Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Analysis of Machine Learning in Classifying Bank Profitability with Corruption Factor

Suraya Masrom[1], Ratna Septiyanti[2], Airuddin Ahmad[3], Rahayu Abdul Rahman[4,*], Norhayati Sulaiman[4]

[1] Computing Science Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia
[2] Faculty of Economics and Business, University of Lampung, Lampung, Indonesia
[3] Faculty of Computing and Multimedia, Universiti Poly-Tech Malaysia, Kuala Lumpur, Malaysia
[4] Faculty of Accountancy, Universiti Teknologi MARA, Perak Branch, Tapah, Malaysia

ARTICLE INFO

Article history:
Received 22 June 2023
Received in revised form 17 October 2023
Accepted 21 October 2023
Available online 28 February 2024

Keywords:
Machine Learning; Bank Profitability; Corruption; Classification

ABSTRACT

Corruption, which is generally defined as the abuse of authority for personal benefit, is not a recent phenomenon and has become a major issue in almost all countries throughout the world. Indeed, a high level of corruption increases bank non-performing loans which in turn reduces profitability and intensifies the fragility of the banking industry. Given the adverse impacts, corruption has been used as one of the factors in bank performance evaluation. As research on corruption-bank performance with machine learning techniques is rarely reported in the literature, this paper presents the empirical comparison of different machine learning algorithms for classifying bank profitability. Besides machine learning performance comparisons, this paper presents the analysis of machine learning features importance to justify the effect of corruption factor in the different machine learning algorithms for classifying bank profitability. The results indicated that all the tested machine learning algorithms present a good ability of classifying bank profitability at accuracy percentages above 70% but corruption index has contributed very minimal effect to the machine learning performances. The framework of this research is highly reproducible to be extended with a more in-depth analysis, particularly on the bank profitability factors as well as on the machine learning algorithms.

1. Introduction

Corruption has grown to be a serious problem in almost all countries throughout the world which affect the economy, security, and social wellbeing of the population. According to International Transparency, no country is immune to corruption, although the extent of corruption varies across cultures and national contexts [1]. In line with it, the media frequently exposes various corruption scandals in many countries, including developed countries like Japan and the United States as well as developing countries like Malaysia, Indonesia, Mexico, and the Philippines. Prior studies highlight that corruption affects bank profitability [2-5]. Corruption enhances the possibility of diverting

* *Corresponding author.*
E-mail address: rahay916@uitm.edu.my

https://doi.org/10.37934/araset.40.2.1321

loanable funds to bad projects [6]. This is because bank managers may approve bad loans just for private benefits arising from bribery and corruption despite the knowledge of the risky nature of such loans. Moreover, loans involving corruption are usually characterized by high default and credit risk in bank portfolios [7]. As a result, non-performing loans might be increased which in turn reduces profitability of the bank [8].

Therefore, predicting the bank profitability with corruption is becoming important for the stakeholders as well as the regulators. Recently, machine learning has become a vital predictive approach in a variety of domains including in education [9,10], agriculture [11], medical [12-14], business [15], fraud detection [16], and energy management [17]. In addition, several studies have proven the capability of machine learning in generating higher accurate results in bank profitability prediction [18-20]. For example, [18] predict the determinants of 13 Development and Investment Banks' profitability in Turkey from 2002 to 2014 using three generally-preferred tree-based machine learning methods; namely DStump, RTree, and REPTree as the base learners and benchmark models. In addition, the study develops three ensemble learning models using bagging ensemble learning methods (i.e., Bag-DStump, Bag-RTree, Bag-MLP and Bag-REPTree). The findings reveal that bagging ensemble models are superior to their base learners and could improve the prediction accuracy of individual machine learning models (i.e., DStump, RTree, REPTree). Meanwhile, [19] develop prediction models on bank performance using boosting regression trees. The study employs data from rural banks in 30 provinces in China and finds that the gradient boosting regression tree algorithm obtains better classification results as it overcomes the shortcomings of low accuracy and poor generalization ability of the existing regression decision tree model. Further, study by [20] aims to predict bank performance of 50 Turkish banks, 30 American banks and 20 European banks. To construct the prediction model, the study employs machine learning classifiers; Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR) and are combined with Genetic Algorithm (GA). Overall, the results reveal that GA+SVM hybrid model with optimization but without feature selection provides best accuracy among all the models which is 100% test accuracy.

Despite the widespread use of machine learning in banking performance prediction, the inclusion of corruption factor is difficult to be found in the literature. Recent studies on corruption-bank performance mostly relied on conventional essential approaches such as multiple linear regression [21-25]. Therefore, the objective of this study is to extend the state-of-the-art of corruption by focusing on machine learning bank profitability prediction models. The existence of advanced techniques of machine learning can be deployed and to start with a few of them is highly crucial. Besides identifying the machine learning performances through prediction accuracy and time efficiency, another important question that needs to be answered is how the corruption aspect contributed to the performances of different machine learning models. This issue has not been broadly discussed in the existing research reports.

The following section provides a brief description on the data set of the concerned problem and machine learning implementation methodology. Section 3 describes and discusses the experimental results for the representative compared algorithms. Finally, section 4 presents the conclusions and future research directions.

2. Methodology

2.1 Data Collection and Dataset

This study uses 358 of bank year observations listed on the Malaysian stock exchange as a sample dataset. Similar to prior research [21-25], this study uses return on average asset (ROAA) to measure bank profitability. Based on the ROAA, two classes of bank profitability were created namely *High*

Bank Profit and *Low Bank Profit*, which were used as the target label for the machine learning. *High Bank Profit* denotes that the bank profitability or the ROAA is 1 and above. Otherwise, the ROAA class is 0 (*Low Bank Profit*). In other words, ROAA class is the dependent variable of the prediction models. Machine learning uses prediction probability to classify the class label in such a way that if the prediction probability is above 0.5 and above, the given case will be classified as 1. Table 1 shows the definition of the dependent variable.

Table 1
Auto model specifications

Variable	Meaning	Class label
ROAA class	ROAA is defined as net income over the average assets. If the ROAA is below 1, the ROAA is classified as 0, otherwise 1.	1- High Bank Profit 0- Low Bank Profit

Table 2 indicates the set of features (independent variables) used to construct the prediction models for classifying bank profitability. Six features have been included representing bank characteristics and macroeconomics factors of the particular banks in Malaysia. The correlation coefficient of each feature to the dependent variable based on Pearson Correlation test is also given in Table 2.

Table 2
List of features

Features/Variables	Meaning	Measurement method	Correlation
Corruption index (CI)	Corruption index (CI) is the perceived levels of public-sector corruption in a country	CI=100–CI index scores by Transparency International.	0.051
GDPG	Gross domestic product (GDP) growth rate	GDP growth rate	0.036
Size	Bank size	The bank size is defined as the logarithm of total assets.	0.526
Liquidity	Liquidity ratio	The liquidity ratio equals the loan divided by customer's deposits.	0.117
NPL	Non-performing loan	Non-performing loan (NPL) is defined as impaired loans to gross loans ratio	0.188
MEf	Management efficiency	Management efficiency (MEf) is measured by dividing the operating expenses by the operating income generated, signalling bank efficiency in managing their operation by reducing costs and increasing profit.	0.84

The features of the prediction models are grouped into two main categories. First category is specific characteristics of the bank that include (bank size (*size*), bank liquidity (*liquidity*), management efficiency (*MEf*) and bank non-performing loan (*NPL*)) while the second category is macroeconomics based on corruption (*CI*) and gross domestic product growth rate (*GDPG*)). Outside the machine learning algorithms, corruption and GDP growth rate present low correlation to the bank profitability. Is the corruption remaining as the least feature importance in the machine learning algorithm is the main question of this research. It is interesting to observe which features will be the most important in the different machine learning algorithms is another issue. The findings are given in Section 3. The samples of data are depicted in Figure 1.

	A	B	C	D	E	F	G	H
1	**CI**	**GDPG**	**Mef**	**Liquidity**	**NPL**	**Size**	**ROAA**	**ROAAIndex**
2	67	5.390988327	61.42	697342	37.19	15.51	0.35	0
3	63	5.788499279	57.19	1254500	35.22	15.53	0.38	0
4	50	6.783437734	60.2	1004158	26.40	15.37	0.85	0
5	49	5.332139144	48.25	1037275	17.17	15.70	0.76	0
6	50	5.584847069	49.33	1777460	16.92	15.82	0.62	0
7	49	6.298785927	52.77	1816057	15.21	15.83	0.69	0
8	49	4.831769895	48.26	1418158	5.84	15.90	1.11	1
9	55	-1.513685083	44.1	1445525	3.64	16.00	1.02	1
10	56	7.425970496	44.48	2209081	3.57	16.26	1.05	1
11	57	5.293784657	45.51	2132893	2.68	16.35	0.97	0
12	51	5.473454192	45.13	1590197	2.16	16.43	1.1	1
13	50	4.713453716	43.97	1902758	1.88	16.44	1.17	1
14	48	5.992609342	44.62	1398856	1.78	16.44	1.02	1
15	64	8.858868104	36.3	287447	16.32	15.12	0.74	0
16	64	0.517675303	44.04	486868	18.56	15.25	0.82	0
17	67	5.390988327	45.79	473447	15.74	15.30	1.01	1
18	63	5.788499279	39.55	850579	13.47	15.48	0.97	0
19	50	6.783437734	38.24	943974	10.72	15.55	0.89	0
20	49	5.332139144	52.65	1174579	8.92	15.59	-0.8	0
21	50	5.584847069	49.81	1889525	5.26	15.77	0.49	0
22	49	6.298785927	50.67	1614306	7.09	15.91	1.52	1

Fig. 1. Sample of data

ROAA is the net income over the average assets while ROAAIndex is the binary class of the bank profit in such that ROAAIndex is 0 when the ROAA is below 1, otherwise ROAAIndex is 1. For confidentiality, names of the bank are hidden and not included in the prediction model. The distribution of the DV is given in Table 3, which most of the cases from the dataset is occupied with High Bank Profit. Therefore, it is important in this research to observe the performances of each machine learning algorithm in predicting case 0 or Low Bank Profit by measuring the precision and recall for the class. Section 2.3 describes precision and recall for each class.

Table 3
Distribution of the bank profit class as dependent variable

Class	Count	Percentage
1 (Bank High profit)	229	64.15%
0(Bank Low profit)	129	35.85%

2.2 The Machine Learning Algorithms

By using AutoModel RapidMiner, ten suggested machine learning algorithms were provided for the given dataset but only the five best accurate were selected. This research used three types of the tree-based machine learning algorithms namely Decision Tree (DT) Random Forest (RF) and Gradient Boosted Trees (GBT) to be compared with another non-family tree-based algorithms namely Logistic Regression (LR) and Generalized Linear Model (GLM). Unlike Logistic Regression and Generalized Linear Model, hyper-parameters preliminary analysis is essential for tree-based machine learning. As a tree-based algorithm, the common hyper-parameter is Maximal Depth. Number of Trees is an additional hyper-parameter for RF and GBT while Learning Rate is only used in GBT. Table 4 lists the optimal setting for the hyper-parameters.

Table 4
The optimal hyper-parameters for Decision Tree and Random Forest

Machine Learning Algorithm	Hyper-parameters	Best Error rate %	Hyper-parameters	Worst Error rate (%)
DT	Maximal Depth=10	27.1	7	35.5
RF	Number of Trees=140 Maximal Depth=2	26.2	Number of Trees=20 Maximal Depth=2	32.2
GBT	Number of Trees=90 Maximal Depth=2 Learning Rate=0.100	23.8	Number of Trees=30,90,150 Maximal Depth=4,7 Learning Rate=0.001	30.8

As depicted in Table 4, the highest error rate reached 35.5%, 32.3% and 30.8% for the three algorithms respectively. The best error rate for DT was 27.1% with Maximal Depth 10 and 26.2% for RF with Maximal Depth 2. Additional hyper-parameter for RF and GBT is Number of Trees. RF can produce best results with Number of Trees 140 while GBT uses 90 Number of Trees. More than the two hyper-parameters, GBT has Learning Rate, which the optimal setting was 0.1.

2.3 Training and Testing the Machine Learning

For separating the training and testing datasets, the research used a split training approach with a ratio of 71:29 percentages. Therefore, from the 357 data, 254 of them were used for the machine learning training and 103 were used in the machine learning testing. The common metrics used to evaluate machine learning algorithms are accuracy and classification error. Accuracy and classification error present how good the algorithm is in doing the prediction without specifying the class group. Thus, precision and recall are used to measure the performances of algorithms belonging to a specific class. For example, for High Bank Profit, precision is the fraction of Class 1 that are predicted from all the cases while the class recall is the number of High Bank Profit cases that are correctly predicted as Class 1. Additionally, Area Under Curve (AUC) measures the ability of machine learning to distinguish between the two classes. The trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) distinguish the machine learning performances for the two classes. TPR is the proportion of correct prediction for High Bank Profit. On the other hand, FPR is the total number of incorrect predictions for Low Bank Profit. Table 5 is the confusion matrix to illustrate FPR and TPR. Additionally, the precision and recall for each class are also given in Table 5. Time to Complete (TTC) is to measure the efficiency of machine learning in completing all the processes.

Table 5
The confusion matrix of the bank profit classification in the prediction model

	Real class 1 (High Bank Profit)	**Real class 0 (Low Bank Profit)**	**Class precision**
Predicted as 1 (High Bank Profit)	True Positive (TP)	False Positive (FP)	*TP/(FP+TP)*
Predicted as 0 (Low Bank Profit)	False Negative (FN)	True Negative (TN)	*TN/(TN+FN)*
Class Recall	*TP/(TP + FN)*	*TN/(TN + FP)*	

TP is the number of correct predictions for High Bank Profit while TN is the number of correct predictions for Low Bank Profit. On the other hand, FN is the number of incorrect (false) predictions for High Bank Profit while FP is the number of incorrect (false) predictions for Low Bank Profit. The formula to calculate TPR and FPR is depicted in Eq. (1) and Eq. (2) respectively.

TPR=Total TP/Total Positive Cases = TP/(TP+FN) (1)

FPR=Total FP/Total Negative Cases = TP/(FP+TN) (2)

3. Results and Discussion

There are three sets of results presented from the study. Firstly, the results of performances of the machine learning to classify both cases of bank profitability either High Bank Profit or Low Bank Profit is provided. Secondly, the precision and recall of each algorithm will be presented. Thirdly, the variance of features in each of the machine learning will be discussed.

Table 6
The performances result

Algorithm	Accuracy (%) (+-Std.Dev)	Classification Error (%) (+-Std.Dev)	AUC (+-Std.Dev)	Time To Complete (s)
GLM	71.9 (0.095)	28.1 (0.095)	0.838 (0.095)	0.825
LR	71.9 (0.095)	28.1 (0.095)	0.838 (0.095)	0.869
DT	74.8 (0.062)	25.2 (0.062)	0.818 (0.077)	1
RF	75.7 (0.098)	24.3 (0.098)	0.812 (0.076)	13
GBT	76.7 (0.061)	23.3 (0.061)	0.819 (0.084)	13

In general, all machine learning algorithms have achieved good accuracy results (above 70%) with considerably less errors (lower than 30%), mainly DT, RF and BGT that used a tree-based paradigm for constructing the classification of Bank Profit. The most outperformed algorithm is BGT with the highest accuracy and lowest classification error (76.7%, 23.3%). However, in terms of AUC, linear-based machine learning from GLM and LR presented better performance than the tree-based algorithms. Presented by small Standard Deviation from all the results indicated the reliability of the prediction models by all the machine learning algorithms. Due to the in complex structure of the linear algorithm compared to tree algorithms, the TTC from the GLM and LR were slightly faster than DT, RF and GBT. But all considerations are efficient within less than a minute.

Second set of results is that the precision and recall for each class of Bank Profit can be measured based on the confusion matrix (as labelled in Table 5) that were generated from each machine learning algorithm as listed in Table 7. As expected, the class precision for detecting Low Bank Profit in all machine learning algorithms is lower than the results for predicting the High Bank Profit. However, even with the very small numbers that are given for the machine learning training with the Low Bank Profit class, the precision results from GLM, LR and DT are considerably good enough (64% and above) and high from RF and GBT (above 77%). Similarly, to precision, recall of Low Bank Profit was rather low compared to recall of High Bank Profit but DT has achieved 61.11%, which indicates good recall ability.

Table 7
Confusion matrix with precision and recall

	Real class 1 (High Bank Profit)	Real class 0 (Low Bank Profit)	Class precision (%)
	GLM		
Predicted as 1 (High Bank Profit)	58	20	74.36
Predicted as 0 (Low Bank Profit)	9	16	64.00
Class Recall (%)	86.57	44.44	
	LR		
Predicted as 1 (High Bank Profit)	58	20	74.36
Predicted as 0 (Low Bank Profit)	9	16	64.00
Class Recall (%)	86.57	44.44	
	DT		
Predicted as 1 (High Bank Profit)	55	14	79.71
Predicted as 0 (Low Bank Profit)	12	22	64.71
Class Recall (%)	82.09	61.11	
	RF		
Predicted as 1 (High Bank Profit)	62	22	73.81
Predicted as 0 (Low Bank Profit)	3	16	84.21
Class Recall (%)	95.38	42.11	
	GBT		
Predicted as 1 (High Bank Profit)	61	19	76.54
Predicted as 0 (Low Bank Profit)	5	17	77.27
Class Recall (%)	92.54	47.22	

Based on the results in Table 7, the F1 score for each machine learning algorithm in predicting each class is depicted in Table 8. The F1 score is a useful metric for evaluating the performance of the classification model, particularly when there is an uneven distribution of classes in the data. Eq. (3) is the formula for F1 score that considers both precision and recall.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{3}$$

The results in Table 8 confirm that the model is better at predicting high bank profits (class 1) than low bank profits (class 0). This is indicated by the higher F1 score for class 1 compared to class 0, which indicates that the model is better able to balance precision and recall for class 1 than for class 0. This may be due to a variety of factors, such as the lack of relevant data or variables for predicting low bank profits or the presence of noisy data that could affect the accuracy of the model. To improve the accuracy of the model for predicting low bank profits, additional data and variables could be added to the analysis to increase the predictive ability of the model.

Table 8
F1 score

Algorithm	Class 1 (**High Bank Profit**)	Class 0 (Low Bank Profit)
GLM	0.80	0.52
LR	0.81	0.64
DT	0.81	0.60
RF	0.83	0.56
GBT	0.84	0.58

Lastly, the third set of results listed in Table 9 explains the features importance of each machine learning algorithm. It can be depicted that the most important feature came from management efficiency (MEf) followed by the size of the bank. It seems that bank characteristics are more important than corruption when the weights correlations from CI is the third lower out of the six features in GLM and LR, second lower in DT and became the lowest in RF and GBP.

Table 9
The weights of correlations of each feature in bank profitability

Features	GLM	LR	DT	RF	GBT
CI	0.046	0.046	0.029	0.023	0.018
CDPG	0.028	0.029	0.033	0.017	0.033
MEf	0.315	0.314	0.205	0.271	0.211
Liquidity	0.004	0.005	0.053	0.108	0.035
NPL	0.005	0.005	0.026	0.044	0.039
Size	0.102	0.103	0.067	0.061	0.121

4. Conclusions

This paper presents significant findings of research that are concerned with corruption factor in understanding bank profitability based on Malaysia evidence. Acknowledging that machine learning techniques can be used to support fast and reliable prediction tasks, to identify which algorithms are suitable and how important corruption features can be seen from the prediction model is a valuable research initiative for further in-depth analysis.

To improve the accuracy of the predictive model, there is a possibility of incorporating more data and variables related to corruption. This could involve identifying the most important and relevant corruption features and incorporating them into the machine learning algorithms used in the analysis. This research will be of great interest to researchers in the business and banking industry, as well as to the artificial intelligence domain, as it has the potential to expand our understanding of the relationship between corruption and bank profitability using machine learning techniques.

Acknowledgement
We acknowledge the Universiti Poly-Tech Malaysia and Accounting Research Institute (HICoE) and Ministry of Higher Education, Malaysia for providing the necessary financial assistance for this study.

References

[1] Zimelis, Andris. "Corruption research: A need for an integrated approach." *International Area Studies Review* 23, no. 3 (2020): 288-306. https://doi.org/10.1177/2233865920926778

[2] Bolarinwa, Segun Thompson, and Funmi Soetan. "The effect of corruption on bank profitability." *Journal of Financial Crime* 26, no. 3 (2019): 753-773. https://doi.org/10.1108/JFC-09-2018-0102

[3] Yakubu, Ibrahim Nandom. "Does corruption grease or sand the wheels of bank profitability in Ghana?." *Cogent Economics & Finance* 7, no. 1 (2019): 1701909. https://doi.org/10.1080/23322039.2019.1701909

[4] Asteriou, Dimitrios, Keith Pilbeam, and Iuliana Tomuleasa. "The impact of corruption, economic freedom, regulation and transparency on bank profitability and bank stability: Evidence from the Eurozone area." *Journal of Economic Behavior & Organization* 184 (2021): 150-177. https://doi.org/10.1016/j.jebo.2020.08.023

[5] Ali, Muhammad, Amna Sohail, Lubna Khan, and Chin-Hong Puah. "Exploring the role of risk and corruption on bank stability: evidence from Pakistan." *Journal of Money Laundering Control* 22, no. 2 (2019): 270-288. https://doi.org/10.1108/JMLC-03-2018-0019

[6] Bougatef, Khemaies. "Determinants of bank profitability in Tunisia: does corruption matter?." *Journal of Money Laundering Control* 20, no. 1 (2017): 70-78. https://doi.org/10.1108/JMLC-10-2015-0044

[7] Akins, Brian, Yiwei Dou, and Jeffrey Ng. "Corruption in bank lending: The role of timely loan loss recognition." *Journal of Accounting and Economics* 63, no. 2-3 (2017): 454-478. https://doi.org/10.1016/j.jacceco.2016.08.003

[8] Farooq, Mohammad Omar, Mohammed Elseoud, Seref Turen, and Mohamed Abdulla. "Causes of non-performing loans: the experience of gulf cooperation council countries." *Entrepreneurship and Sustainability Issues* 6, no. 4 (2019): 1955-1974. https://doi.org/10.9770/jesi.2019.6.4(29)
[9] El Mostafa, Hambi, and Faouzia Benabbou. "A deep learning based technique for plagiarism detection: a comparative study." *IAES International Journal of Artificial Intelligence* 9, no. 1 (2020): 81. https://doi.org/10.11591/ijai.v9.i1.pp81-90
[10] Tammouch, Ilyas, Abdelamine Elouafi, Souad Eddarouich, and Raja Touahni. "Centroid competitive learning approach for clustering and mapping the social vulnerability in Morocco." (2022).
[11] Aliyu, Muhammad Abdu, Musa Mohd Muhammad Mokji, and Usman Ullah Ullah Sheikh. "Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning." *IAES International Journal of Artificial Intelligence* 9, no. 4 (2020): 670. https://doi.org/10.11591/ijai.v9.i4.pp670-683
[12] Abdali-Mohammadi, Fardin, Maytham N. Meqdad, and Seifedine Kadry. "Development of an IoT-based and cloud-based disease prediction and diagnosis system for healthcare using machine learning algorithms." *IAES International Journal of Artificial Intelligence* 9, no. 4 (2020): 766. https://doi.org/10.11591/ijai.v9.i4.pp766-771
[13] Skouta, Ayoub, Abdelali Elmoufidi, Said Jai-Andaloussi, and Ouail Ochetto. "Automated binary classification of diabetic retinopathy by convolutional neural networks." In *Advances on Smart and Soft Computing: Proceedings of ICACIn 2020*, pp. 177-187. Springer Singapore, 2021. https://doi.org/10.1007/978-981-15-6048-4_16
[14] Chan, Nicholas Khin-Whai, Angela Siew Hoong Lee, and Zuraini Zainol. "A framework for predicting employee health risks using Ensemble Model." *International Journal of Advanced and Applied Sciences* 8, no. 9 (2021): 29-38.
[15] Makruf, Moh, Arif Bramantoro, Hasan J. Alyamani, Sami Alesawi, and Ryan Alturki. "Classification methods comparison for customer churn prediction in the telecommunication industry." *International Journal of Advanced and Applied Sciences* 8, no. 12 (2021). https://doi.org/10.21833/ijaas.2021.12.001
[16] Barik, Lalbihari. "Data mining approach for digital forensics task with deep learning techniques." *Int J Adv Appl Sci* 7, no. 5 (2020): 56-65. https://doi.org/10.21833/ijaas.2020.05.008
[17] Slimene, Marwa Ben, and Abdulaziz Aljaloud. "Neural Network Estimation of Photovoltaic System Based on MPPT Controller." *IJAAS journal* (2019).
[18] Erdal, Hamit, and İlhami Karahanoğlu. "Bagging ensemble models for bank profitability: An emprical research on Turkish development and investment banks." *Applied soft computing* 49 (2016): 861-867. https://doi.org/10.1016/j.asoc.2016.09.010
[19] Wei, Jun, Tao Ye, and Zhe Zhang. "A machine learning approach to evaluate the performance of rural bank." *Complexity* 2021 (2021): 1-10. https://doi.org/10.1155/2021/6649605
[20] Ainan, Ummey Hany, and Md Nur-E-Arefin. "Prediction of Bank Performance Using Machine Learning and Genetic Algorithm Hybrid Models." In *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, pp. 1-6. IEEE, 2022. https://doi.org/10.1109/ICAEEE54957.2022.9836440
[21] Bolarinwa, Segun Thompson, and Funmi Soetan. "The effect of corruption on bank profitability." *Journal of Financial Crime* 26, no. 3 (2019): 753-773. https://doi.org/10.1108/JFC-09-2018-0102
[22] Ali, Mohamed Sami Ben, Fredj Fhima, and Ridha Nouira. "How does corruption undermine banking stability? A threshold nonlinear framework." *Journal of Behavioral and Experimental Finance* 27 (2020): 100365. https://doi.org/10.1016/j.jbef.2020.100365
[23] Murharsito, Murharsito, Fitri Ella Fauziah, Emanuel Kristijadi, and Rr Iramani. "Provincial corruption and local development bank performance." *Economic Journal of Emerging Markets* 9, no. 1 (2017): 66. https://doi.org/10.20885/ejem.vol9.iss1.art7
[24] Ali, Muhammad, Amna Sohail, Lubna Khan, and Chin-Hong Puah. "Exploring the role of risk and corruption on bank stability: evidence from Pakistan." *Journal of Money Laundering Control* 22, no. 2 (2019): 270-288. https://doi.org/10.1108/JMLC-03-2018-0019
[25] Yunan, Zuhairan Yunmi. "Does corruption affect Islamic banking? Empirical evidence from the OIC countries." *Journal of Financial Crime* 28, no. 1 (2020): 170-186. https://doi.org/10.1108/JFC-06-2020-0101