Comparing Machine Learning Techniques for Predicting Bank Failure

Armen Eghian[1]

Abstract:

The study of machine learning has helped create and refine many types of predictive models. These models have been applied to countless problems, but the importance of banking stability has led many researchers to create models predicting if a bank will be active in future years. Neural networks, support vector machines, and traditional regression models, have been used to predict bank failure, but random forests have not been fully explored. This paper shows random forests can offer prediction power greater than neural networks and logit models when used to predict bank failure. This aligns with work comparing random forests and neural networks in other disciplines. The readability of a random forest offers a distinct advantage over neural networks when used for research purposes.

JEL classification: G33, C01, C45

Key Words: Failure Prediction, Decision Trees, Machine Learning

[1] Economics Undergraduate. Bryant University, 1150 Douglas Pike, Smithfield, RI 02917.
Email: aeghian@bryant.edu

1. Introduction

Because predicting bank failure fits the role of machine learning models, has been extensively analyzed with traditional statistics, and serves an important role in protecting the health of the economy, the use of bank data is an excellent tool in comparing the ability of predictive models. Many studies have shown the benefit of looking towards financial ratios over other forms of bank data because they allow for researchers to cut down on the dimensionality of a model without losing information (Beaver 1966). As a result, studies looking to use banking data to compare predictive models tend to use financial ratios as inputs (Tam and Kiang 1992). While many different models have been compared to predict bank failure few studies directly compare decision trees and neural networks. Multiple researchers have commented on the predictive power of neural networks, but always discuss difficulties in the model's interpretability and standardization. Random forests can create a much more understandable model while only sacrificing little predictive power (Jacobsen, Perner, and Zscherpel 2000).

This paper aims to explore three different machine learning techniques to better understand how they can be used to predict bank failures. The 3 models used will be a lasso logit approach, random forest, and neural network. The logit model and random forest can be easily interpreted through a few different techniques. While some techniques have been developed to interpret neural networks, these complex solutions fall outside the scope of this paper. The models will use 10 different banking ratios to predict the failure rate of 253 banks. The ratios were recorded during the years 2010-2016. Each year will create a new model to predict bank failure for the year 2016. More prediction data will allow the models to be compared more concretely. The relevance of this study is in understanding the tradeoffs to using each of the three listed machine learning models.

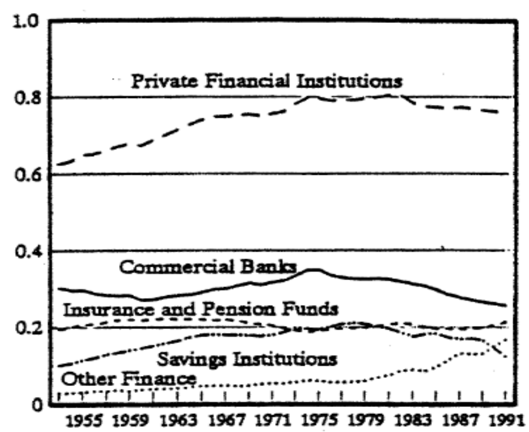
This paper is guided by three research objectives. The first being to evaluate the predictive ability of random forests against neural networks and logit models. Second, it aims to compare the human readability of each model. Finally, by utilizing a small dataset this research can help others decide which modeling technique is best for predicting bank failure given a small set of data. While much research has investigated predicting bank failure, few articles have directly compared random forests against other machine learning techniques.

The rest of the paper is organized as follows: Section 2 shows some trends in banking stability and business practice. Section 3 gives a brief overview on the current literature comparing decision trees to neural networks, analyzing bank data through traditional statistics, and using predictive models to better understand bank failure. The data and methodology are outlined in Section 4. Section 5 shows the results of all models. Section 6 ends the paper by discussing the implications of the study.

2. Trends in Commercial Banking Business Operations

Commercial banking in America has been put under stress for many interrelated reasons. The first being banks are losing their share of the credit market. Banks are also undergoing more risky business operations to make up for this loss in the credit market (Boyd and Gertler 1993). **Figure 1** demonstrates each financial institution's share of the credit market over time. Commercial banking saw a peak in its share of total credit in the 1970s but has seen a steady decline since the 1980s. During this time, a sharp rise in other financing methods such as bond sales has been rising dramatically.

Figure 1: Share of Financial Institutions in Total Credit[a]

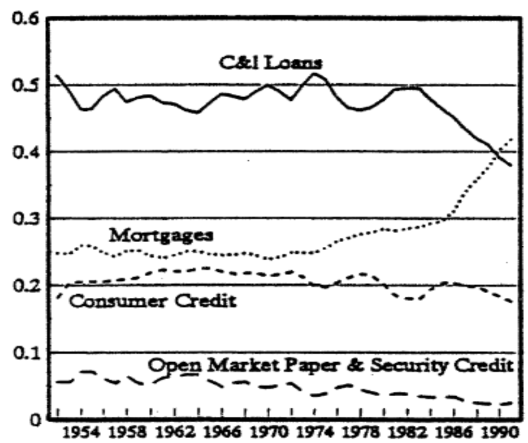


[a]Fractions of total funds advanced in credit markets to domestic nonfinancial sectors.

Source: Boyd and Gertler 1993

The researchers found this decrease in credit share has been occurring since the 1980s. They attribute this loss to the rise of open market credit operations and nonbank intermediaries. Both researchers found this decrease in the credit market was followed by a higher than usual period of bank failures. In the 1980s bank failures jumped dramatically (Boyd and Gertler 1993). Their work also looked at the composition of bank assets over time. By looking at the types of loans given out, the researchers found a large drop in commercial and industrial loans. **Figure 2** shows the composition of loans held by commercial banks. The most noteworthy trends include the percent of commercial and industrial loans and mortgages. Offshore banks took a lot of business in commercial and industrial loans. American commercial banks turned to more risky markets like mortgages to compensate for these losses.

Figure 2: Composition of Commercial Bank Loans[a]



[a]Fractions of total loans.

Source: Boyd and Gertler 1993

The work of McCauley and Seth (1992) shows this uptick in offshore banking loans was due to low reserve requirements overseas. These low reserve requirements allowed offshore banks to target large companies and offer them better lending rates than American banks. Boyd and Gertler (1993) believe American banks, being unable to compete with offshore banks in the commercial lending market, used their vast amounts of data to trade in less liquid more volatile assets. Large bank's access to data allowed them short term success in this risky market. Boyd and Gertler (1993) believe American banks become more incentivized to take on risky loans as laws changed. They believed banks had too many safety nets for dealing with this much risk.

Boyd and Gertler (1993) also comment on the growing trend of off-balance sheet activities. From 1979 to 1991 fee income as a percentage of bank assets doubled. One of the largest contributors to the rapid increase in off-balance sheet activities is the derivative market. While the high risk of this market is enough to make regulators weary, the main concern is how interwoven the largest derivative trading banks are. Not only are their actions tied to each other, but only 7 banks control most of the derivative activities in America. This makes predicting the failure of banks a critical task. One of these large banks failing could bring the entire market with it.

3. Literature Review

The following section looks to cover some of the literature surrounding the comparison of random forests and neural networks and prediction of bank failure using various modelling techniques. It begins by discussing some of the trade offs between random forests and neural networks. Literature surrounding common methods of predicting bank failure is reviewed next. The literature review concludes by talking about works that have used machine learning to predict bank failure.

3.1 Random Forests and Neural Networks

Neural networks and decision trees have been compared in many different fields. They are both adept at solving classification problems. Work by Jacobsen, Perner, and Zscherpel (2000) aimed to explore the practical differences between neural networks and random forests by looking at welding data. The goal of the algorithm was to determine if a defect was present in a welded seam. Tests were run in two different ways. One consisted of data extracted from the photos to act as inputs. The other test was run with the photos directly used as input. Both classification options had different tradeoffs in simplicity. It is more complicated to interpret a neural network than random forest, but the random forest was only as accurate as the neural network if the data were preprocessed and extracted from the pictures. Neural networks worked better with the pictures, because analyzing a picture requires many features in the model. Neural networks easily handle data with lots of dimensions. (Demyanyk and Hasan 2010). Some data scientists have attempted to bridge the gap in neural network interpretation by translating a neural network into a random forest. Hinton and Frosst (2017) describe a methodology for training a random forest that could represent a neural network. To do this, researchers used the results of a neural network to train a random forest. Overall, they were able to break down filters within a neural network to better understand how decisions were being made, but the random forest failed to be a replacement for the classification ability of the neural network.

3.2 Common Bank Failure Predictors

Financial ratios are one of the most important places to start when trying to predict bank failure. Beaver (1966) showed the usefulness of financial ratios to predict bank failure. One of the most important conclusions of his work is how the usefulness of accounting data can be measured through its predictive ability. Beaver (1966) was only able to create predictive models utilizing one ratio at a time. He commented on how a multi ratio approach could lead to more accurate results. Many different scores have been created to represent a combination of accounting data. Z scores have been used to view the stability of a bank. Chiaramonte, Croci, and Poli (2015) wanted to understand how the Z-score compared to the CAMELS rating system. They took a sample of European banks from the years 2001 to 2011 and used probit models with these factors to predict a bank's likelihood of failure. They found both rating systems were good predictors without one being clearly better. The researchers did note how Z scores can be less data demanding and help be a better predictor as the bank becomes more sophisticated. Large commercial banks tend to fall under this category.

3.3 Making Models for Predicting Bank Failure

Demyanyk and Hasan (2010) have taken a high-level view at the different methods used to analyze bank failure. Their research shows that neural networks are the predictive model of choice when researchers opt to use machine learning. Other machine learning techniques mentioned in the article tended to be improvements on neural networks. Tam and Kiang (1992) looked to compare neural networks against k nearest neighbor, logistic regression, and other predictive models not based on machine learning. While neural networks tended to perform better in prediction and offer less interpretability as expected, some other concerns and positive features were mentioned. The researchers noted how the lack of standardization in neural networks can leave more guesswork in the process than those using the models may like. On the positive side, the researchers mentioned how neural networks easily absorb new information. As financial data is created the network can keep learning and be more robust in the future. Martin (1977) tried to use logit models to predict bank failure. He noted how the model's accuracy depended on the financial climate. Claiming that it could give the best results during periods of light financial strain. Ideally, the model would work in many different climates. Le and Viviani (2018) were able to make a model with more consistent predictions. They also tried comparing many traditional prediction models and machine learning techniques. They noticed neural networks and k nearest neighbors performed the best. They also suggested these models could be used in various financial climates. The researchers did not create perfect models, but they noted their models only used accounting data and not market data. The addition of this data could lead to more robust models.

4. Methodology

The overall goal of this paper is to compare the usage of different machine learning techniques by predicting bank failure. This is done by gathering banking data from 2010-2016 and making seven separate models for each type of machine learning technique. Each model uses the information from one year to predict if a bank is still open in 2016. Since three separate models were used with this dataset, this section is broken up in the following way. The section begins by talking about the data used in the model. Next, the methodology for the logit model will be explained. After, details will be given on the random forest used for prediction. Finally, parameters used in the neural network are described.

4.1 Data

The data for this model was gathered through the Compustat annual bank fundamentals dataset. Because the work of multiple researchers has shown the predictive power of financial ratios, 10 financial ratios were used in the prediction process (Le and Viviani 2018; Beaver 1966). These ratios included loan loss provision/net interest revenue, tier 1 capital ratio, total capital ratio, equity/total assets, equity/deposit & short-term funding, equity/liabilities, capital funds/total assets, capital funds/deposit & short-term funding, capital funds/liabilities, and net interest margin. Unfortunately, the work of Le and Viviani (2018) was able to utilize over 25 financial ratios, but datasets this extensive required paid membership to access. The dependent variable used in all models is the binary active status coded 1 if a bank is still active and 0 if a bank has closed. The Compustat dataset had information from around 400 banks per year over the seven years (2010-2016) I wanted to measure. This had to be trimmed down to 253 banks per year, because the original data set was heavily biased towards active banks. I removed some of these entries randomly to not skew the dataset. Only 52% of the banks were still active in this new dataset. A train-test ratio of 80% was used across all models. A threshold of 0.5 was also used across all models.

4.2 Logit Model

STATA was used to create a logit model to represent the banking data. While logit regressions can be used with traditional statistical models, using machine learning to estimate the bias parameter lambda can help increase predictive power. Since the goal of this research is to compare both human readability and predictive power, the model estimated lambda to minimize EBIC. Minimizing EBIC has multiple benefits. The first being time. Because lambda is estimated the model solves in a relatively short period of time. Unlike k fold cross validation which can take longer as a dataset grows. Estimating lambda in this way also allows the researcher to easily plot lambdas' effect on different variable coefficients. This information can show the relationship between different variables and the model.

4.3 Random Forest

The random forests in this paper were grown using the fastai library for python. A random forest was chosen because it can avoid many of the issues associated with traditional decision trees. Single decision trees tend to overfit. Instead of a single tree overfitting the data, many trees made of a subset of the data can help produce enough random error to offset some of this overfitting. This model takes slightly longer to solve because each forest requires 5000 trees to be grown. Each individual tree was created by sampling 95 random banks and randomly splitting leaf nodes until each node could no longer be split into 2 nodes with more than five observations.

4.4 Neural Network

Fastai was also used to create and train the neural network. A neural network has many features to consider that may affect its implementation. One of the most significant factors is the number of hidden layers and the nodes within them. The networks used in this research were two-layer models with ten nodes in each layer. The loss function is also a key defining factor. This is the equation that tells the function how incorrect its predictions are so it can accurately train itself during each cycle. The loss function chosen for this network is 1-Sigmoid(prediction). Each network was trained on the entire dataset 250 times. The learning rate chosen for this process was 0.01.

5. Empirical Results

It was found that each model had similar predictive accuracy even though the random forest approach did slightly better than the logit model or neural network. Unfortunately, no model was able to predict bank failure more accurately than the work of Le and Viviani (2018). This was the expected result. The models in this paper are trained using significantly less data.

The more important aspects to analyze are the aspects of interpretability. It was revealed that random forests were using all the data to make its predictions, but only 3 or 4 variables had a strong effect on the prediction in the logit model.

Result 1: *Predictive accuracy across all models was low, but the random forest approach had the highest accuracy the most and the lowest accuracy the least.*

Figure 3 shows the predictive results of the model. Numbers on the left are results from the training set while numbers on the right are from the test set. As expected, the shallow dataset used to train the model produced poor results. The work of Le and Viviani (2018) produced models with a correct percent above 70. The highest percent correct in the test set was predicted by a random forest for the year 2011. It was able to accurately classify 69.7% of the test set. The random forest model produced the best test set accuracy for 3 of the 7 years tested. The neural network and logit model had the best accuracy for 2 years. It should also be noted that the logit model gave the worst test set accuracy for 4 of the 7 years tested. The random forest model only produced the worst test accuracy once.

Figure 3: Percentage of Classifications Correct, Type I Error, or Type II Error Predicting the Active Status for the Year 2016

	2010	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	65.6//54	26//34	8.4//12
Random Forest	83.8//54.6	10.1//33.3	6.1//12.1
Neural Network	86.1//63.6	8//15.9	5.8//20.5

	2011	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	61//44.8	31.1//41.4	7.9//10.3
Random Forest	86.2//69.7	7.5//15.2	6.25//15.2
Neural Network	78.8//45.3	15.2//40.5	6//14.3

	2012	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	60.2//51.4	29.5//40.5	10.2//8.1
Random Forest	84.3//50	9.1//33.3	6.7//16.7
Neural Network	79.4//51.3	12.2//17.1	8.1//31.7

	2013	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	61.3//34.1	26//41.5	12.7//24.4
Random Forest	86.8//61.7	8.6//21.3	4.6//17
Neural Network	81//56.6	5.3//7.5	13.6//35.8

	2014	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	61.2//56.3	21.9//25	16.9//18.8
Random Forest	81.3//43.1	10.7//36.4	8//20.5
Neural Network	69.7//45.2	18//22.6	12.4//32.1

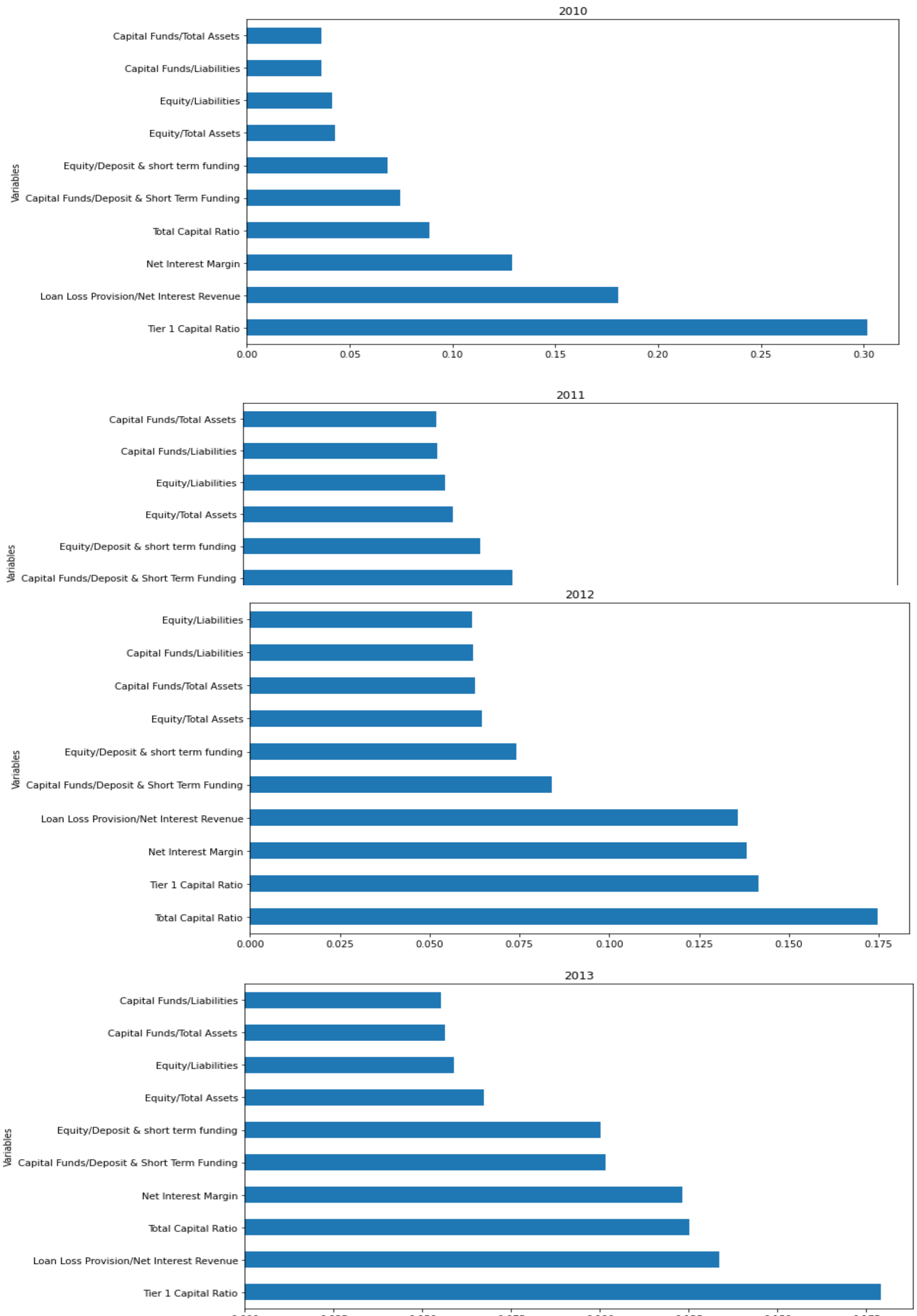
	2015	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	58.2//61.4	18//11.4	23.7//27.3
Random Forest	85.2//62.8	5.6//4.7	9.2//32.6
Neural Network	77.8//54.5	9.3//25	12.9//20.5

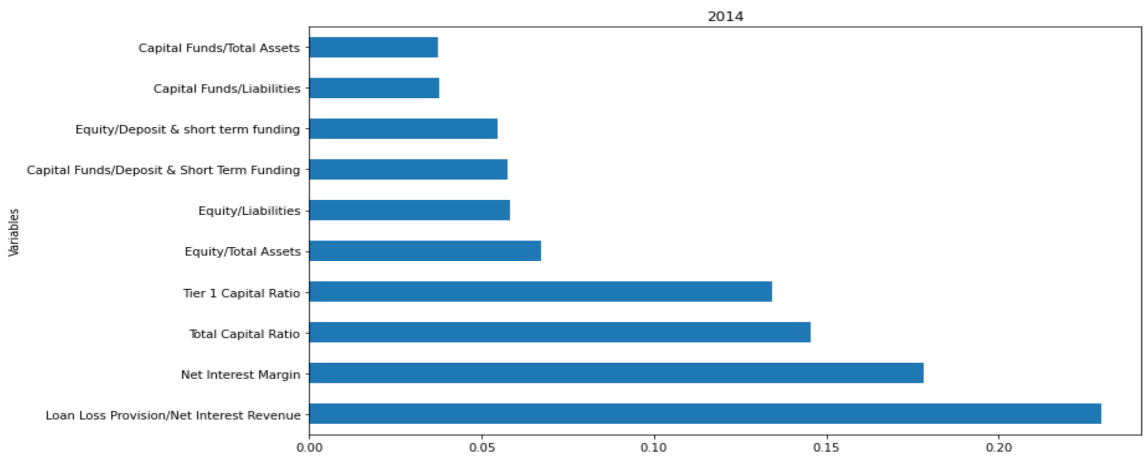
	2016	Key: Training Set%//Test Set%	
Model	Correct	Type I Error	Type II Error
Logit	57.5//52.6	27.2//26.3	15.8//21.1
Random Forest	81.5//55	10//15	8.5//30
Neural Network	70.5//60	10.5//10	18.9//30

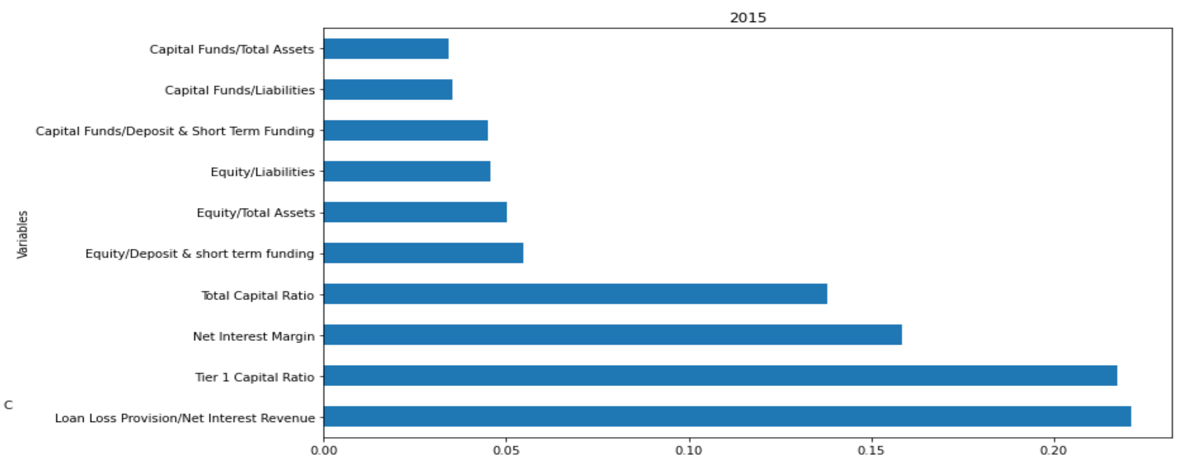
Result 2: *All features had some effect on the prediction results of the random forest model.*

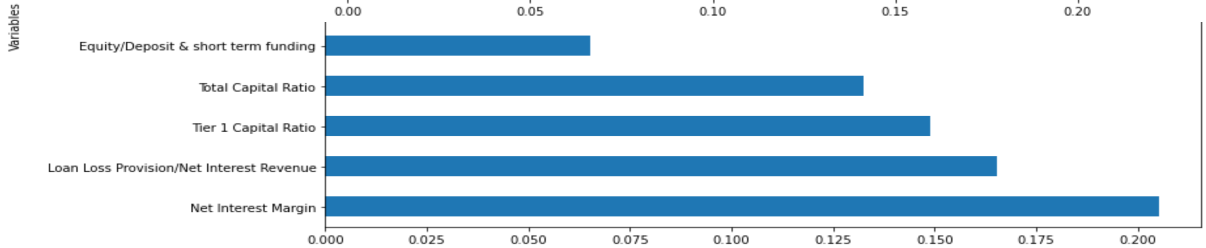
Another goal of this study was to evaluate the interpretation of each model. Unfortunately, very advanced techniques are needed to evaluate the methods of a neural network so that will be viewed as a black box within the scope of this study. **Figure 4** shows the importance of each independent variable in the random forest model. This is calculated by solving for the change each variable contributes to the results of each tree within the forest. All this change is added together and normalized to show importance. A variable of more importance has a larger effect on the results within the model. **Figure 4** charts the relative importance of each variable. Total capital ratio, tier 1 capital ratio, net interest margin, and loan loss provision/net interest revenue were the most important within every random forest. The other 6 variables trailed behind. An important aspect to note across all models is how all variables are used within the models. This shows that the model is using information from every variable to help make predictions. This is in contrast with the interpretation of the logit model which will be shown in result 3. Result 3 reveals only few variables are used to make the predictions. This is most likely why the logit model held the lowest predictive accuracy across the most years.

Figure 4: Relative Importance of Each Variable Across All Random Forests







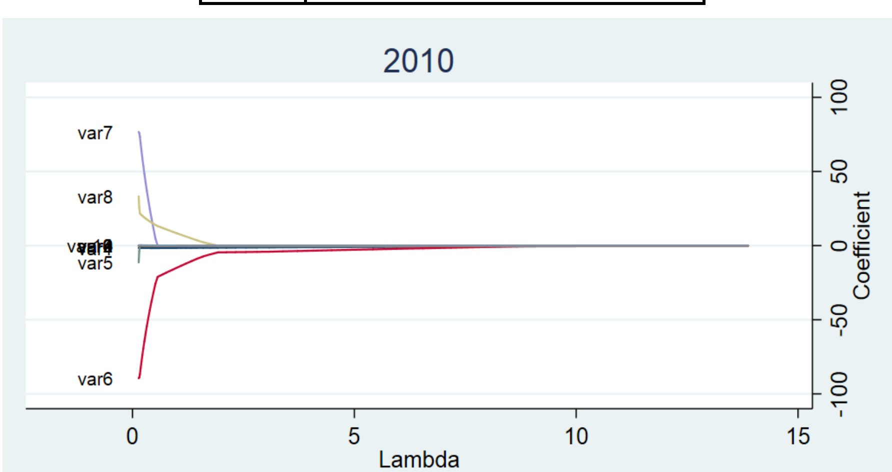


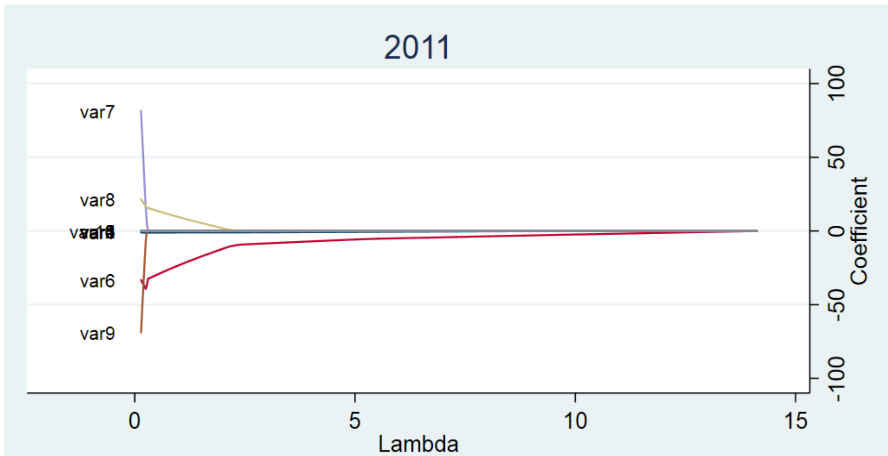
Result 3:

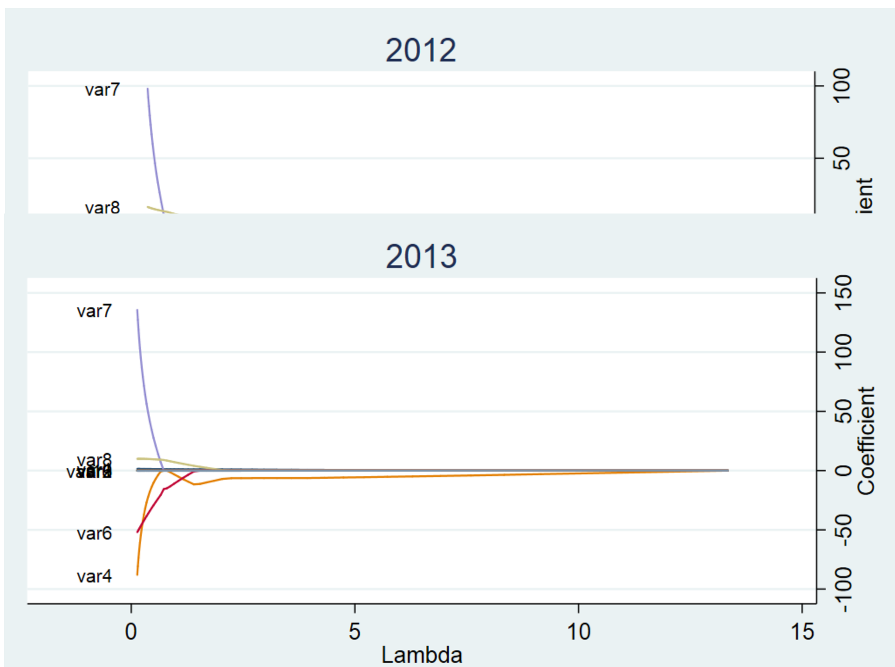
Figure 5 show lambda's effect on the coefficient of variables. Lambda is a bias factor that helps adjust the logit model. As the model becomes more biased the effect of some variables is lessened and their coefficients move towards 0. The further away from 0 a coefficient is or the slower it moves towards 0, indicates the variable plays a stronger role in predicting the active status of the bank. Since the variable names are very long, they are shortened to numbers. The key at the top shows which numbers represent which variables. Capital funds/total assets, capital funds/deposit & short-term funding, capital funds/liabilities, and net interest margin tended to be the variables of interest within the logit models. By looking at how the logit and random forest models function we can see how similar their prediction techniques are. For example, of the 4 most important variables to both modeling techniques only net interest margin was shared by both. This means the models extracted information from the data differently. The same variables had different effects within each model. One benefit of looking at the effect of lambda against the importance of each variable in the random forest is the coefficient reveals if the relationship between the independent variable of interest and the dependent variable is positive or negative. Capital funds/total assets and capital funds/deposit & short-term funding both had positive coefficients across all models. This makes sense considering a banks lack of capital funds would make it insolvent. Equity/liabilities consistently had a negative correlation. This suggests financing activities through debt is generally safer than financing through equity.

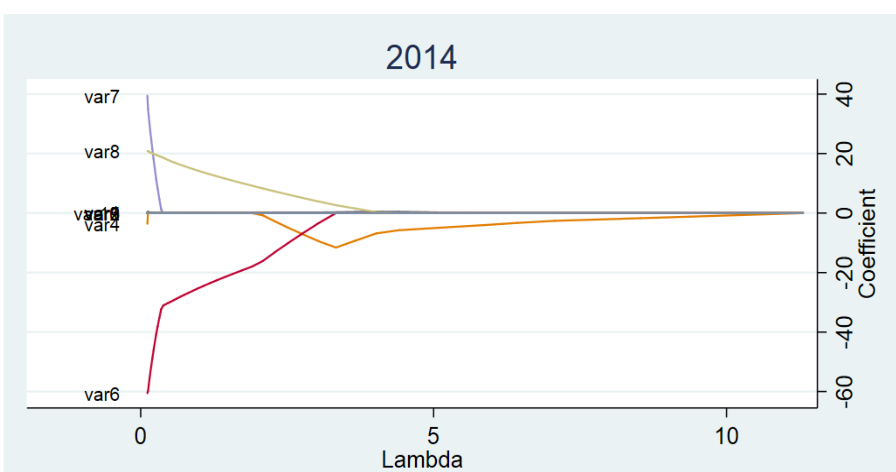
Figure 5: Relationship Between Variable Coefficient and Lambda Across all Logit Models

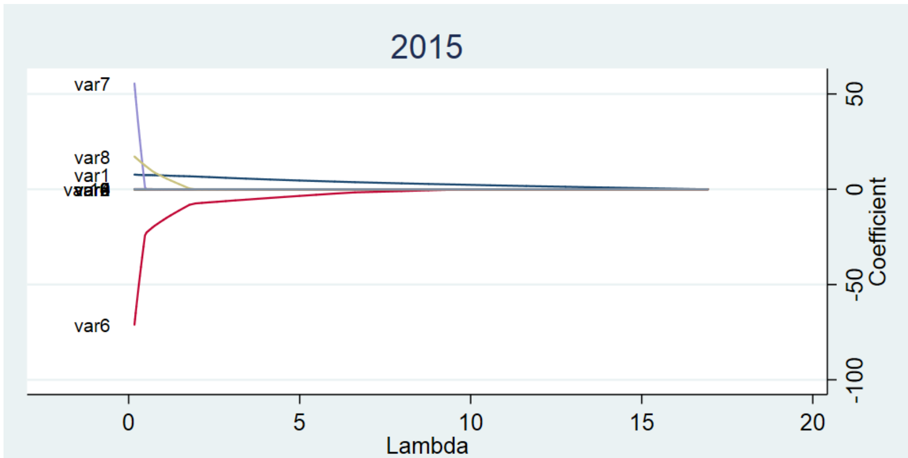
Placeholder	Variable Name
var1	Loan Loss Provision/Net Interest Revenue
var2	Tier 1 Capital Ratio
var3	Total Capital Ratio
var4	Equity/Total Assets
var5	Equity/Deposit & Short-Term Funding
var6	Equity/Liabilities
var7	Capital Funds/Total Assets
var8	Capital Funds/Deposit & Short-Term Funding
var9	Capital Funds/Liabilities
var10	Net Interest Margin

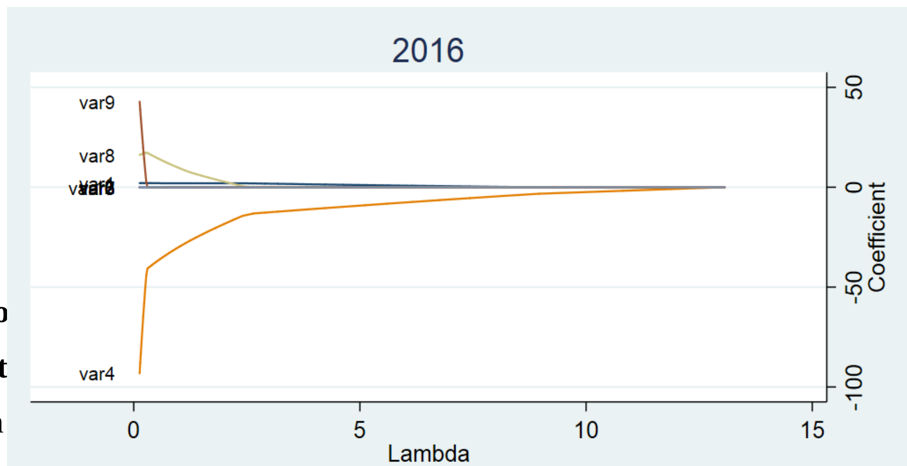



2011
var7
var8
var6
var9
100
50
0
-50
-100
Coefficient
0
5
10
15
Lambda


2012
var7
var8
100
50
2013
var7
var8
var6
var4
150
100
50
0
-50
-100
Coefficient
0
5
10
15
Lambda


2014
var7
var8
var4
var6
40
20
0
-20
-40
-60
Coefficient
0
5
10
Lambda





6. Conclusio

6.1 Applicat

Even cy, the small dataset used to train the models showed that neural networks did not have a significant advantage in predictive ability against the random forest. This finding suggests that random forests should be a more commonly used model not only for bank failure prediction but any prediction problem with low dimensionality. If a researcher's goal is to understand how the dimensions affect the outcome, a random forest it is much easier to interpret than the neural network. A random forest may also be a good pick for a dataset with high dimensionality because researchers can view the importance of the dimensions to understand which ones should be removed. Interpreting the results from the logit model showed the models consistently relied on less than half the given independent variables. I believe a logit model will lose predictive power as dimensionality grows. The random forest model on the other hand, utilized all independent variables to make a prediction. This is probably why the random forest model tended to have better accuracy than the logit model. While this research supports the idea that a random forest model is best suited to research this problem, logit models and neural networks can have some advantages. Logit models are the easiest to create and tweak. They have very few options to adjust so a researcher can realize quickly if they need to gather more data or just change the model type. Another positive feature of logit models and neural networks is that they give prediction results and train quickly. If a regulatory organization wanted to create an application for predicting bank failure that could be trained on incoming datasets a logit model or neural network could quickly incorporate this new information and give output without any lag for those using the application. In general, a neural network may be better suited for this task because incoming data might carry new dimensions and increase the complexity of the dataset. Otherwise, a random forest seems to offer the most merits for research.

6.2 Shortcomings

The biggest challenge in this research was acquiring a robust dataset. Unfortunately, many banking datasets are behind premium memberships and paywalls. As a result, I could not access the same information used to train the models in the work of Le and Viviani (2018). The first issue this caused was a lack of data for the models to train on. Not only did the dataset in the work of Le and Viviani (2018) contain over 3000 observations, but it had over 10 more dimensions than the dataset used to train the models in this paper. This extra dimensionality would make the comparison better by seeing how well these models can predict as the information is scaled up. Having this dataset could have also offered more insights because the research in this paper would be able to focus on just creating a random forest to compete with the neural network and logit model made by Le and Viviani (2018). This would have allowed for much more intricate tweaking into more minute aspects of the model. Even though this paper was not able to compare fully optimized models to each other, time to get results is an important factor when working with machine learning. Random forests were able to offer strong results with minimal adjustments while the creation of neural networks even for simple binary classification problems is a nonstandard process. This means neural networks can undergo many different adjustments to the nodes, hidden layers, loss function, and other features. As a result, I only tested one loss function and used the recommended number of hidden layers from the work of Le and Viviani (2018).

6.3 Future work

Besides just receiving a more robust dataset a few other improvements could be made on this research to better understand the pros and cons of each type of model. The first aspect I would like to explore further is the scalability of the models. The work of Jacobsen, Perner, and Zscherpel (2000) shows that neural networks tend to offer better predictive accuracy than random forests for datasets with high levels of dimensionality. I wonder if the neural network will significantly outperform the random forest as more and more features are added to the model. Another aspect that could be further researched is the use of more variables external to the banks. For example, closures may occur because of certain outside events that a bank with otherwise strong financial indicators may succumb to. Not only could this increase the accuracy across models but interpreting the models could offer new insights. Adding new features unrelated to the features currently being used in the models would result in a different relationship between features. A random forest would be a great model to test these new features on because of its predictive power and easy interpretation. Overall, random forests have proven to be a useful tool for researchers looking for predictive power and an understanding of what features contribute the most to that prediction.

Bibliography

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71. doi:10.2307/2490171

Boyd, J., & Gertler, M. (1993). U.S. commercial Banking: TRENDS, cycles, and policy. doi:10.3386/w4404

Calomiris, C. W., & Mason, J. R. (2003). Fundamentals, panics, and BANK distress during the depression. *American Economic Review, 93*(5), 1615-1647. doi:10.1257/000282803322655473

Chiaramonte, L., Croci, E., & Poli, F. (2015). Should we trust the Z-SCORE? Evidence from the European banking industry. *Global Finance Journal, 28*, 111-131. doi:10.1016/j.gfj.2015.02.002

Demyanyk, Y., & Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods. *Omega, 38*(5), 315-324. doi:10.1016/j.omega.2009.09.007

Frosst, N., & Hinton, G. (2017). Distilling a Neural Network Into a Soft Decision Tree. *Cornell University*.

Haldane, A. G., & May, R. M. (2011). Systemic risk in banking ecosystems. *Nature, 469*(7330), 351-355. doi:10.1038/nature09659

Jacobsen, C., Zscherpel, U., & Perner, P. (1999). A comparison between neural networks and decision trees. *Machine Learning and Data Mining in Pattern Recognition*, 144-158. doi:10.1007/3-540-48097-8_12

Le, H. H., & Viviani, J. (2018). Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research in International Business and Finance, 44*, 16-25. doi:10.1016/j.ribaf.2017.07.104

Martin, D. (1977). Early warning of bank failure. *Journal of Banking & Finance*, *1*(3), 249-276. doi:10.1016/0378-4266(77)90022-x

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank Failure Predictions. *Management Science, 38*(7), 926-947. doi:10.1287/mnsc.38.7.926

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank Failure Predictions. *Management Science, 38*(7), 926-947. doi:10.1287/mnsc.38.7.926