

The Integration of Big Data and Artificial Neural Networks for Enhancing Credit Risk Scoring in Emerging Markets: Evidence from Egypt

Osama Wagdi¹ & Yasmien Tarek²

¹ Finance and Accounting Department; Division of Business Administration; International Academy for Engineering & Media Science - IAEMS, Egypt

² Financial Institutions Department, Faculty of Management, Modern University for Technology & Information - MTI, Egypt

Corresponding author: Osama Wagdi, E-mail: osamawagdi_ta@yahoo.com; ORCID 0000-0003-0451-9726; ResearcherID: D-4898-2019; Scopus Author ID: 57212470180.

Received: December 8, 2021

Accepted: December 31, 2021

Online Published: January 7, 2022

doi:10.5539/ijef.v14n2p32

URL: <https://doi.org/10.5539/ijef.v14n2p32>

Abstract

This study investigates the effectiveness of technology models in credit risk scoring modeling in emerging markets. The study proposes evaluation methods for credit risk scoring modeling for current and potential borrowers through an investigation into the Egyptian banking industry by offering and examining a framework for the integration of big data and artificial neural networks based on systematic and unsystematic risk for both the macroeconomic environment and characteristics of current and potential borrowers. The data for the borrowers under examination covers the period from 2015 to 2019 for 75 firms, excluding 2020 and 2021 data to isolate the impact of COVID-19 on the results of the inferred statistics. Artificial Neural Networks was training within 25 firms under NeuroXL program but examination for 50 firms. The study found the ability of artificial neural networks to rank the commitment of borrowers in Egyptian banks under big data about the firm and Egyptian economy. Additions to discrepancy between the proposed model against some traditional models. Finally; The Integration of Big Data and ANN can help banks to bring out the value of data within create a level of financial stability for banks. Especially in emerging markets characterized by information inefficiency

Keywords: Big Data, Artificial Neural Network, Credit Risk Scoring, banking industry; emerging markets

1. Introduction

In the banking industry, Default risk is a significant issue that may have a significant impact on the decisions of banks. It may utilise accurate forecasts to identify hazardous borrowers and adjust their lending practises accordingly. (Moscatelli et al., 2020).

Lenders use a credit scoring technique to gather information about a borrower's payment habits in order to assess the level of risk involved (Trivedi, 2020). Some banks' biggest challenges, however, are deciding which borrowers are eligible for loans and which ones aren't. Building an automated artificial intelligence credit score prediction model based on artificial intelligence classifiers like machine learning and neural networks is critical for making such an important judgement (Trivedi, 2020). The financial industry, which includes banks, is attempting to stay competitive in the face of an enormous rise of data around the globe (Yldrm et al., 2021). Banks and FinTech companies are increasingly relying on artificial intelligence tools to anticipate and classify customers (Hurlin & P éignon, 2019; ACPR, 2020).

Big data analytics is becoming an integral aspect of data science because of the rapid expansion of financial data (Lee, 2017). Scalability and storage limitations have prevented traditional data approaches from extracting the desired information from a large volume and variety of financial data. Consequently, big data analytics provides a strong platform for doing deep analyses and delivering more accurate and dependable results for extracting information on economic activities (Narayanan, 2014; Oussous et al., 2018).

Banks have spent a lot of time and money over the last four decades establishing internal models for credit risk scoring. Bank authorities have applauded and encouraged these efforts. Banks have recently expanded their

efforts to include credit risk scoring modeling; however, an important question for both banks and their regulators is evaluating the accuracy of credit risk scoring modeling. Using a panel data approach, the study proposes evaluation methods for credit risk scoring modeling based on cross-sectional simulation for borrowers through an investigation into the Egyptian banking industry by offering and examining a framework for the integration of big data and artificial neural networks based on systematic and unsystematic risk for both the macroeconomic environment and borrowers.

2. Literature Review and Theoretical Framework

During the last quarter of the twentieth century, the pace of changes in the business environment accelerated based on the developments that occurred in many fields, especially the technological aspects, in addition to the liberalisation of international trade and capital movements, the world have experienced.

Egypt responded by The policies of economic openness have been underway since the seventies, as the levels of economic activities for the private sector, which was reflected in thire contribution of the private sector to the GDP, as this contribution by the private sector to various economic activities resulted in more transactions between the private sector and banking system on the other hand, which took many forms such as facilities beside bank loans, and in light of the many shocks that the Egyptian economy has been subjected to, both internal (due to terrorism, interest rate fluctuations, and exchange rate fluctuations), or external as southeast asian countries crises 1998, Mortgage Crisis 2008 and Pandemic Crisis (Covid-19) led to the exposure of many individuals and businesses to The risks of stopping its operating activities, which usually requires intervention from external parties, in order to ensure the continuation of work in that facility without losing jobs and cutting the supply chain, to provide goods and services to the Egyptian economy, usually through banking flotation.

Traditional financial methods such as financial ratios have been used for this purpose, in addition to a number of other methods that mix financial and non-financial methods at the level of credit to the business establishment (loans and banking facilities for companies) in exchange for proof of income and verifying the existence of a surplus to repay loans to individuals (Loans and personal facilities for individuals) in light of the controls of regulatory and supervisory authorities such as the Central Bank of Egypt.

2.1 Concept of Credit Scoring

Banks formulate their credit policies in a way that achieves good and safe use of the available funds, whether from depositors or shareholders of that unit, while achieving an appropriate return. Credit is defined as the ability to grant loans in light of the confidence that the banking unit places on its customers in making a certain amount of money available for its use. For a specific purpose during a certain period, and it is paid on certain terms in return for an agreed upon material return.

Bank credit patterns vary, as bank credit patterns can be determined according to several factors, including the maturity period (short/medium/long-term), according to the type of collateral (credit with in-kind guarantees/without in-kind guarantees), according to the payment method (equal payments/unequal payments/one payment), according to the legal form of the credit recipient (natural person/legal person), according to the type of credit currency (local currency/foreign currency), according to the number of parties providing credit (single credit/joint credit), and according to the economic sectors of the credit recipient (production activity: industrial/agricultural production/service/commercial).

Credit risk scoring is seen as one of the quantitative tools that can be used by banks to reduce their risks. Credit risk scoring is an extension of financing theories in general, and credit management in particular. Through these methods, the customer is viewed with a critical view, as the ability to pay is the basis for this assessment.

A debtor's financial flexibility and creditworthiness may be determined objectively using credit scoring. Debtors' exposure to credit, market, liquidity and interest rate risk is one of several aspects taken into account in this judgement. (Pertaia et al., 2021). The purpose of credit scorings is to provide a uniform and transparent system for the market candidates in order to understand the repayment capacity and to facilitate an efficient debt space (Szetela et al., 2019).

Credit scoring plays a key role in identifying default prediction of companies (Nehrebecka, 2018). In the retail credit sector, credit scoring is used for developing empirical models to support decisions (Crook et al., 2007). The huge numbers of individual and mortgage loans indicate that financial institutions require quantitative tools to process such big data to take credit decisions (Calabrese, 2014).

Traditional methods used to credit scoring include linear discriminant analysis and logit analysis. Emerging artificial intelligence techniques such as artificial neural networks, genetic algorithm and support vector machine were applied with positive results for both consumer and company credit ratings (Lai et al., 2006a).

Credit scoring models capture probability of the default behavior of borrowers in the future. Therefore, banks use predictive models called scorecards in application scoring, to evaluate the probability of non-payment. (Hand & Henley, 1997) it also have benefits for stakeholders and entities, as they help in optimizing the financing cost, and provide better access to raise capital (Alanis, 2020). In credit risk modeling, many prediction tasks occurs. According to Basel II Capital accord is necessary for banks to estimate like: Probability of default (PD); Exposure at default (EAD) and Loss given default (LGD). The recent trending models for research topic are EAD and LGD models (Yao et al., 2015).

In addition to the above, some studies have presented some variables affecting credit risk. These variables can be classified into two groups. The first is general variables called sources of systematic risk, versus private variables called sources of unsystematic risk.

The sources of systematic risk relate to credit risk such as Macroeconomic environment (Jakubik, 2007), interest rate (Chen et al., 2013), and the business cycle (Alessandrini, 1999); but the sources of unsystematic risk relate to credit risk such as Capital structure (Hackbarth et al., 2006); Liquidity (Ericsson & Renault, 2006). According to (Ham & Koharki, 2016) that a positive correlation between “the corporate general counsel” promotions to “senior management” and increases in business unit credit risk. Finally; Credit default swaps leads to looser loan terms (less collateral and looser covenants), consistent with a reduction in monitoring incentives (Shan et al., 2019)

2.2 Artificial Neural Network and Credit Risk Scoring

The field of artificial intelligence is based on research in biology, neuroscience, psychology, mathematics, and other related sciences, as it focuses on how the human mind performs its functions and how humans themselves learn from experience, as well as the way they think in different situations and circumstances, and the last quarter of the twentieth century witnessed the emergence of multiple decision-making methods based on artificial intelligence in the fields of economics, finance, and business, and these emerging sciences use computer systems designed on the basis of human thinking and intelligence. These methods include: Neural Networks, Genetic Algorithms, Genetic Programming, Evolutionary Programming, Classifier Systems, Agent-based Modeling, Fuzzy Logic, Wavelets, Molecular Computing.

Artificial neural networks are a method by which data is analyzed by tracking data trends, studying the complex and interrelated relationships between them, while learning to adapt to environmental changes. With it, and these artificial neural networks are modeled to resemble the human brain, and with the way it works.

When artificial neural networks emerged, they were limited to engineers only, and with the beginning of the use of artificial neural networks as software, they began to invade the financial field, and due to the fact that many banks and pension funds as well as insurance companies and international investment funds managed billions of dollars in the financial markets, in the business of buying, selling and speculating on stock prices. And currencies, and the managers of those facilities and institutions need for modern scientific methods to help predict what will happen in the future, in order to direct their investments towards the right direction of investment, hence the use of artificial neural networks in the field of accounting to predict stock prices, or changes in them.

At the end of the nineties of the twentieth century, the applications of artificial neural networks varied in accounting and finance, which required dividing them into multiple types: in financial accounting, auditing, cost accounting, management accounting, as well as financial investments.

With the high rates of interest in artificial neural networks, their industrial applications increased, and with the multiplicity of software ready for artificial neural networks, the fields applying artificial neural networks in the field of applied and social sciences also increased.

The vast majority of Artificial Neural Network applications can be placed in prediction application classifications, data filtering, classification, and data correlation, as well as sample and pattern recognition, and data modeling.

A neural network is a massively parallel-distributed processor that has a natural propensity for storing experiential knowledge and making it available for use, Credit scoring was one of the first areas of economics and finance to use machine learning and artificial intelligence techniques. is a related field of Business Intelligence which refers to as techniques, technologies, systems, and tools that enable accessing to diverse data, manipulating and transforming such data to provide managers of productivity and competitiveness. It also helps improving decision making progress to perform accurate actions in the future (Chen et al., 2012, 2013).

Some early studies were on using artificial neural networks in the financial arena like (Tam & Kiang, 1992) used managerial applications of neural networks for the case of bank failure predictions, (Desai et al., 1996)

conducted a comparison of neural networks and linear scoring models in the credit union environment; (West, 2000) investigated the credit scoring accuracy of five neural network models, (Yobas et al., 2000) compared the predictive power of linear discriminant analysis, neural networks, genetic algorithms and decision trees in distinguishing between good and slow payers of bank credit card accounts. (Bahrammirzaee, 2010), confirmed that neural networks have proven the ability to predict credit worthiness rather than its counterparties.

According to Hosaka (2019), banks can use Convolutional Neural Network (CNN) through imaged financial ratios to predict bankruptcy, and compare findings with other techniques such as DT, linear discriminant analysis, SVM, MLP, and Altman's Z-score. The results showed higher performance results with CNN in comparison with other machine learning methods.

But Khandani et al. (2010) construct a credit score model via machine learning, but their model only incorporates bank account transactions and credit bureau information. on other side; Nwulu et al. (2012) used support vector machine (SVM) and Artificial Neural Network (ANN) for building credit-scoring models. Results showed that ANN outperforms SVM with significant high accuracy.

In addition to Bequ   and Lessmann (2017) used artificial neural network to build extreme learning machine (ELM) credit score predication model. but Trivedi (2020) conducted a study on credit scoring modeling with different feature selection and machine learning approaches.

One of the main shortages of Neural Networks techniques in the credit scoring industry comes from their lack of explain ability and interpretability. Because most of such Machine learnings tools, are considered as "black boxes "in the sense that the corresponding scorecards and credit approval process cannot be easily explained to customers and regulators. (Dupont et al., 2020), this explains why the logistic regression remains the standard approach in credit industry, because of its simplicity and intrinsic interpretability (EBA, 2020; Bracke et al., 2020).

Criticisms of Artificial Neural Networks; There are many criticisms of artificial neural networks, which the researcher can summarize in the following points:

- a) The most common criticism of artificial neural networks is that they work like black boxes, they work in an unclear way, and they do the main calculations in the hidden layer, and there is no apparent way to process that information in the hidden layer, so the rules of operation of this network cannot be fully understood.
- b) The process of training an artificial neural network requires a lot of effort and time, compared to the results of an artificial neural network or the cost of preparing a knowledge base by experts, this is not the problem of the system, the more experiments, the greater the confidence in the results produced by the network We must point out that despite the limited training time required for artificial neural networks especially when using multiple layers, the speed involved in software development overcomes most of the shortcomings.
- c) There are no mathematical theories that can be applied to ensure the successful performance of an applied artificial neural network, and those theories are still under development or in the process of being created.
- d) Sometimes artificial neural networks are biased, even if there is no relationship between these data, it will try to fit the data curve and visualize the relationship between certain data, that is, exaggerating some data even in the absence of a relationship between these data, and this is called "Over Fitting", This occurs when the network is overtrained, which results in poor network comprehensiveness.

Despite the previous weaknesses, they do not reduce the success of artificial neural networks in any accounting and financial fields, as they have proven their effectiveness and efficiency in the work of estimation and forecasting.

2.3 Big Data analytics and Credit Risk Scoring

Data analytics defined as technologies and algorithms that facilitate the analysis of data; the big data technology has been successfully used in the Internet of Things, and the financial industry hopes to use such advanced technology to integrate and improve internal and external data related to credit risks (Chunhui et al., 2021).

With shedding light on consumer lending process, financial institutions have begun using large-scale external data to evaluate the creditworthiness of potential borrowers. (Jiang et al., 2021). The big data credit score system has greater coverage and uses thousands of variables that cannot be easily manipulated by traditional rating techniques, which relies heavily on official credit records that are not available for most individuals (Jiang et al., 2021). The big data credit score include an individual's very frequent and realtime behavioral information, such as online cash loan applications, online shopping, and internet browsing, which could reflect an individual's general profile accurately (Einav & Levin, 2014, Garmaise, 2015).

Traditional data analytic platforms are challenged by storage, management, and analysis challenges as the volume of data grows exponentially. Decentralized and distributed processing is provided by BDA; it call as emerging Big Data Analytics (Yildirim et al., 2021).

According to Gandomi and Haider (2015); Big data analytics consists of five sub-stages: “Acquisition and Recording”, “Extraction, Cleanup and Commentary”, “Integration, Clustering, Representation”, “Modeling and Analysis”, and “Electronic Interpretation”. By analyzing a huge amount of financial data, banks can obtain valuable information to determine their strategic plans such as risk control, crisis management or growth management (Yildirim et al., 2021).

However, there are certain drawbacks to using big data in credit evaluation. The first difficulty is how to collect, store, and manage massive amounts of data in a cost-effective manner. Using typical econometric methods to calculate credit scores from large datasets might also be difficult. high-dimensional data analysis, which is computationally expensive and susceptible to overprocessing, is another challenge (Stanmirova et al., 2007; Yu et al., 2009).

3. Examine and Results

The study seeks to demonstrate the effectiveness of the integration between big data and artificial neural networks to support the decisions of banking units towards credit success in the Egyptian business environment, as it is one of the emerging markets, unlike the majority of previous studies that dealt with developed markets.

3.1 Study Design

Big data in credit evaluation usually refers to applying machine learning algorithms that have high predictive power. The study built a credit-scoring model based on integration between the FinTech field and the artificial intelligence field. Regarding FinTech, we have focused on big data analytics as a branch of FinTech activities due to its advantages that enable accessing diverse data, manipulating and transforming such data for rational decisions. Regarding AI techniques, we incorporated artificial neural networks (ANNs) rather than other AI techniques due to their ability to better classify and predict huge amounts of data, in addition to their ability to capture nonlinear relationships in incomplete, random, and hugely diverse data sets of loan candidates. Therefore, the study create an integrated system for loan approval in rational and secured practises for entities, which ends up mitigating default risks largely.

This study investigates the improvements in credit risk scoring modeling accuracy by under Integration of Big Data and Artificial Neural Networks, with two real world data sets partitioned into training and independent test sets using panel data approach. The study proposes evaluation methods for credit risk scoring modeling based on cross-sectional simulation for borrowers through an investigation into the Egyptian banking industry by offering and examining a framework for the integration of big data and artificial neural networks based on systematic and unsystematic risk for both the macroeconomic environment and borrowers.

The study used a training set establish the credit scoring model’s parameters, while the independent holdout sample is used to exam the generalisation capability of the model. The examination covers the database of the Egyptian Credit Bureau “I-Score” In addition to the advice data from the “Central Bank of Egypt” and “Central Agency for Public Mobilization and Statistics-Egypt” and “IHS market” for period from 2015 to 2019, excluding 2020 and 2021 data to isolate the impact of COVID-19 on the results of the inferred statistics; Figure 1 illustrates proposes credit risk scoring modeling based on the integration of big data and artificial neural networks in the Egyptian business environment.

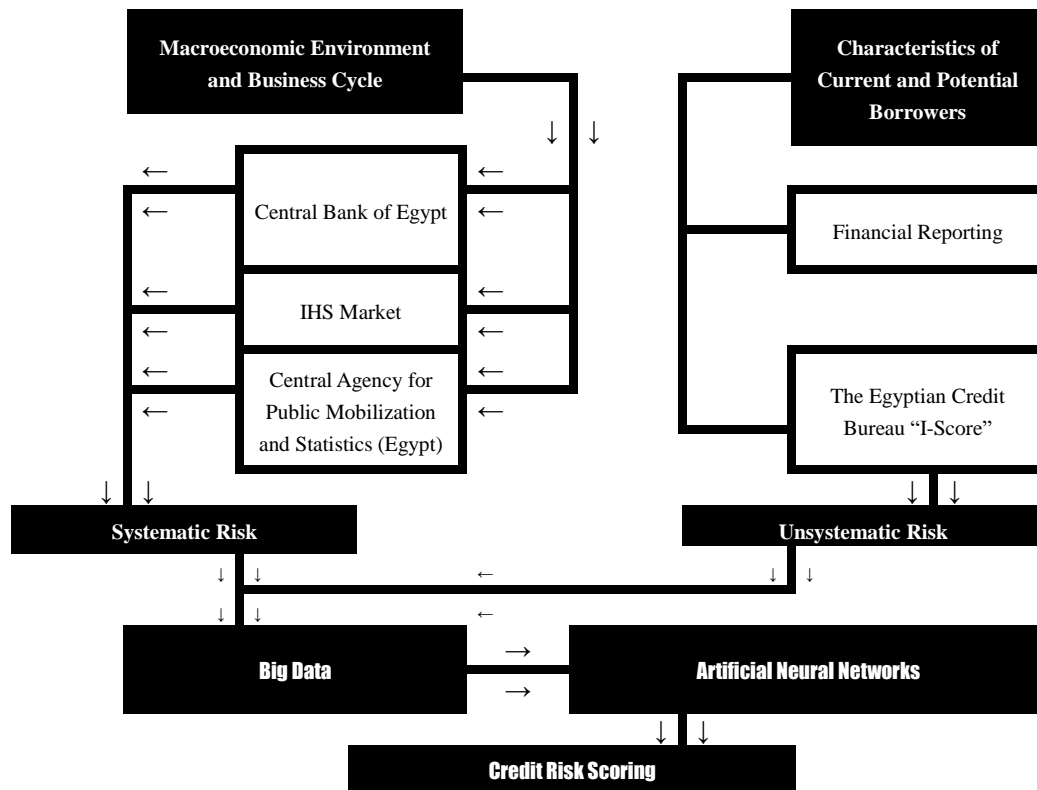


Figure 1. Proposes credit risk scoring modeling based on the Integration of Big Data and Artificial Neural Networks

3.2 Study Variables

The proposed model for credit evaluation includes two main groups of variables, including systematic and unsystematic risks; the first group includes economic variables and business cycle (Economic Growth Rate, Interest Rate, Exchange Rate, Inflation Rate, and Purchasing Managers' Index). Second group includes two types of variables: financial indicators and control variables. The credit risk scoring modeling indicator system established of four sources of data, covering the status of credit. Table 1 illustrates study variables.

Table 1. Study variables

Sources	Symbol	indicator	Variable
Financial Reporting	F1	Gross profit margin	Independent
	F2	Operating profit margin	
	F3	Accounts receivable turnover	
	F4	Inventory turnover	
	F5	Accounts payable turnover	
	F6	Operating cash flow for fixed expenses	
	F7	Operating cash flow for free cash flow	
	F8	Debt to equity ratio	
	F9	Financial leverage	
	F10	Operating leverage	
	F11	Current ratio	
	F12	Quick ratio	
	F13	Cash ratio	
	F14	Revenue Growth Rate	
	F15	Net profit growth rate	
I-Score	ORR	Obligor Risk Rating	
CBE, CAPMS & IHS market	E1	Economic Growth Rate	
	E2	Interest Rate	
	E3	Exchange Rate	
	E4	Inflation Rate	
	E5	PMI, Purchasing Managers' Index	

Table 1. Study variables (continue)

Sources	Symbol	indicator	Variable
Financial Reporting	C1	logarithm of Sales	Control
	C2	logarithm of total assets	
	C3	Logarithm of total owner's equity	
	C4i	Sector i	
	C5	weight of owner's equity for Institutional	
ANN	CRS	Credit Risk Scoring	Dependent

In addition to the previous variables, and some verification of the efficiency of the model, the firms were estimated according to some traditional models (Beaver, 1966; Edmister, 1972).

3.3 Training Artificial Neural Networks

Artificial Neural Network was training within 119 firms under NeuroXL program (see Figure 2).

Author: Yaskov Anatoliy
 Company: OLSOFT LLC.
 Title: NeuroXL Predictor
 Comments: Neural network-based forecasting and estimating tool.
 Version 3.1.2

Figure 2. NeuroXL program version

It is built by entering the data that makes up the artificial neural network; the parameters are shown Figure 3; But Figure 4 illustrates Artificial Neural Networks under training within 119 firms.

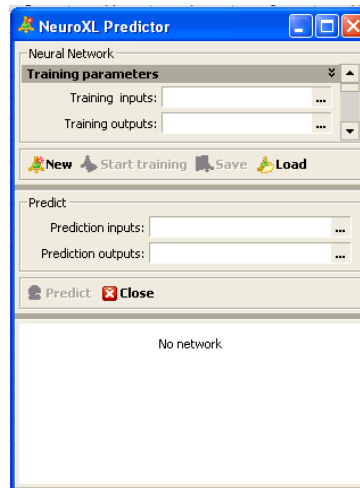


Figure 3. The parameters of artificial neural network

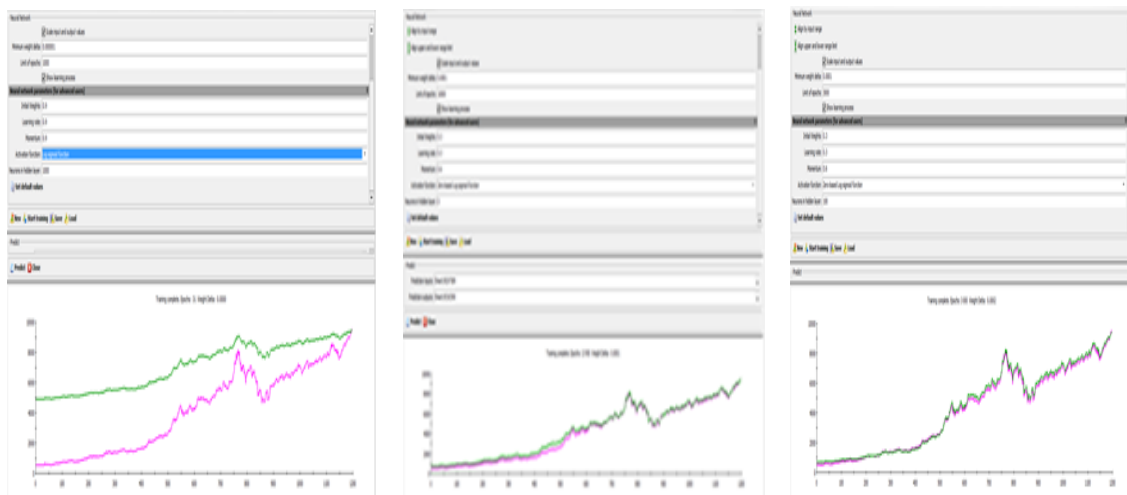


Figure 4. Artificial Neural Networks under training

3.4 Examining the Integration of Big Data and Artificial Neural Networks to Rank Credit Commitment of Borrowers

This hypothesis examines the ability of artificial neural networks to rank credit commitment of borrowers in an Egyptian bank under big data about the firm and Egyptian economy. The study used a cross-sectional units test to examine this hypothesis based on fixed-effects. This examines includes credit risk scoring –the rank commitment of borrowers based on artificial neural networks – as independent variable (Prediction value) against obligor credit – based on credit report from bank- as dependent variable (Actual value).

Table 2. Credit risk scoring (Prediction value) against obligor credit (Actual value)

Model 1: Fixed-effects, using 250 observations

Included 50 cross-sectional units

Time-series length = 5

Dependent variable: OBLIGOR

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
Const.	66.7340	3.55130	18.79	<0.0001	***
CRS	29.5000	3.79298	7.778	<0.0001	***
Mean dependent var	94.22800	S.D. dependent var		19.17621	
Sum squared resid	5725.900	S.E. of regression		5.364081	
LSDV R-squared	0.937466	Within R-squared		0.233111	
LSDV F(50, 199)	59.66497	P-value(F)		8.25e-96	
Log-likelihood	-746.1464	Akaike criterion		1594.293	
Schwarz criterion	1773.887	Hannan-Quinn		1666.574	
rho	-0.042992	Durbin-Watson		1.478981	

Joint test on named regressors -

Test statistic: $F(1, 199) = 60.49$

with p-value = $P(F(1, 199) > 60.49) = 3.91048e-013$

Test for differing group intercepts -

Null hypothesis: The groups have a common intercept

Test statistic: $F(49, 199) = 6.03645$

with p-value = $P(F(49, 199) > 6.03645) = 3.01116e-020$.

According to the statistical results from table 2, the study rejects the null hypothesis, as the study found the ability of artificial neural networks to rank the commitment of borrowers in an Egyptian bank under big data about the firm and Egyptian economy. This was not significant at the 0.01 level. According “within R-squared”; the credit risk scoring –the rank of the commitment of borrowers based on artificial neural networks – explain of 23.31% of change in obligor credit of borrowers – based on credit report from bank-.

3.5 Examining the Rank Credit Commitment of Borrowers Based on “Artificial Neural Networks And Big Data” Against Some Traditional Models

This hypothesis compares the credit rating according to the proposed model based on integration of big data and artificial neural networks against some traditional models. The study used a Friedman Test for that within Beaver (1966) and Edmister (1972).

Table 3. Credit risk scoring against Beaver and Edmister models

Ranks	
Mean Rank	
BM	2.03
EM	1.88
CRS	2.10
Test Statistics ^a	
N	250
Chi-Square	45.489
df	2
Asymp. Sig.	0.00
a. Friedman Test	

According to the statistical results from table 3, the study rejects the null hypothesis, as the study found a discrepancy between the proposed model - based on integration of big data and artificial neural networks- against some traditional models for both Beaver (1966) and Edmister (1972).

4. Conclusion and Recommendations

4.1 Conclusion

Machine learning and other expert systems in the banking industry are being used to speed up loan decisions while also controlling increasing risk. Lenders have relied on credit ratings for a long time to make loan choices for individuals as well as businesses. In essence, most credit rating systems are based on information provided by financial institutions, such as payment history and transaction data. Banking and other lenders are now relying on a variety of non-structured data, including text message and mobile phone use activity data as well as social media usage, to better assess for persons as retail clients. In order to improve credit risk scoring in emerging markets, the researchers advocated integrating big data and artificial neural networks for enhancing credit risk Scoring of companies as institution clients. These issues are intended for the current study.

The study proposed a dynamic model for credit risk scoring in emerging markets; this model has Integration of Big Data and Artificial Neural Networks that includes the factors of credit risk based on financial, non-financial and economic variables. Our model can accommodate ever-changing uncertain financial, non-financial and economic factors. Big Data can help banks to bring out the value of data so better decisions can be taken related to credit. Through this approach, banks would have less credit risk when forecasting which borrowers will be successful in their payments.

The data for the borrowers under examination covers the period from 2015 to 2019 for 75 firms, excluding 2020 and 2021 data to isolate the impact of COVID-19 on the results of the inferred statistics. Artificial Neural Networks was training within 25 firms under NeuroXL program but examination for 50 firms. The study found the ability of artificial neural networks to rank the commitment of borrowers in Egyptian banks under big data about the firm and Egyptian economy. Additions to discrepancy between the proposed model against some traditional models – Beaver, 1966 and Edmister, 1972-. So; The Integration of Big Data and ANN can help banks to bring out the value of data within create a level of financial stability for banks. Especially in emerging markets characterized by information inefficiency.

4.2 Recommendations for Egyptian Banking Industry

Emerging markets and Egypt too are highly ready to accept plans of domestic and multinational firms for more projects. so it is a positive factor to bank's need for Enhancing Credit Risk Scoring models, Therefore, the study suggests the following recommendations:

- a. Egypt needs restructuring to information environment. The development of regulatory restrictions play a critical role to make the information more suitable for analysis, the study finds that this mission is assigned to the Central Bank of Egypt and the Financial Supervisory Authority.
- b. Banks have to close the gap between current technological infrastructures and the required technological credit risk scoring systems. This requires more investment in financial technology.
- c. In order to effectively integrate current cases and the needed abilities, banks must allocate money for seminars and session centres to train their staff on credit risk scoring systems.

4.2 Recommendations for Future Research

Researchers could pay attention to the Fintech category of Robo-advising, which includes systems that provide automated advice to many parties in the banking system and firms. Using technologies like artificial intelligence, big data, and machine learning all together for building data mining techniques advisor system. In addition, test fuzzy logic for credit scoring because it is accurate in uncertainty modeling with machine learning and big data analytics. Therefore, the study suggests the following recommendations as a title for future studies:

- a. Data Mining Techniques from Social Media and Credit Card Risk Scoring model.
- b. Effective Comparative Study of Credit Risk Scoring between international and Emerging Markets.
- c. Emerging market requirements for the application of machine learning systems in banks

References

- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182, 113-131. <https://doi.org/10.1016/j.ijpe.2016.08.018>

- Alanis, E. (2020). Is there valuable private information in credit ratings? *The North American Journal of Economics and Finance*, 54, 101293. <https://doi.org/10.1016/j.ijpe.2016.08.018>
- Alessandrini, F. (1999). Credit risk, interest rate risk, and the business cycle. *The Journal of Fixed Income*, 9(2), 42-53. <https://doi.org/10.3905/jfi.1999.319259>
- Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19(8), 1165-1195. <https://doi.org/10.1007/s00521-010-0362-z>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 71-111. <https://doi.org/10.2307/2490171>
- Bequ   A., & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications*, 86, 42-53. <https://doi.org/10.1016/j.eswa.2017.05.050>
- Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). *Machine learning explain ability in finance: An application to default risk analysis*. Working Paper No. 816, Bank of England. <https://doi.org/10.2139/ssrn.3435104>
- Calabrese, R. (2014). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry*, 30(2), 99-114. <https://doi.org/10.1002/asmb.1932>
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188. <https://doi.org/10.2307/41703503>
- Chen, R. R., Cheng, X., & Wu, L. (2013). Dynamic interactions between interest-rate and credit risk: Theory and evidence on the credit default swap term structure. *Review of Finance*, 17(1), 403-441. <https://doi.org/10.2139/ssrn.779445>
- Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447-1465. <https://doi.org/10.1016/j.ejor.2006.09.100>
- Desai, V. S., Crook, J. N., & Overstreet Jr, G. A. (1996). A comparison of neural net- works and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95(1), 24-37. [https://doi.org/10.1016/0377-2217\(95\)00246-4](https://doi.org/10.1016/0377-2217(95)00246-4)
- Dumitrescu, E., Hue, S., Hurlin, C., & Tokpavi, S. (2021). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178-1192. <https://doi.org/10.1016/j.ejor.2021.06.053>
- Dupont, L., Fliche, O., & Yang, S. (2020). *Governance of Artificial Intelligence in Finance*. Discussion document; Banque De France. https://acpr.banque-france.fr/sites/default/files/medias/documents/20200612_ai_governance_finance.pdf
- EBA. (2020). *Report on big data and advanced analytics*. European Banking Authority, January. <https://www.cambridge.org/core/product/identifier/>
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, 7(2), 1477-1493. <https://doi.org/10.2307/2329929>
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210). <https://doi.org/10.1126/science.1243089>
- Ericsson, J., & Renault, O. (2006). Liquidity and credit risk. *The Journal of Finance*, 61(5), 2219-2250. <https://doi.org/10.1126/science.1243089>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Garmaise, M. J. (2015). Borrower misreporting and loan performance. *The Journal of Finance*, 70(1), 449-484. <https://doi.org/10.1111/jofi.12156>
- Hackbarth, D., Miao, J., & Morellec, E. (2006). Capital structure, credit risk, and macroeconomic conditions. *Journal of Financial Economics*, 82(3), 519-550. <https://doi.org/10.1016/j.jfineco.2005.10.003>
- Ham, C., & Koharki, K. (2016). The association between corporate general counsel and firm credit risk. *Journal of Accounting and Economics*, 61(2-3), 274-293. <https://doi.org/10.1016/j.jacceco.2016.01.001>
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.

- <https://doi.org/10.1111/j.1467-985X.1997.00078.x>
- Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, 287-299. <https://doi.org/10.1016/j.eswa.2018.09.039>
- Hurlin, C. , & P é rignon, C. (2019). Machine learning et nouvelles sources de données pour le scoring de crédit. *Revue d'économie financière*, (3), 21-50. <https://doi.org/10.3917/ecofi.135.0021>
- Jakubik, P. (2007). Macroeconomic environment and credit risk. *Czech Journal of Economics and Finance*, 57(1-2), 60-78. https://doi.org/10.1111/j.1460-2466.2006.00329_5.x
- Jiang, J., Liao, L., Lu, X., Wang, Z., & Xiang, H. (2021). Deciphering big data in consumer credit evaluation. *Journal of Empirical Finance*, 62, 28-45. <https://doi.org/10.1016/j.jempfin.2021.01.009>
- Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787. <https://doi.org/10.1016/j.jbankfin.2010.06.001>
- Lai, K. K., Yu, L., Wang, S., & Zhou, L. (2006, September). Credit risk analysis using a reliability-based neural network ensemble model. *International Conference on Artificial Neural Networks*, 682-690. https://doi.org/10.1007/11840930_71
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303. <https://doi.org/10.1016/j.bushor.2017.01.004>
- Moscattelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications*, 161, 113567 <https://doi.org/10.1016/j.eswa.2020.113567>
- Narayanan, V. (2014). Using big-data analytics to manage data deluge and unlock real-time business insights. *The Journal of Equipment Lease Financing*, 32(2), 1-7. <https://doi.org/10.1016/j.procs.2016.05.152>
- Nehrebecka, N. (2018). Predicting the default risk of companies. comparison of credit scoring models: Logit vs support vector machines. *Econometrics*, 22(2), 54-73. <https://doi.org/10.15611/eada.2018.2.05>
- Nwulu, N. I., Oroja, S., & Ilkan, M. (2012). A comparative analysis of machine learning techniques for credit scoring. *Information, an International Interdisciplinary Journal*, 15(10), 4129-4146. <http://hdl.handle.net/11129/2207>
- Oskarsd ottir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, 74, 26-39. <https://doi.org/10.1016/j.asoc.2018.10.004>
- Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Pertaia, G., Prokhorov, A., & Uryasev, S. (2021). A new approach to credit ratings. *Journal of Banking & Finance*, 106097. <https://doi.org/10.1016/j.jbankfin.2021.106097>
- Shan, C., Tang, D. Y., & Winton, A. (2019). Do banks still monitor when there is a market for credit protection? *Journal of Accounting and Economics*, 68(2-3), 101241. <https://doi.org/10.1016/j.jacceco.2019.101241>
- Srivastava, U., & Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. *Procedia Computer Science*, 50, 643-652. <https://doi.org/10.1016/j.procs.2015.04.098>
- Stanimirova, I., Daszykowski, M., & Walczak, B. (2007). Dealing with missing values and outliers in principal component analysis. *Talanta*, 72(1), 172-178. <https://doi.org/10.1016/j.talanta.2006.10.011>
- Szetela, B., Mentel, G., & Brożyna, J. (2019). Modelling European sovereign default probabilities with copulas. *Economic research-Ekonomska istraživanja*, 32(1), 1716-1726. <https://doi.org/10.1080/1331677X.2019.1629325>
- Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, 38(7), 926-94. <https://doi.org/10.1287/mnsc.38.7.926>
- Trivedi, S. K. (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society*, 63, 101413. <https://doi.org/10.1016/j.techsoc.2020.101413>
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356-365. <https://doi.org/10.1016/j.jbusres.2016.08.009>

- West, D. (2000). Neural network credit scoring models. *Computers & operations research*, 27(11-12), 1131-1152. [https://doi.org/10.1016/S0305-0548\(99\)00149-5](https://doi.org/10.1016/S0305-0548(99)00149-5)
- Yao, X., Crook, J., & Andreeva, G. (2015). Support vector regression for loss given default modelling. *European Journal of Operational Research*, 240(2), 528-538. <https://doi.org/10.1016/j.ejor.2014.06.043>
- Yıldırım, M., Okay, F. Y., & Özdemir, S. (2021). Big data analytics for default prediction using graph theory. *Expert Systems with Applications*, 176, 114840. <https://doi.org/10.1016/j.eswa.2021.114840>
- Yobas, M. B., Crook, J. N., & Ross, P. (2000). Credit scoring using neural and evolutionary techniques. *IMA Journal of Mathematics Applied in Business and Industry*, 11, 111-125. <https://doi.org/10.1093/imaman/11.2.111>
- Yu, K., Zhang, T., & Gong, Y. (2009). Nonlinear learning using local coordinate coding. *NIPS*, 22, 2223-2231. Retrieved from <http://hdl.handle.net/1783.1/98329>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).