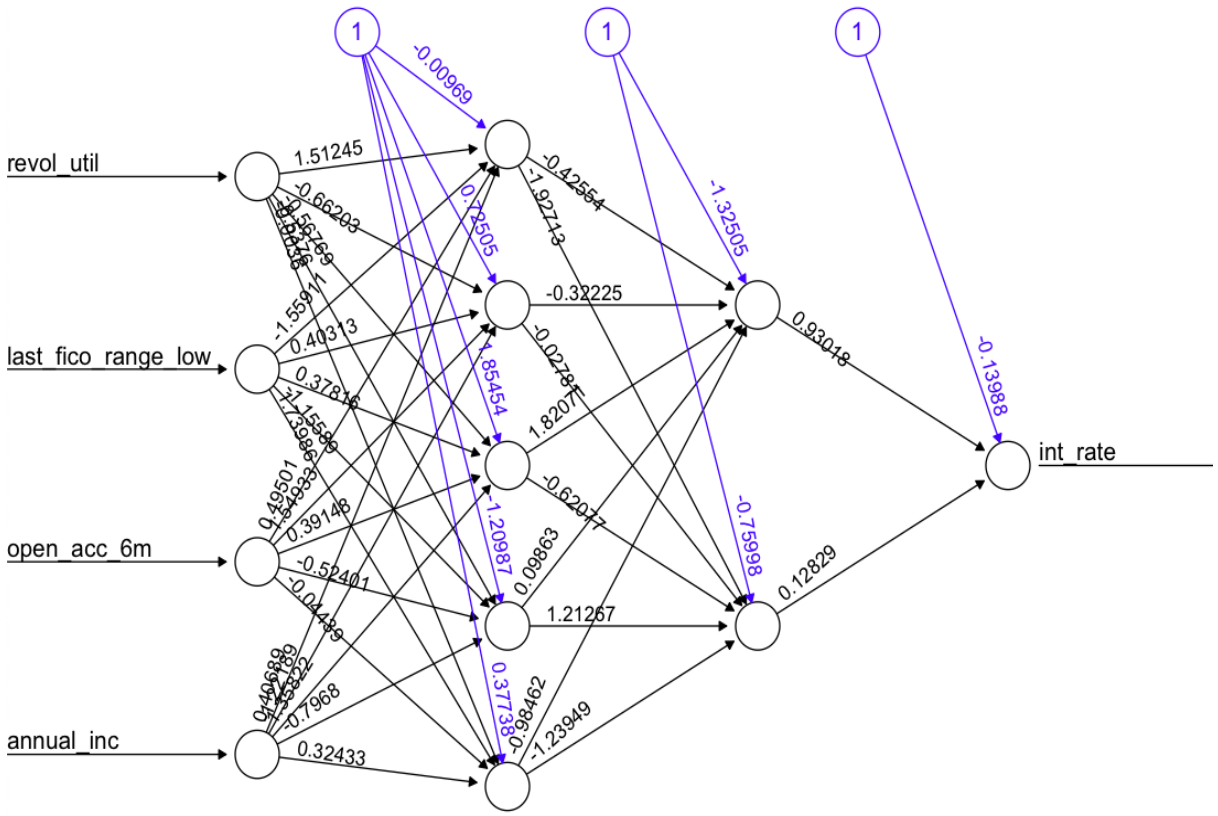
- Revolving Credit Utilization
- Number of Open Trades in the last 6 months
- Last FICO Range Low
- Annual Income

Model 4: (2 hidden layers of 5 and 2 nodes)

ANN Plot



Error: 4.641084 Steps: 39



Error: 4.525617 Steps: 162

Models Assessment

Linear Regression

- Testing MSE
- Model 1: 0.0029
- **Model 2: 0.0014**
- Model 3: 0.0015

K-nearest neighbors (KNN)

- Accuracy
- Model 1 (K = 5): 0.3028
- Model 2 (K =10): 0.3267
- Model 3 (K = 5): 0.3319
- **Model 4 (K = 10): 0.3418**
- No Information Rate: 0.33

Artificial Neural Network (ANN)

- Testing MSE
- Model 1 (1 hidden layer of 2 nodes): 0.002518
- Model 2 (2 hidden layers of 5 and 2 nodes): 0.002519
- **Model 3 (1 hidden layer of 2 nodes): 0.002516**
- Model 4 (2 hidden layers of 5 and 2 nodes): 0.002518

Conclusion

- While Linear Regression and ANN were used to predict Interest Rate for a loan application, KNN was used to classify the application by specific Grade
- Linear Regression and ANN were assessed by comparing testing MSE
  - LR Model 3 - MSE of 0.0022
  - ANN Model 3 - MSE of 0.0025
- KNN models with different configurations
  - Model 4 (K = 10, p = 4) has the highest accuracy among three models

References

- About Us: Save with LendingClub. (n.d.). Retrieved from https://www.lendingclub.com/company/about-us
- Harrison, O. (2019, July 14). Machine Learning Basics with the K-Nearest Neighbors Algorithm. Retrieved from https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761
- Python. (2019, September 19). neuralnet: Train and Test Neural Networks Using R. Retrieved from https://datascienceplus.com/neuralnet-train-and-test-neural-networks-using-r/
- john@hranalytics101.com. (2020, April 18). Tutorial: How to Assess Model Accuracy. Retrieved from https://www.hranalytics101.com/how-to-assess-model-accuracy-the-basics/

Thank you!

Any Questions?