HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 4 (2022)

An Optimization of Machine Learning Approaches in the Forecasting of Global Financial Stability

Adyant Ranjan[1] and Guillermo Goldsztein[#]

[1]Seminole High School, USA
[#]Advisor

ABSTRACT

In the current data-driven world, the significance of machine learning as a mechanism for making predictions is vital. This research dives into how supervised learning techniques can be used to predict whether a banking crisis will occur in areas of Africa, which can be generalized to determining the status of financial stability in all areas around the world. By applying different machine learning mechanisms, along with tuning the hyperparameters, the optimal machine learning technique was found to be a neural network with two hidden layers, both hidden layers having the ReLU activation function. These results demonstrate that through widespread implementation of this neural network, governmental and financial organizations can develop significant trends and predict when a state is in economic peril, allowing for sufficient financial, social, or other aid to be administered before situations deteriorate.

Introduction

A recurring disruption in the financial situation of nations are banking crises, characterized when banks are unable to both redistribute assets and gain short-term profits (World Bank Group, 2017). This leads to the inability to secure funds for long-term holdings in an effort to sustain the banking organization. These crises are often a precipice towards complete financial disasters, including the inability to successfully utilize loans, an incline on interest rates, and an eventual recession in the economy. Predicting these banking crises presents a prohibitory effect on these crises from emerging, as knowing when these crises will occur can allow for swift domestic reallocation of funds to protect banks or allow for foreign financial intervention to be administered in an effort to prevent banking crisis prospects from inducing further ramifications on the economy. However, successful predictions are not feasible with traditional methods that consider all associated factors.

Machine learning, a prominent prediction technique overtaking all lenses of innovation, is a subset of artificial intelligence involving the training of predictive models to computationally predict labels. Machine learning has shown promising results in predicting aspects across all areas of study, including the prediction of fraud (Lokanan & Sharma, 2022), the prediction of heart disease (Nagavelli et al., 2022), and even the prediction of political elections (Coletto et al., n.d.). In fact, there has been previous research exploring how financial stability can be analyzed using methods of machine learning and machine learning's implications on the financial sector (Gensler & Bailey, 2020; Alessi & Savona, n.d.; Fouliard, et al., 2020). To complement these previous research endeavors, this research will focus on the application of machine learning techniques to provide an accurate predictor for banking crises in Africa as a means of predicting financial stability in countries around the world, enhancing the field in the physical testing process, as this research describes an approach to simulating the different techniques, comparing these techniques with each other, and adjusting them to achieve the maximum accuracy and minimal overfitting possible.

Methods

Interpreting and Cleaning the Data Set

A typical machine learning data set can be characterized into three different components: examples, features, and a label. For the purposes of this research, "examples" will refer to the different trials of measurement or the instances of something occurring. "Features" refer to the different types of data that are influencing the "label," which is the quantity or quality that one is trying to predict.

The dataset utilized was found on the compilation website Kaggle, but the data is truly derived from Harvard Business School's research on "Global Crises Data" by Carmen Reinhart and associates (Chiri, 2019). There are distinct political, economic, and social factors in the dataset that provide a glimpse as to the factors influencing the emergence of banking crises. The data set chosen has 1060 rows which corresponds to 1050 examples across different countries (from listed order: Algeria, Angola, Central African Republic, Ivory Coast, Egypt, Kenya, Mauritius, Morocco, Nigeria, South Africa, Tunisia, Zambia, and Zimbabwe), different regions, and different time periods. The data set has only binary or numerical features, no qualitative features. These include USD exchange rate (numerical), domestic debt (binary), sovereign external debt (binary), weighted GDP (numerical), annual inflation CPI (numerical), independence (binary), currency crises (binary), and inflation crises (binary). This data set incorporates factors across multiple avenues of study, from politics and social factors in referring to the independence, to financial factors when referring to inflation and GDP, to international economic factors in exchange rate and external debt. These factors culminate in predicting the label, which is whether or not there is a banking crisis.

In order for the data set to be viable for machine learning, cleaning measures must take place. Firstly, irrelevant columns in terms of the machine learning model were dropped. The country name was decided to not be a factor in the machine learning model as the country's name itself should not attest to its banking crisis susceptibility, rather the political, social, and financial climate. Along the same lines, the names of different regions were removed as the name itself is not a key factor in the determination of the label. The case number column was removed as it is irrelevant to the model. The "systemic_crisis" column was also removed as the factor:label correlation is too high to the point of direct causation and would alter the machine learning model by a large magnitude. Secondly, missing values were checked using the dropna() method to create a holistic data set. Lastly, quantitative conversion was used for the label as the data of "crisis" or "no_crisis" is a categorical variable that must be converted. Thus, "0" was set to represent there being no crisis while "1" represented that there is a crisis. After scanning the data, it was evident that no rows were dropped, meaning that there were no missing values in any of the columns and rows.

Preparing the Data for Machine Learning

After the data was preprocessed, computational synthetization of the data was administered. Numpy arrays were established for both the labels and factors using the logic of directly including the labels into the first numpy array of labels and including everything but the labels into the second numpy array of features. The arrays were verified.

Through these numpy arrays, it was evident that 94 entries had a banking crisis while 965 did not have a banking crisis. To balance out the data set, copies were made of the crisis entries so that the final condensed data set, which includes both the training and validation sets, is 940 with a banking crisis and 965 without a banking crisis.

Lastly, this condensed set was separated into both the training and the validation sets. A training set is the data that is used to develop the machine learning model while the validation set will be used as a comparison to the model once developed. The validation set cannot be used in the building of the model, so the condensed set was split so that 75% of the data is the training set and 25% is the validation set. These proportions were administered in a completely random manner. The feature numpy array was scaled using the training set to allow for a standard of comparison for the machine learning model. The label does not need to be scaled as it is the predicted value.

Characterizing the Machine Learning Problem and Evaluation Parameters

This research is a demonstration of implementing supervised learning, a subset of machine learning that has input-output layers and alters the model as more data is analyzed (IBM Cloud Education). A subset of supervised learning are classification problems, where the model filters through and places data into pre-determined categories. In this case, the dataset is a binary classification problem in which the two categories are that there is no banking crisis or there is a banking crisis. For a binary classification problem, the error that needs to be minimized is the binary cross-entropy error (BCE).

In order to deem the effectiveness of the computational model, a classification report of the training and validation sets must be analyzed and compared to each other. The classification report has four major categories that this paper focused on: precision, recall, F1-score, and accuracy. These values utilize four different scenarios for a classification problem which is commonly defined as true positives, true negatives, false positives, and false negatives.

1. True Positive (TP). This corresponds to there being a banking crisis and the model predicting that it is a banking crisis.
2. True Negative (TN). This is when there is no banking crisis, and the model predicts that there is no banking crisis.
3. False Positive (FP). This scenario is when the model predicts a banking crisis when there truly is no banking crisis.
4. False Negative (FN). This occurs when the model predicts that there is no banking crisis when there is a banking crisis.

When defining these different outcomes, recall, precision, F1-score, and accuracy can then be represented mathematically (Kanstrén, 2021).

$$recall = \frac{TP}{(TP + FN)}$$

$$precision = \frac{TP}{(TP + FP)}$$

$$F1 - score = \frac{(2 \times precision \times recall)}{(precision + recall)}$$

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Conceptually speaking, a high recall means that the majority of positive cases are being identified, even if these cases may be negative, while a high precision means that the majority of negative cases are reported, even if some of them may be positive. F1-score and accuracy metrics provide a culmination of these measures.

When analyzing these computational models for effectiveness, an important relationship between the accuracy and the binary cross-entropy error must be established. Although seemingly correlated in concept, each value has clear distinctions and elicit unique meaning in describing the model. The accuracy refers to solely the percentage of classifications that are correct in relation to all classifications, evident through the accuracy formula defined above. The numeric binary cross-entropy error is the identified error that needs to be minimized and is different in that it analyzes how the probabilities correlate with the classification. In this case, if a computational model has an extremely high probability that there is a banking crisis when in reality there is no banking crisis, then the BCE will increase substantially. This investigation considered both the BCE and accuracy when determining the efficacy of each model.

Lastly, a common predicament that arises in machine learning is overfitting. Overfitting occurs when the machine learning program becomes accustomed to the training set and learns the extraneous attributes that distinguishes the training set from the validation set. Because of this, the training set accuracy will increase substantially while the validation set accuracy will not. Overfitting leads to the inability for the computational model to be applied to real-life scenarios that involve different data points. When analyzing each technique, overfitting was analyzed in addition to its overall performance.

Defining the Machine Learning Techniques

Logistic regression is a machine learning technique that utilizes no hidden layers in its prediction and a one node output layer. The output layer is a sigmoid function which outputs a designated number between 0 and 1. This sigmoid function thus presents a probability for each predicted class, and logistic regression is interpreted as if $\sigma(x) < 0.5$, then the label will be "0" for a binary classification problem and conversely, if $\sigma(x) > 0.5$, then the label will be "1".

Conceptually, logistic regression can be defined using this sigmoid function along the basis of parameters and the different factors. The input of the sigmoid function condenses the sum of each of the parameters scaled by its corresponding factor. The sigmoid function then outputs a number which aligns with the binary classification. Mathematically, this can be defined as $\hat{y} = \sigma(\sum_1^n X_n w_n + c)$, where X_n is defined as the "nth" factor found in the machine learning data set and w_n is defined by the parameter of the regression line which is determined by minimizing the binary cross-entropy error of the model.

The other techniques that were analyzed are multiple variations of neural networks. Neural networks are a type of machine learning technique that utilize different hidden layers to map out a path from the input layer to the output layer. There are various hyperparameters of neural networks, which are different parameters of the computational model that are used to implement the methodical development of the machine learning model. One important hyperparameter is the relationship between the input layer, output layer, and a fixed number of hidden layers. The input layer is quantified as "Layer 0," the output layer is defined as "Layer 'L,'" and the number of hidden layers is $L - 1$ for a total of L layers. Additionally, the input layer will have the number of nodes (also referred to as neurons) equal to the number of features in the data frame while the output layer will have a number of nodes equal to the number of labels. The last hyperparameter to note is that for a binary classification problem, the output activation function must be the sigmoid function ($\sigma(x)$).

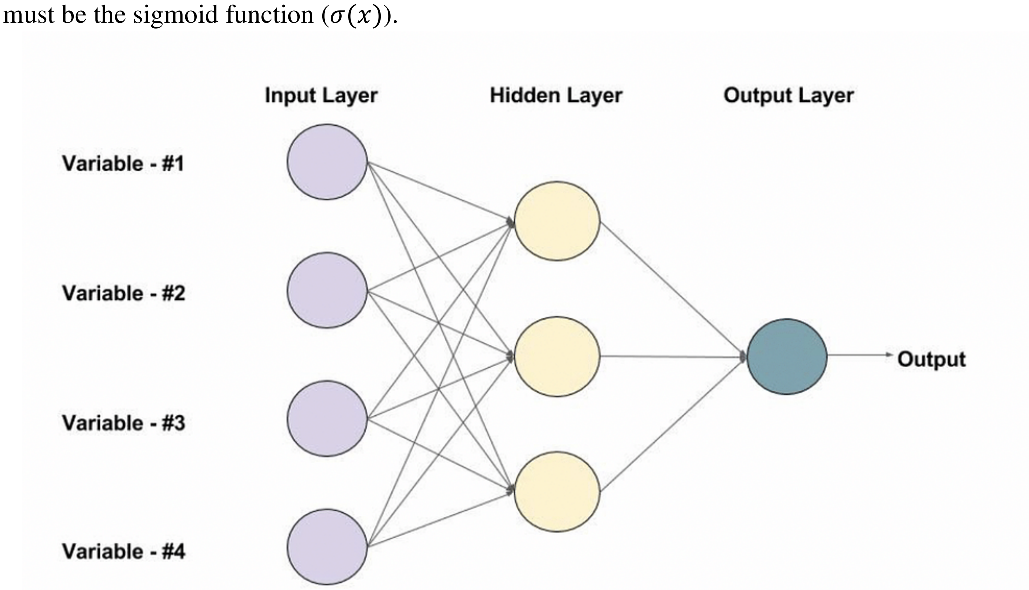


Figure 1: An Illustration of a Neural Network (Gupta, 2021)

This research focuses on investigating the number of hidden layers that are appropriate for this data analysis. For this data set, the number of nodes was held constant for each corresponding layer that is kept, while the number of hidden layers changed. As this is a binary classification problem, the output activation function is still the sigmoid function. After the techniques were compared with each other, the hyperparameters were adjusted in the second phase of the experiment through the activation function. The sigmoid function remained utilized for the output layer, while the ReLU function and the tanh function were used in combination with each other for the hidden layers. An overview of the mathematical framework of each function is displayed.

Table 1: A Characterization of the Activation Functions Utilized

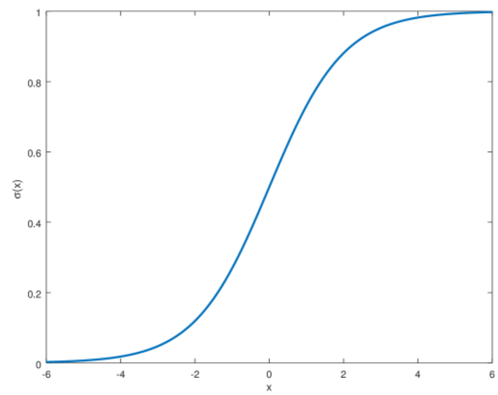
<table>
<tr><th>Activation Function</th><th>Equation</th><th>Graph</th></tr>
<tr><td>Sigmoid</td><td>$\sigma(x) = \dfrac{1}{1 + e^{-x}}$</td><td>

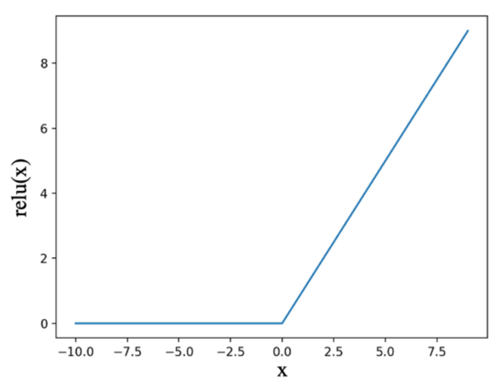
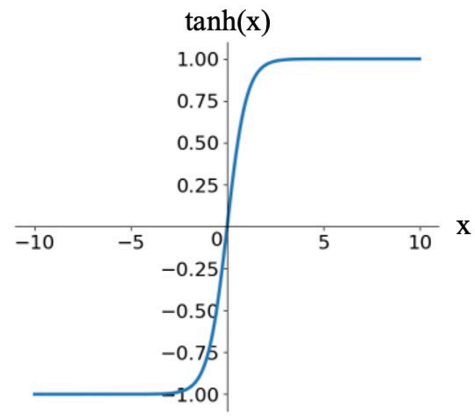
Figure 2: Sigmoid Function Graph (Hvidberrrg)</td></tr>
<tr><td>ReLU</td><td>$relu(x) = \begin{cases} x, if\ x > 0 \\ 0, if\ x \leq 0 \end{cases}$</td><td>

Figure 3: ReLU Function Graph (Brownlee, 2020)</td></tr>
</table>

Activation Function	Equation	Graph
Tanh	$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	 *Figure 4: Tanh Function Graph (Patwari, 2021)*

Results: Applying Techniques and Testing Hidden Layer Effectiveness

Computational models were developed for variations of the machine learning techniques. Each model was instantiated with the specific number of hidden layers, nodes, and activation functions. Each model's error was visualized through the matplotlib.pyplot library and adjusted for the number of epochs (steps). The error graphs displayed have an x-axis of the number of epochs, while the y-axis is the error.

Logistic Regression

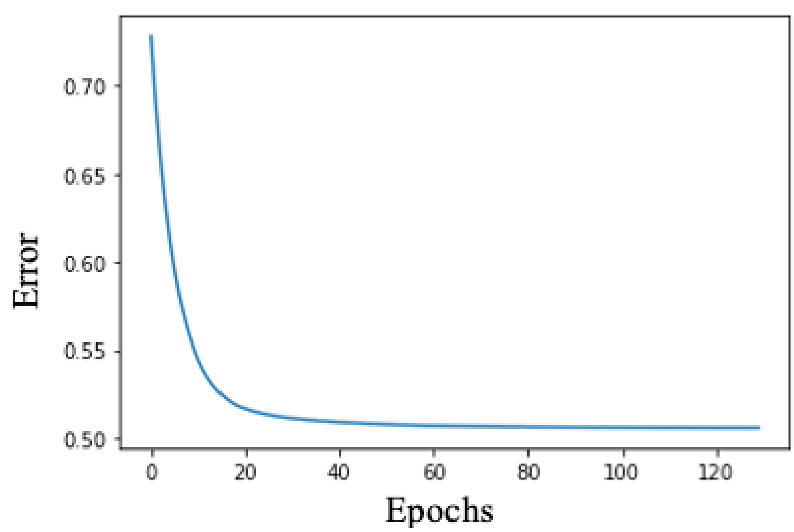


Figure 5: Error vs. Epochs Graph for Logistic Regression

$$Validation\ Set\ BCE\ =\ 0.5104677$$
$$Training\ Set\ BCE\ =\ 0.5055834$$

Table 2: Training Set Classification Report – Logistic Regression

	Precision	Recall	F1-score	Support
No Crisis	075	0.78	0.76	721
Crisis	0.77	0.74	0.75	707
Accuracy			0.76	1428
Macro Avg.	0.76	0.76	0.76	1428
Weighted Avg.	0.76	0.76	0.76	11428

Table 3: Validation Set Classification Report – Logistic Regression

	Precision	Recall	F1-score	Support
No Crisis	0.74	0.77	0.75	244
Crisis	0.75	0.72	0.73	233
Accuracy			0.74	477
Macro Avg.	0.74	0.74	0.74	477
Weighted Avg.	0.74	0.74	0.74	477

Neural Network (One Hidden Layer: ReLU)

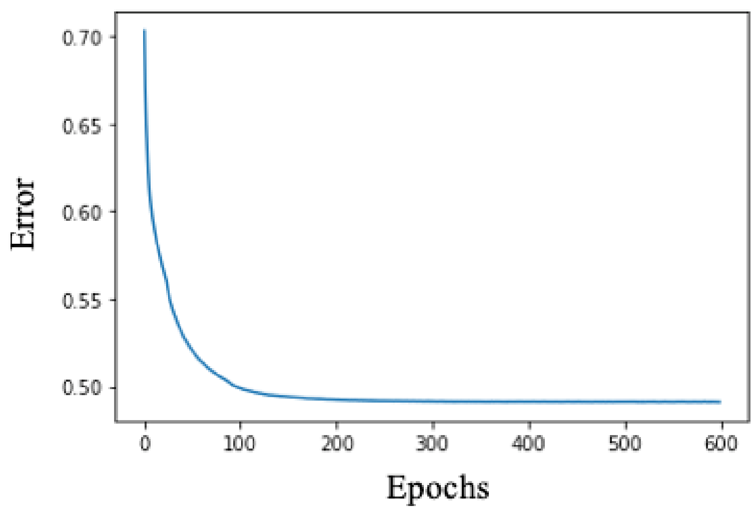


Figure 6: Error vs. Epochs Graph for the One Hidden Layer Neural Network (ReLU)

$$Validation\ Set\ BCE\ =\ 0.49274427$$
$$Training\ Set\ BCE = 0.48543516$$

Table 4: Training Set Classification Report – One Hidden Layer Neural Network (ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.74	0.77	0.76	721
Crisis	0.76	0.73	0.74	707
Accuracy			0.75	1428
Macro Avg.	0.75	0.75	0.75	1428
Weighted Avg.	0.75	0.75	0.75	1428

Table 5: Validation Set Classification Report – One Hidden Layer Neural Network (ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.74	0.75	0.74	244
Crisis	0.74	0.72	0.73	233
Accuracy			0.74	477
Macro Avg.	0.74	0.74	0.74	477
Weighted Avg.	0.74	0.74	0.74	477

Neural Network (Two Hidden Layers: ReLU, ReLU)

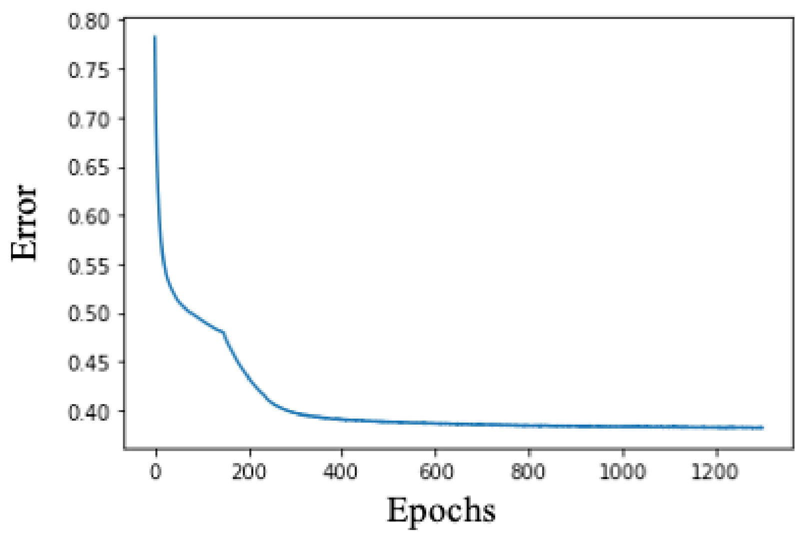


Figure 7: Error vs. Epochs Graph for the Two Hidden Layer Neural Network (ReLU, ReLU)

$$Validation\ Set\ BCE\ =\ 0.3869956$$
$$Training\ Set\ BCE\ =\ 0.3824707$$

Table 6: Training Set Classification Report – Two Hidden Layer Neural Network (ReLU, ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.95	0.76	0.84	721
Crisis	0.80	0.95	0.87	707
Accuracy			0.86	1428
Macro Avg.	0.87	0.86	0.86	1428
Weighted Avg.	0.87	0.86	0.86	1428

Table 7: Validation Set Classification Report – Two Hidden Layer Neural Network (ReLU, ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.96	0.77	0.85	244
Crisis	0.80	0.97	0.87	233
Accuracy			0.86	477
Macro Avg.	0.88	0.87	0.86	477
Weighted Avg.	0.88	0.86	0.86	477

Neural Network (Three Hidden Layers: ReLU, ReLU, ReLU)

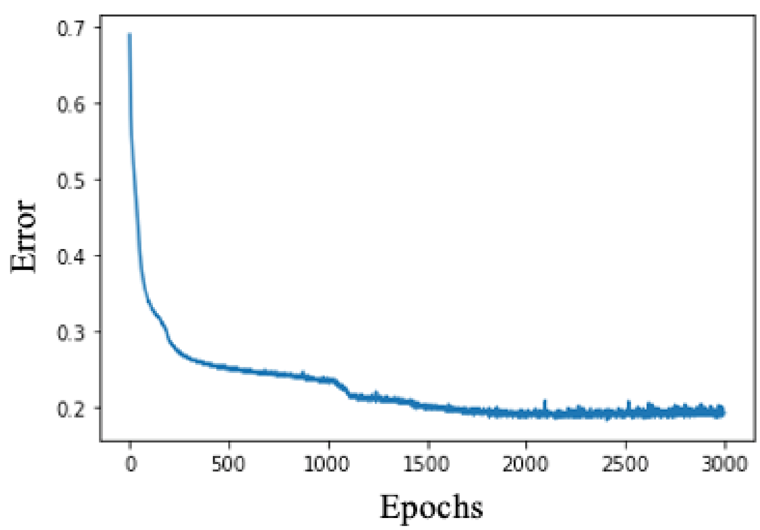


Figure 8: Error vs. Epochs Graph for the Three Hidden Layer Neural Network (ReLU, ReLU, ReLU)

$$Validation\ Set\ BCE\ =\ 0.24401085$$
$$Training\ Set\ BCE\ =\ 0.19012333$$

Table 8: Training Set Classification Report – Three Hidden Layer Neural Network (ReLU, ReLU, ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.99	0.87	0.92	721
Crisis	0.88	0.99	0.93	707
Accuracy			0.93	1428
Macro Avg.	0.93	0.93	0.93	1428
Weighted Avg.	0.93	0.93	0.93	1428

Table 9: Validation Set Classification Report – Three Hidden Layer Neural Network (ReLU, ReLU, ReLU)

	Precision	Recall	F1-score	Support
No Crisis	1.00	0.82	0.90	244
Crisis	0.84	1.00	0.91	233
Accuracy			0.91	477
Macro Avg.	0.92	0.91	0.91	477
Weighted Avg.	0.92	0.91	0.91	477

Results: Testing Activation Functions in a Neural Network of Two Hidden Layers

The two renowned activation functions that were tested in combination with each other were the ReLU activation function and the tanh activation function. With two hidden layers, keeping the nodes constant for each model, there are 4 combinations of activation functions: ReLU and ReLU, ReLU and tanh, tanh and ReLU, and tanh and tanh, where the order of the functions is characterized by their layers, as the first activation function listed is for Layer 1, and the second one listed is for Layer 2. The first combination of (ReLU, ReLU) was already observed in the previous investigation. Error graphs were similarly visualized, and classification reports were developed for the other combinations.

Neural Network (Two Hidden Layers: ReLU, tanh)

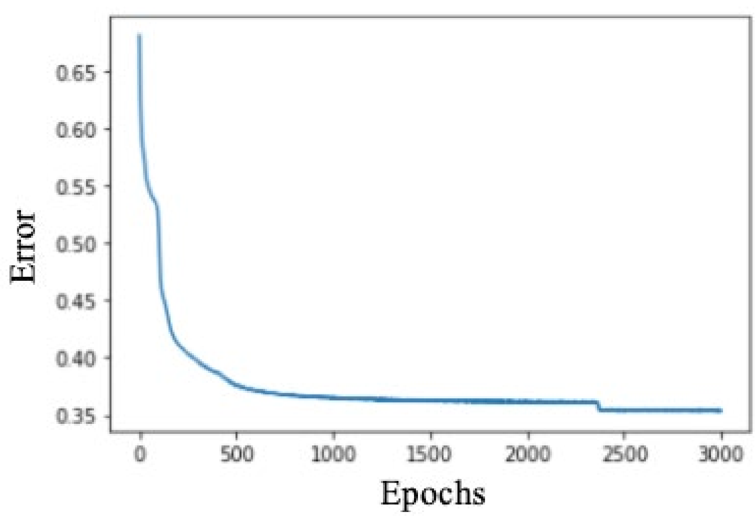


Figure 9: Error vs. Epochs Graph for the Two Hidden Layer Neural Network (ReLU, tanh)

$$Validation\ Set\ BCE\ =\ 0.39182386$$
$$Training\ Set\ BCE\ =\ 0.35286307$$

Table 10: Training Set Classification Report – Two Hidden Layer Neural Network (ReLU, tanh)

	Precision	Recall	F1-score	Support
No Crisis	0.96	0.67	0.79	721
Crisis	0.74	0.97	0.84	707
Accuracy			0.82	1428
Macro Avg.	0.85	0.82	0.81	1428
Weighted Avg.	0.85	0.82	0.81	1428

Table 11: Validation Set Classification Report – Two Hidden Layer Neural Network (ReLU, tanh)

	Precision	Recall	F1-score	Support
No Crisis	0.95	0.69	0.80	244
Crisis	0.75	0.97	0.84	233
Accuracy			0.82	477
Macro Avg.	0.85	0.83	0.82	477
Weighted Avg.	0.85	0.82	0.82	477

Neural Network (Two Hidden Layers: tanh, ReLU)

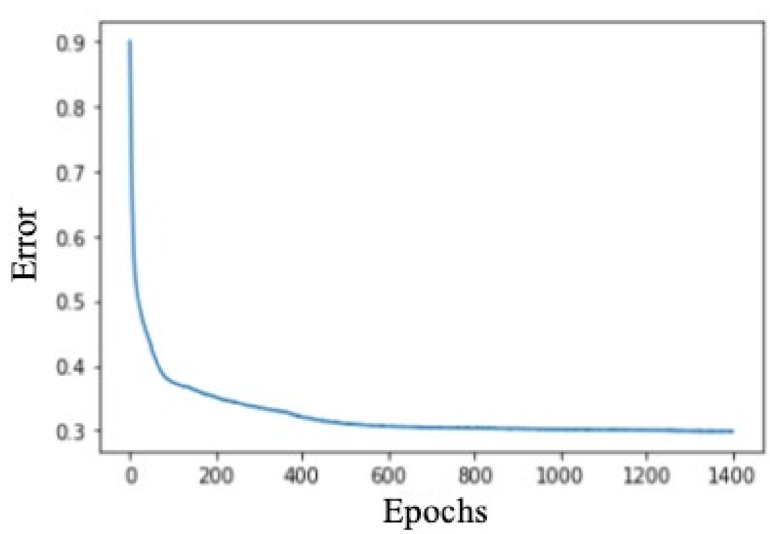


Figure 10: Error vs. Epochs Graph for the Two Hidden Layer Neural Network (tanh, ReLU)

$$Validation\ Set\ BCE\ =\ 0.34254894$$
$$Training\ Set\ BCE\ =\ 0.29801306$$

Table 12: Training Set Classification Report – Two Hidden Layer Neural Network (tanh, ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.95	0.81	0.88	721
Crisis	0.83	0.96	0.89	707
Accuracy			0.88	1428
Macro Avg.	0.89	0.88	0.88	1428
Weighted Avg.	0.89	0.88	0.88	1428

Table 13: Validation Set Classification Report – Two Hidden Layer Neural Network (tanh, ReLU)

	Precision	Recall	F1-score	Support
No Crisis	0.96	0.80	0.87	244
Crisis	0.82	0.96	0.88	233
Accuracy			0.88	477
Macro Avg.	0.89	0.88	0.88	477
Weighted Avg.	0.89	0.88	0.88	477

Neural Network (Two Hidden Layers: tanh, tanh)

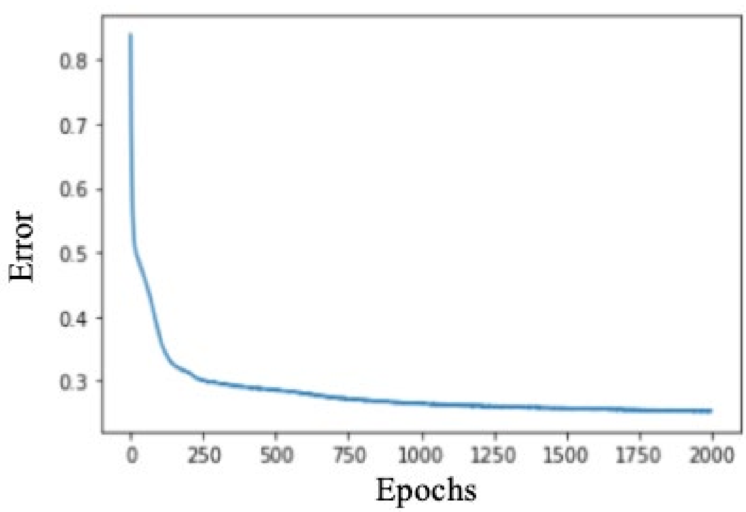


Figure 11: Error vs. Epochs Graph for the Two Hidden Layer Neural Network (tanh, tanh)

$$Validation\ Set\ BCE\ =\ 0.3087736$$
$$Training\ Set\ BCE\ =\ 0.24944979$$

Table 14: Training Set Classification Report – Two Hidden Layer Neural Network (tanh, tanh)

	Precision	Recall	F1-score	Support
No Crisis	0.94	0.87	0.90	721
Crisis	0.87	0.94	0.91	707
Accuracy			0.90	1428
Macro Avg.	0.91	0.91	0.90	1428
Weighted Avg.	0.91	0.90	0.90	1428

Table 15: Validation Set Classification Report – Two Hidden Layer Neural Network (tanh, tanh)

	Precision	Recall	F1-score	Support
No Crisis	0.95	0.86	0.90	244
Crisis	0.86	0.95	0.91	233
Accuracy			0.90	477
Macro Avg.	0.91	0.90	0.90	477
Weighted Avg.	0.91	0.90	0.90	477

Discussion

This investigation involved the prediction of banking crises in different areas of Africa, presenting the role of applied machine learning in the projection of financial stability around the world. Firstly, different machine learning techniques were tested. Logistic regression proved to be a decent computational model with a training set accuracy of 0.76, a validation set accuracy of 0.74, and comparable error values. The minimal overfitting evident meant that more complexity can be established to better train the model. Thus, a neural network was established with one ReLU hidden layer, which decreased the error but had comparable accuracy values to that of the logistic regression model. Even so, more complexity can be administered to the model as both the training set and validation set errors were similar in magnitude.

Next, multiple hidden layers were created to increase complexity. Solely ReLU functions were utilized to prevent the extraneous variable of the different types of activation functions to factor into the results. A two hidden layer neural network with two ReLU functions was established that proved to substantially increase accuracy of both the training and validation sets by 0.11 and 0.12, respectively, and the binary cross-entropy error decreased prominently to ~0.38 for both sets. Two hidden layers proved to be better than logistic regression and a one hidden layer neural network by a large margin, so a three hidden layer neural network was instantiated. This computational model of three hidden layers presents an error that is smaller for both the training set and the validation set, but this measure is implicitly inaccurate in interpretation. Firstly, the differences between the validation set error and training set error is significant (~0.054). Thus, this machine learning model exhibits overfitting as the model is learning the attributes of the training set more than the trends in the data itself, not allowing it to be applied to separate arrays of values. The validation set error still decreased however, despite the model exhibiting overfitting. This is because of the copies evident in the data set. Because copies were made of values that had "1" as a label, the validation set contained copies that are similar to the training set, thus making both sets not completely separate in values. As the computational model learns about the training set more effectively, it learns more about the validation set in the process because of these copies, allowing for a decrease in error for both sets. Thus, despite the accuracy increase due to the copies, there is a large degree of overfitting that renders this three hidden layer model useless towards additional predictions.

Through these observations of logistic regression and different types of neural networks, it is evident that for the dataset, the two hidden layer neural network was the best technique as it allowed for the most accuracy without the dangers of overfitting from substantially occurring. However, another variable emerges in these investigations: activation functions. The prior investigation analyzed neural networks in terms of only the ReLU activation function, yet there are other prominent activation functions present. Through these conclusions, hyperparameters were adjusted for a two hidden layer neural network to account for differences in the activation functions.

The computational model of ReLU as the first layer and tanh as the second layer had a low accuracy of 0.82 for each set, in comparison to the accuracies of the other two hidden layer neural networks and had definitive overfitting with the BCE of the validation set being ~0.04 higher than that of the training set. The model with tanh as the first layer and ReLU as the second layer experienced similar setbacks with a higher accuracy of 0.88 for both sets but that is coupled with overfitting as the BCE of the validation set was ~0.05 higher than that of the training set. The model with both hidden layers having the tanh function proved to have the highest accuracy among the four combinations (0.90 for both sets), but there was significant overfitting, as the difference in error was ~0.06. Out of all four neural networks with two hidden layers, the computational model of two ReLU functions proved to be the best at limiting overfitting with a difference of error of only ~0.004. The only setback was the accuracy as it was 0.86 for both sets, which is a smaller accuracy in comparison to the first layer tanh/second layer ReLU model and the two tanh model.

Through these conclusions, it is evident that the best technique was a neural network with two hidden layers, and from an applicatory standpoint, the model with two ReLU function hidden layers is the best overarching computational model as it has virtually no overfitting with a relatively high accuracy, for a machine learning model, of 0.86.

Despite the successful results in developing accurate computational models, there is a main source of error that can be improved upon in future investigations. This is found in the copying of examples in the dataset. The dataset chosen was inherently unbalanced in terms of the number of banking crises examples to the number of non-banking crises examples. Because of this, banking crises values had to be copied before the model was implemented, possibly creating an inaccurate portrayal of the data. In further research, gathering a larger and more balanced data set is key for an accurate demonstration of banking crisis prediction.

By applying this machine learning model, great humanitarian benefits can be administered. Financial specialists can analyze the different features in the dataset, such as GDP, inflation CPI, independence, etc., of all areas around the world and utilize machine learning to predict if a banking crisis will occur in real-time, allowing governments and humanitarian services to administer economic relief in a time-effective manner. By utilizing this data-based approach, sustainability and positive financial permanence can be administered globally.

References

Alessi, L., & Savona, R. (n.d.). *Machine Learning for Financial Stability*. Retrieved from https://www.researchgate.net/publication/352278218_Machine_Learning_for_Financial_Stability

Brownlee, J. (2020, August 20). *A gentle introduction to the rectified linear unit (ReLU)*. Machine Learning Mastery. Retrieved from https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

Chiri. (2019, July 21). *Africa Economic, banking and Systemic Crisis Data*. Kaggle. Retrieved from https://www.kaggle.com/datasets/chirin/africa-economic-banking-and-systemic-crisis-data

Coletto, M., Lucchese, C., Orlando, S., & Perego, R. (n.d.). *Electoral Predictions with Twitter: a Machine-Learning approach*. Retrieved from https://core.ac.uk/download/pdf/53182811.pdf

Fouliard, J., Howell, M., & Rey, H. (2020, December). *Answering the Queen: Machine Learning and Financial Crises*.

Gensler, G., & Bailey, L. (2020). *Deep Learning and Financial Stability*. Retrieved from https://mitsloan.mit.edu/shared/ods/documents?PublicationDocumentID=7644

Gupta, V. (2021, April 20). *Understanding feedforward neural networks*. LearnOpenCV. Retrieved from https://learnopencv.com/understanding-feedforward-neural-networks/

Hvidberrrg@GitHub. Is That All There Is? (n.d.). Retrieved from https://hvidberrrg.github.io/deep_learning/activation_functions/sigmoid_function_and_derivative.html

IBM Cloud Education. (n.d.). *What is supervised learning?* IBM. Retrieved from https://www.ibm.com/cloud/learn/supervised-learning#:~:text=Supervised%20learning%2C%20also%20known%20as,data%20or%20predict%20outcomes%20accurately

Kanstrén, T. (2021, May 19). *A look at precision, recall, and F1-score*. Medium. Retrieved from https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec

Lokanan, M. E., & Sharma, K. (2022, January 31). *Fraud prediction using machine learning: The case of investment advisors in Canada*. Machine Learning with Applications. Retrieved from https://www.sciencedirect.com/science/article/pii/S2666827022000111

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022, February 27). *Machine learning technology-based heart disease detection models*. Journal of healthcare engineering. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8898839/#:~:text=Machine%20learning%20may%20be%20used,patients%20while%20avoiding%20serious%20effects

Patwari, D. K. (2021, July 15). *Activation function: Choose the best function for your network*. Medium. Retrieved from https://medium.com/nerd-for-tech/choose-the-best-activation-function-for-your-network-f46154bd9541

World Bank Group. (2017, November 7). *Banking crisis*. World Bank. Retrieved from https://www.worldbank.org/en/publication/gfdr/gfdr-2016/background/banking-crisis#:~:text=A%20(systemic)%20banking%20crisis%20occurs,other%20banks%20in%20the%20system