Estimating and Visualizing Banks Failure using Random Forest

Vincenzo Dentamaro *vincenzo@gatech.edu*, **Daniel Cazzaniga** *dcazzaniga3@gatech.edu*, **Paul Livesay** *plivesey3@gatech.edu*, **Thomas Weldon** *tweldon7@gatech.edu*, **Aswin Gigi** *aswingigi@gatech.edu*, **Sharath Kumar Ravi Kumar** *skumar444@gatech.edu*

Abstract

Financial Distress Prediction (briefly FDP) is the problem that aims to predict whether or not a bank will fall into financial distress based on the current financial data through mathematical statistical or intelligent models. The objective of this project is to develop a method for predicting bank failures using Machine learning technique called Random Forest Regressor and visualization techniques to explore similar banks by exploiting their hidden common patterns, a tool to visualize the resulting decision tree and a search tool for non-safe banks visualization all over the country.

I INTRODUCTION

Banks are the bedrock of our modern global economy, when they fail, the consequences can be both widespread and devastating even to ordinary people's life. If methods are developed to better inform the public about the financial outlook of institutions, market forces might act to reward banks with less risky managerial policies. This is important for those who need tools to assess the solvency of financial institutions. This group includes banking consumers, regulators, and business leaders.

After 10 years from the big default caused by big banks in Manhattan, lots of data can be used for FDP problem. In this work, quarterly data was used to create new features[6][7][16][17][1][20]. Innovations of this paper are:

- Accurate FDP predictor on par or superior to the literature reviewed
- UI for searching non safe banks
- The resulting decision-tree[11] for rule based FDP modelling.
- Graph showing similar banks clustered together allowing to explore hidden common patterns among different banks which leads them to be similar. In this way if a bank in the cluster has high probability of failure, also others in the same cluster need further deepening.

At time of writing, it is the first attempt to develop a visual computer aided FDP system making use of ML techniques on real data. The accuracy of the algorithm will be tested in a backtesting environment on the ground truth data using cross-validation, for what concerning the user interface via user satisfaction surveys.
The biggest risks of creating a system that does not perform as good as the reviewed works or that cannot scale on so much data in feasible time have been surpassed. Thanks to this work, banking consumers will have an additional tool to make vital decisions about the safest place to invest their savings. In order to compete for more deposits, banks will need to institute sound management policies and avoid risky behavior.
The paper is organized as follows: section II contains literature reviews, III experimental setup, IV the achieved results, V the data visualization and insights and VI contains the conclusions.

II LITERATURE REVIEW

In 1968 Altman published a method of assigning a Z-score using a linear combination of five financial

ratios with coefficients determined using Multiple Discriminant Analysis (MDA)[2]. In [6] the authors used MDA to classify with 92% of accuracy. The study focuses on creating a very simple tool detecting the first signs of failure. Amadasu[7] also used Altman's Z-score with other features and found contrary to Altman's findings, that working-capital/total-asset, was the biggest discriminator of bankruptcy[7].

In [8] authors found that models taking management quality into account performed better than those that did not with an accuracy of 95%.

In[15] authors developed first, a framework for evaluating signals of early warning models, and second, the estimation and prediction methods. In[16] authors built a statistical model using probability theory able to show that the interest rate on a loan becomes a worse predictor of default as securitization increases. The paper didn't use any machine learning(ML) technique. In[17] proxies for bank-specific CAMELS variables are combined with data on economic conditions concluding that all CAMELS variables are important for prediction accuracy.[18] showed that logit and probit functional may offer an advantage over the frequently used discriminant analysis on FDP with about 86% accuracy.

In[1] authors combined a ANN and self-organizing maps to display the probability of distress up to 3 years before bankruptcy. They used 32 engineered features and used the Federal Deposit Insurance Corporation (FDIC)with an overall accuracy of 96.15%. Work [5] is a survey on FDP. Authors have classified several works with respect to the ML technique used and the features used. It further explains that MDA is a valuable and easy to implement technique, even if it achieves less accuracy compared to SVM,ANN and other data-mining techniques. In [9] the authors proposed a hybrid system that first uses a rough set approach (removes redundant attributes without any information loss) and then a neural network to achieve ~90% in FDP.

In [10] the authors have created a summary of several economics and works that attempt to explain and predict Financial Distress. Authors compared several statistical and ML techniques without concluding which technique is more effective.

In [12] authors performed feature selection using ANOVA and ANN for classification for FDP problem. In [13] authors show the importance of using SVM and its Empirical Risk Minimization criteria for FDP. In[14] the authors in this paper use a hybrid intelligent system, combining rough set approach and RNNs for FDP problem.

In[19] authors showed that ensemble classifiers, outperform the individual models.
As of now, there is no work showing how to visualize motivations that lead to Financial Distress. In [20] authors have shown correlation among natural disasters and Bank's robustness showing that when a disaster happens, the intrinsic stability of local banks is affected.

I. III EXPERIMENTAL SETUP

The dataset is composed of financial indicators for each US Bank from the year 2000 to 2018 every four months, resulting in 547874 instances. The dataset is first shuffled to reduce the time dependent nature of the instances (since the assumption is that instances are I.I.D) and then it was split in two part: training set, which is the 70% of the whole dataset and test set the remaining 30%. as shown in [1].

No feature scaling technique was applied. The labels, namely non fail = 0 and fail = 1 are computed for each instance in the dataset: if the bank under consideration failed, the time difference in months between the report date (the date where indicators values for that bank were issued) and the date that particular bank has failed has been

computed. If the difference in months is greater than *n* months the label is 0(not failed) otherwise 1 (failed). The number *n* is configurable and can be 12, 24 or 36 months.

Since the dataset is very skewed, only few instances have label 1 (failed) with respect to label 0 (non-failed),thus stratified 10-Fold Cross Validation was used to resolve this class imbalance problem.

Two different experiments for FDP problem have been realized. The first experiment was to setup the system as a classification problem. The second was to configure it as a regression problem, where it predicted a continues(scalar) value between [0,1]. The features set is a mix of selected features taken from [1] and [20] plus some engineered features resulting in the following 21 features:

1. Asset Value
2. Change in balance
3. Change on interests
4. Percentage change in net loans and leases
5. Normalized Trade Value
6. Federal Fund sold
7. Normalized Premises and fixed assets
8. Normalized Intangible
9. Normalized Deposit
10. Normalized Deposit Interest
11. Normalized Total Domestic Deposit
12. Liquidity
13. Foreclosure ratio
14. Income earned not collected on loans
15. bank size (log(asset))
16. Wholesale funding over asset
17. Other Liabilities over asset
18. Total equity-capital
19. Percent change in non-current loans and leases
20. Total risk-weighted asset adjusted
21. Volatility of liabilities

For classification AdaBoost was used with 50 Decision Trees built with maximum depth (as pre-pruning parameter) of 20 and minimum split=5. Entropy was used as splitting criteria.

F1 scoring metric has been employed and it is defined as the harmonic mean between precision and recall of all classes weighting each instance equally, therefore an F1 score is best suited for imbalanced datasets like this. For the regression, Random Forest Regressor was used with 50 regression trees and 10-fold cross validation. Since only 0.5% of the training set had banks that failed, we decided to use this as our threshold by classifying any predicted value greater than 0.5% as a failed bank and below as a bank that will not-fail.

II. IV RESULTS

The F1 score on cross validation for the classification scenario was 0.9974 and on test set was 0.9973 which seems to be high, but the confusion matrix in Fig.1 shows clearly that with the test set the system was able to correctly classify all non-failed banks, but the true negative rate is high, over 50% of banks that have failed, were predicted as non-failed. Results clearly show a low specificity rate (TN/(FP+TN)).

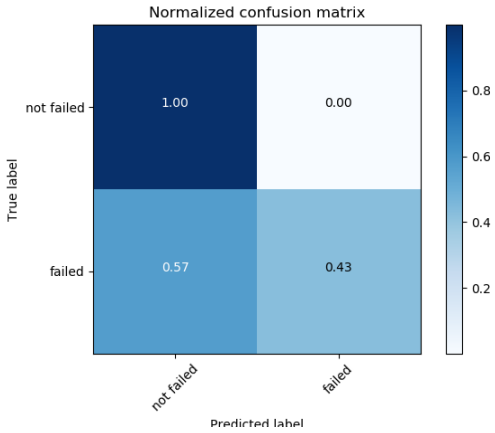


Fig.1 Normalized Confusion Matrix with 12 months parameter as classification problem

Fig.2 shows the confusion matrix results of the FDP as a regression problem setting threshold to 0.005 (0.5%). Both specificity and accuracy are higher with respect to Adaboost for classification, and slightly above the majority of studies reviewed. F1 score is 0.95, thus the solution chosen for the FDP problem is to use an ensemble of random trees regressors. In both cases average training time was about 46 minutes.

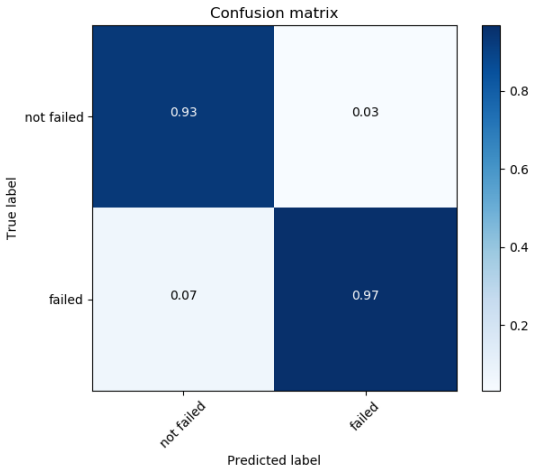


Fig. 2 Normalized Confusion Matrix with 12 months parameter as regression problem

V DATA VISUALIZATION

For the data visualization, three user interfaces have been produced:

1. Clusters visualization
2. Choropleth map of US with possibility to search for banks that more likely will fail in future 36 months.
3. Regression Tree visualization

Clustering

Clusters visualization needs to perform clustering algorithm on data before generating charts. K-Means clustering was used. The number of clusters to be used is selected looking at the elbow curve in Fig.3.

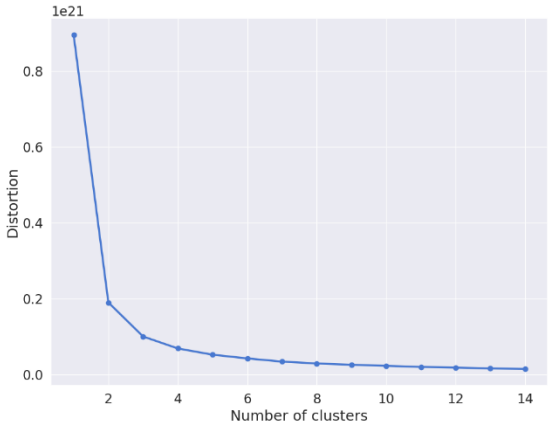


Fig. 3 Elbow curve K-Means

The elbow curve is a method to validate consistency of the variance with respect to the number of clusters. It becomes almost flat if, adding new clusters, it does not allow to better model the data. As it is possible to note from Fig.3 an interesting number of clusters is in the interval [2,4].

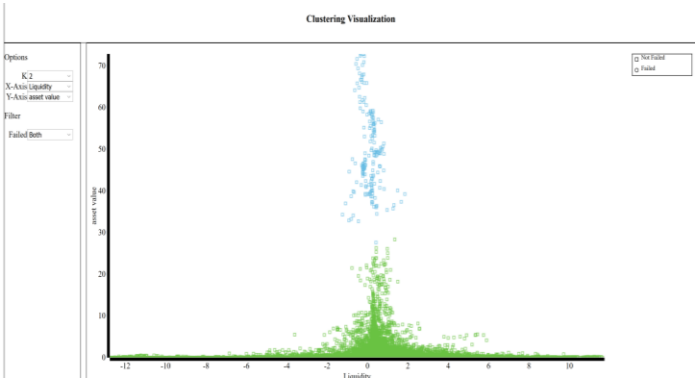


Fig. 4 Clustering with 2 clusters and liquidity vs asset value as axis

Fig.4 shows the plot of two clusters (light blue) and green over Liquidity (x axis) vs Asset Value (y axis). As it is possible to note, the clusters are linearly separable. The result of filtering only on banks that have failed is displayed in Fig.5. As it is possible to note, with such clustering, which is actually the binarization of the FDP problem (failed or non-failed), it correctly clusters all failed banks within one cluster: the green one.

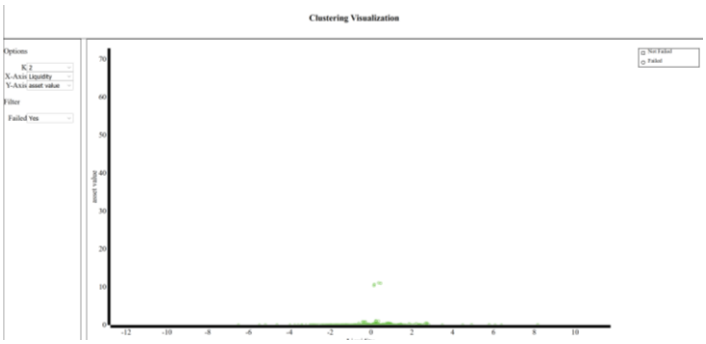


Fig. 5 Clustering with 2 clusters and liquidity vs asset value as axis filtering only on failed banks

Fig.6 shows the clustering applied using 3 clusters. It is possible to observe that dataset is almost linearly separable and there is a distinguishable pattern among data.

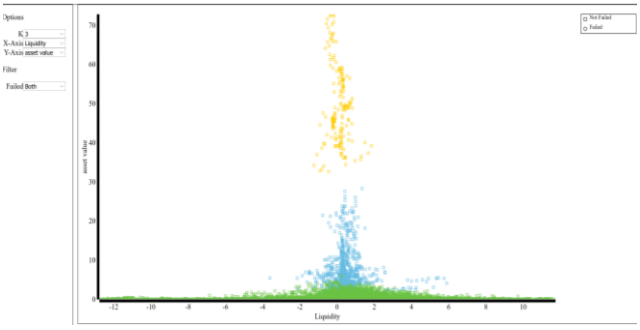


Fig. 6 Clustering with 3 clusters and liquidity vs asset value as axis

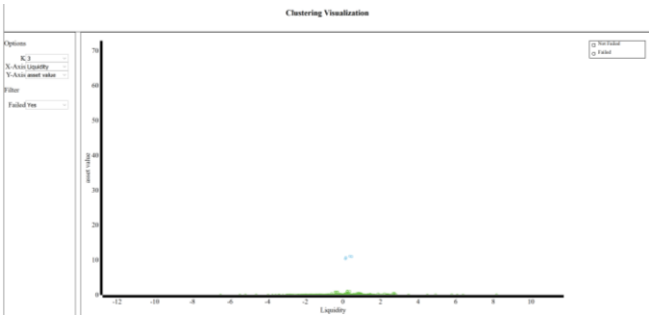


Fig. 7 Clustering with 3 clusters and liquidity vs asset value as axis filtering only on failed banks

Looking at Fig.7 is possible to note that failed banks belongs, for the most part, at the green cluster and only 4 instances to the light blue cluster. None of the failed banks belongs to the orange cluster, showing that this clustering has brought to light an important aspect needed for the correct interpretation of the pattern: banks can be separated in banks that share a common pattern with banks that have failed (green cluster), banks that share a common pattern with lots of non-failed banks and a small minority of banks that have failed, thus are almost safe (light blue cluster) and 100% safe banks (orange cluster). This data exploration is crucial and a key parameter for financial institutions, to understand their safeness level and for savers, which have the opportunity to carefully selects banks in the orange cluster, to take less risks.

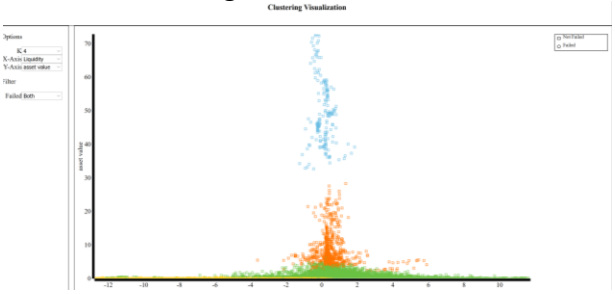
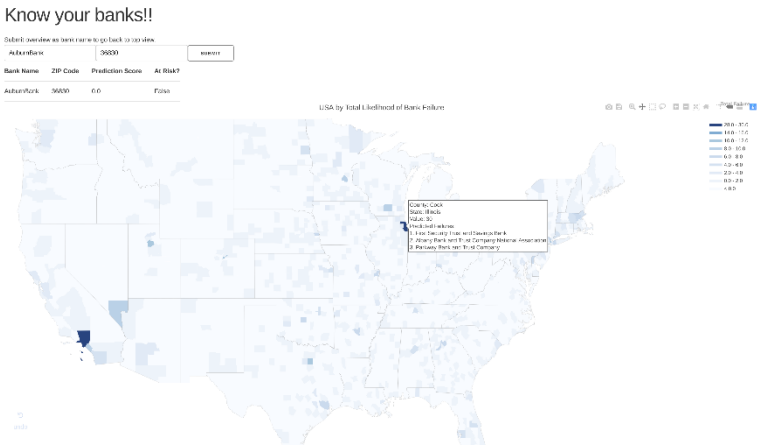


Fig. 8 Clustering with 4 clusters and liquidity vs asset value as axis

Clustering with 4 clusters, as shown in Fig.8 does not really add any new information on pattern analysis, showing that the model interpretation with 3 clusters is, indeed, enough to correctly explain data distribution among clusters.



Choropleth map of US

Fig. 9 Risk meter tool developed for searching and displaying US banks's health

In Fig.9 (and in APPENDIX II with bigger resolution) is represented the screenshot of the developed too used to search Banks and see their health status. The legend shows the total number of banks that have failed per county. The color of each county indicates the number of banks that have failed over a maximum of 30 banks. Each county becomes at more risk if the color moves towards dark blue.

For the selected city (in the example there is Aubum Bank), its prediction score is 0 which means that it is not predicted to be at risk. The threshold value for determining if a bank is predicted to fail is > 0.2.

When the user hovers with the mouse on a city, a box appears showing, for each county, the total number of banks that have failed reporting the top 3. After few seconds the map is resized and the result displayed in Fig.10

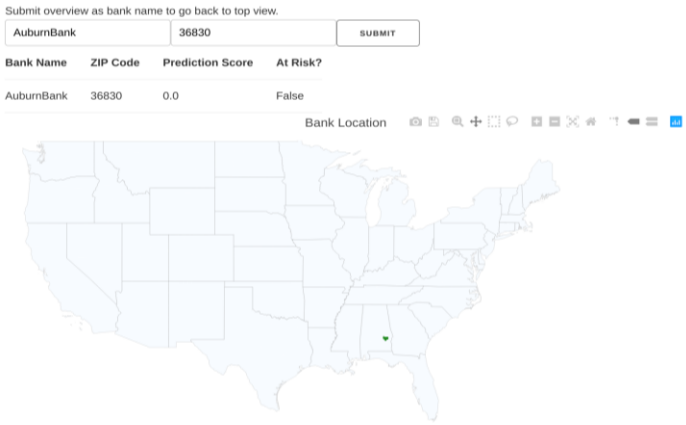


Fig. 10 Choropleth map of the selected bank's county

Fig.10 highlights the county of where the bank is on the map, and since the bank we searched for is safe the county is marked as green, if it was unsafe it would have been red.

Tree Visualization

In APPENDIX I there is the result of the computed tree for FDP problem. It is the first tree of the computed forest. It can be used by financial institutions to create decision rules that with acceptable precision predict automatically, and without using machine learning techniques and additional data, if a bank will fail or not. In particular, if the resulting value is > 0.2, the bank is predicted to fail. If, instead, the resulting value is <= 0.2 the bank is predicted to not fail. Despite all the pre-pruning parameters, the final tree is very wide even if not very deep.

All the UI interfaces have been evaluated by peers and from a verbal analysis it has emerged that the risk meter tool (the choropleth map in Fig.9) needs to be simplified, including also a description on the meaning of total failure. For clustering has emerged that the UI needs to be more responsive, but managing hundreds of thousands of datapoints in a browser is not easy.

VI CONCLUSIONS

This work has demonstrated that by carefully selecting features from several cited works has increased the accuracy of FDP problem to 95% which is above the accuracy of the majority of all other paper reviewed. In addition, risk meter tool and clusters visualization tool help the end user in understanding the health status of a Bank and its connected risk, but also explain the hidden pattern among banks by visually exploring their relation looking at their clusters. Also the choropleth map indicates that Illinois and California have more banks that are predicted to be at risk in 2019. The tree visualization is an additional tool for financial institution that can be used to integrate the computed rules into their risk modelling algorithms and thus notify in case of high risk and potential failure.

Member Contributions

All team members have contributed similar amount of effort.

I. REFERENCES

[1]ITURRIAGA, Félix J. López; SANZ, Iván Pastor. Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. Expert Systems with applications, 2015, 42.6: 2857-2869.

[2]Altman, E. I. (1968), FINANCIAL RATIOS, DISCRIMINANT ANALYSIS AND THE PREDICTION OF CORPORATE BANKRUPTCY. The Journal of Finance, 23: 589-609.

[3]LE, Hong Hanh; VIVIANI, Jean-Laurent. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. Research in International Business and Finance, 2018, 44: 16-25.

[4]TANAKA, Katsuyuki; KINKYO, Takuji; HAMORI, Shigeyuki. Random forests-based early warning system for bank failures. Economics Letters, 2016, 148: 118-121.

[5]SUN, Jie, et al. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowledge-Based Systems, 2014, 57: 41-56.

[6]CLEARY, Sean; HEBB, Greg. An efficient and functional model for predicting bank distress: In and out of sample evidence. Journal of Banking & Finance, 2016, 64: 101-111.

[7]AMADASU, David E. Bank Failure Prediction. AFRREV IJAH: An International Journal of Arts and Humanities, 2012, 1.4: 250-265.

[8]BARR, Richard S.; SEIFORD, Lawrence M.; SIEMS, Thomas F. Forecasting bank failure: A non-parametric frontier estimation approach. Recherches Économiques de Louvain/Louvain Economic Review, 1994, 60.4: 417-429.

[9]AHN, B. S.; CHO, S. S.; KIM, C. Y. The integrated methodology of rough set theory and artificial neural network for business failure prediction. Expert systems with applications, 2000, 18.2: 65-74.

[10]DEMYANYK, Yuliya; HASAN, Iftekhar. Financial crises and bank failures: A review of prediction methods. Omega, 2010, 38.5: 315-324.

[11]ANKERST, Mihael, et al. Visual classification: an interactive approach to decision-tree construction. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 1999. p. 392-396.

[12]ECER, Fatih. Comparing the bank failure prediction performance of neural networks and support vector machines: The Turkish case. Economic research-Ekonomska istraživanja, 2013, 26.3: 81-98.

[13]HÄRDLE, Wolfgang; MORO, Rouslan; SCHÄFER, Dorothea. Predicting bankruptcy with support vector machines. In: Statistical Tools for Finance and Insurance. Springer, Berlin, Heidelberg, 2005. p. 225-248.

[14]AHN, B. S.; CHO, S. S.; KIM, C. Y. The integrated methodology of rough set theory and artificial neural network for business failure prediction. Expert systems with applications, 2000, 18.2: 65-74.

[15]BETZ, Frank, et al. Predicting distress in European banks. Journal of Banking & Finance, 2014, 45: 225-241.

[16]RAJAN, Uday; SERU, Amit; VIG, Vikrant. The failure of models that predict failure: Distance, incentives, and defaults. Journal of Financial Economics, 2015, 115.2: 237-260.

[17]TATOM, John; HOUSTON, Reza. Predicting failure in the commercial banking industry. 2011.

[18]CROWLEY, Frederick D.; LOVISCEK, Anthony L. New directions in predicting bank failures: the case of small banks. North American Review of Economics and Finance, 1990, 1.1: 145-162.

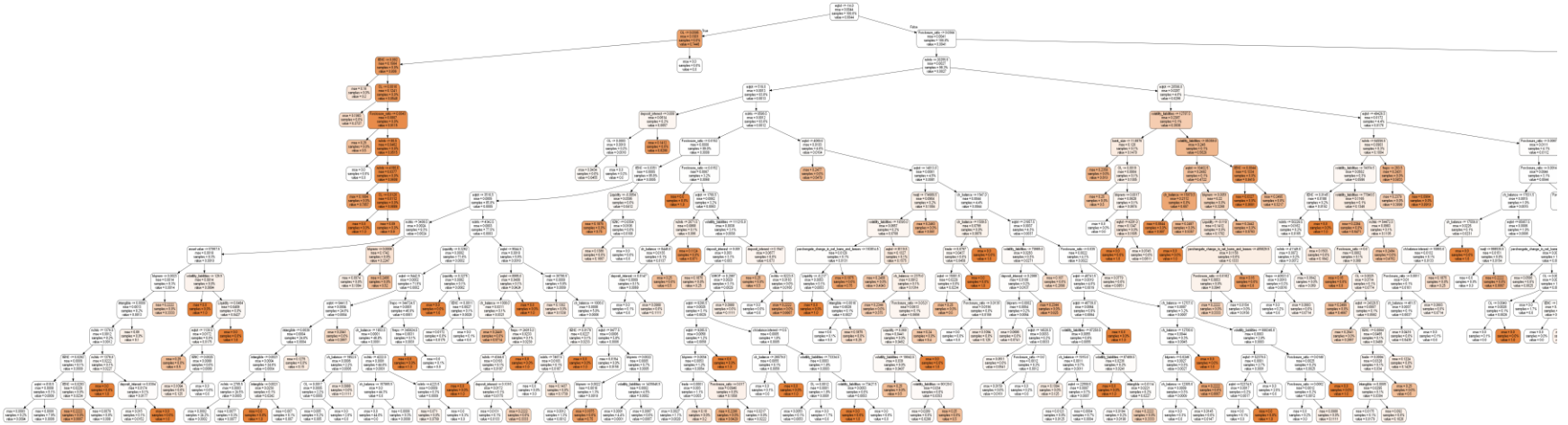
[19]KUMAR, P. Ravi; RAVI, Vadlamani. Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. European journal of operational research, 2007, 180.1: 1-28.

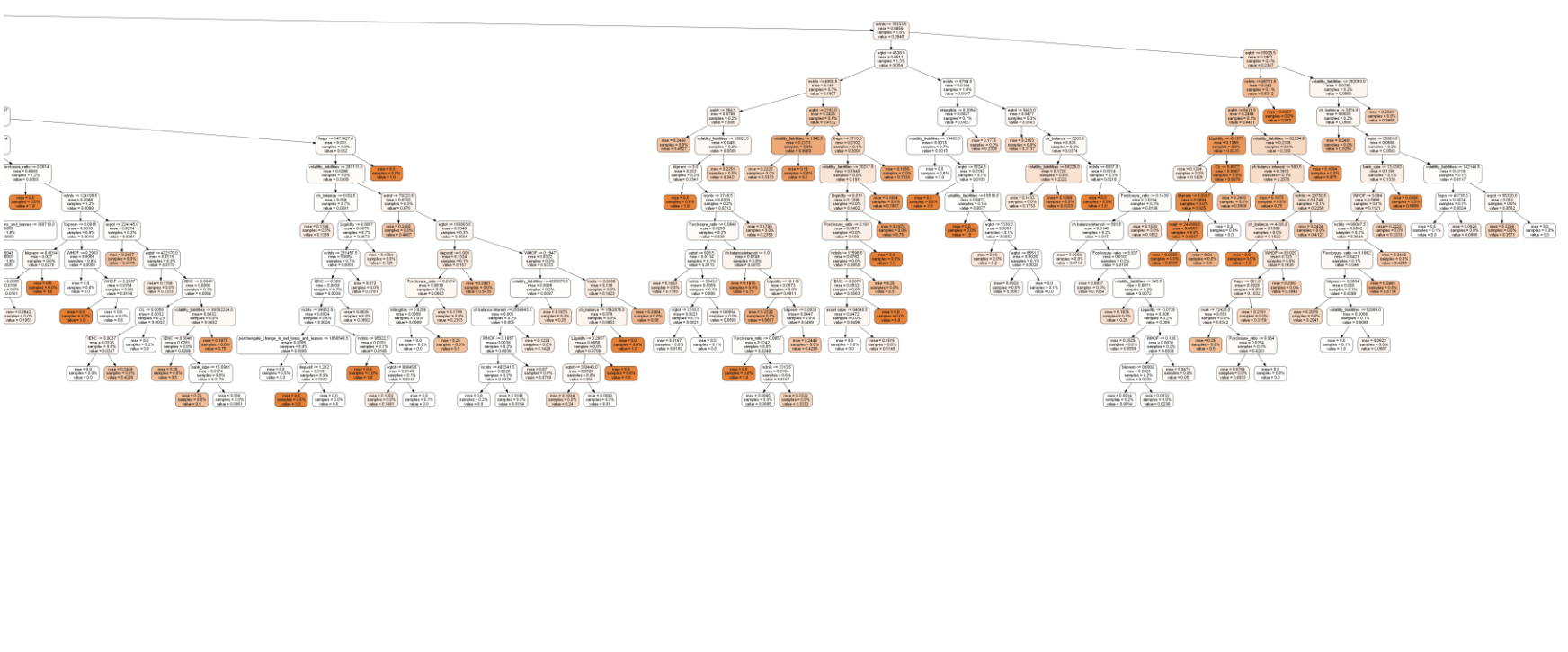
[20]NOTH, Felix; SCHÜWER, Ulrich. Natural disaster and bank stability: Evidence from the US financial system. 2018.

[21]Y. Freund, R. Schapire, “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting”, 1995.

APPENDIX I

Complete plot of the decision tree, divided in two parts: first image represents the tree starting from the left of the root, and the second represents the tree starting from right of the root.





APPENDIX II

Submit overview as bank name to go back to top view.

AuburnBank | 36830 | SUBMIT

Bank Name	ZIP Code	Prediction Score	At Risk?
AuburnBank	36830	0.0	False