Advances in Social Science, Education and Humanities Research, volume 474
Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)

Building an Initiation of Financial Reporting System Using Blockchain and Classification Analysis to Financial Distress

Arum Handini Primandari[1,*], Masthura[1]

[1]*Department of Statistics, Universitas Islam Indonesia, Yogyakarta, 55584, Indonesia*
[*]*Corresponding author: primandari.arum@uii.ac.id*

ABSTRACT
Indonesia's regional government still uses a conventional financial reporting method by manually submitting some documents. As a consequence, there is no digital database provided. Our proposed technology is blockchain to be implemented for distributed tech financial reports. As popular as a cryptocurrency database system, Blockchain is a time-stamped series of immutable data records in chained blocks. As all the involved institutions (i.e., SKPD, PPKD, Regional Secretary, Regent/Mayor, BPK, and DPRD) will become nodes on the financial report system, its transparency and confidentiality are maintained. This research builds an initiation of the system using python programming. Adding the information about the regional financial condition, whether it experiences financial distress, or not will give the report overview. The prediction of financial distress is made by employing Support Vector Machine (SVM) with Radial Basis Function (RBF). In order to overcome the imbalance class, the Synthetic Minority Over-Sampling Technique (SMOTE) is applied. As a result, SVM with SMOTE can classify financial distress with the accuracy of the testing dataset is about 94.44%, the precision of 100%, and the kappa value of 82.34%.
Keywords: *Blockchain, Financial Distress, Financial Reporting System, SMOTE, SVM.*

1. INTRODUCTION

Local government financial reports have very vital functions, one of which is to describe the government's condition and as a way to account for financial performance to the public. The accountability process follows Government Regulation Number 24 of 2005 concerning Government Accounting Standards, which has been revised with government regulation Number 71 of 2010 concerning Government Accounting Standards. Regional financial reports are a means of reporting all regional activities, whether successful or unsuccessful, to measure the achievement of the organization's mission. Satuan Kerja Perangkat Daerah (SKPD) prepares the financial report from their region called LK-SKPD (Laporan Keuangan-SKPD). Badan Pemeriksa Keuangan (BPK) makes the final examination of LK-SKPD, which later becomes LKPD (Laporan Keuangan Pemerintah Daerah). The audit process of local government financial reports is a process of problem identification, analysis, and evaluation carried out independently by the BPK.

Currently, the financial reporting process uses manual methods by submitting documents from one party to another. As the impact, there is no accountable digital data record. Besides, This conventional method has the potential for fraud because the process involves many parties.

This research proposes blockchain technology to overcome these problems. Blockchain was first introduced in the database system contained in bitcoin transactions [0]. Blockchain is an immutable and transparent distributed ledger that makes no one argue with the truth [0]. Blockchain consists of several blocks that are related to one another. It is built sequentially, where each block has a hash value from the previous block. If someone wants to corrupt a blockchain system, they need to change every block in the chain across all of the chain's distributed versions. Therefore, it is almost impossible to change, hack, or cheat the system [0]. The database on the blockchain system is publicly in the network, which means the copy is not only owned by one user but many users. If cheating occurs, the database owned by the user who committed the fraud will be different from the database owned by other users. Thus, the database contained in these users is considered invalid.

We make a blockchain prototype using a python programming on a local computer. There are two central systems in the prototype, namely a user interface for inputting data and a database for storing all financial reporting data. SKPD inputs the initial data published to the network. If there is a correction to a data value, a party can do the mining process and create a new block. All parties on the network will know any changes in the data [0][0].

Regional financial health analysis can be done based on LKPD data. The result of the analysis produces early warning regarding regional financial distress. It is an inability of the government to provide services to the public according to service quality standards that have been previously set. This inability is because the government does not have the availability of funds to invest in the infrastructure used in providing services to the public. This matter is due to the low amount of capital expenditure spent.

The problem of financial distress relates to bankruptcy. If local governments know earlier about the indication of bankruptcy, they can anticipate it sooner.
Support Vector Machines (SVMs) are empowered to predict financial distress. It is one of the supervised learning used for classification and regression [0][0][0]. Put merely, SVMs construct a hyperplane on high dimensional space that can be used for classification, regression, or any task. The suitable kernel function helps to build an optimal hyperplane that separates classes [0].
This study uses the financial data of Aceh Province in 2014-2017. The financial report variables are processed into research variables, namely financial ratio analysis, fiscal decentralization, local revenue effectiveness, shopping efficiency ratio, regional financial efficiency ratio, and capital expenditure efficiency ratio. Based on the labeled data, the local government's proportion experienced financial distress and not experienced financial distress is an imbalance. Therefore, we employ SMOTE (Synthetic Minority Oversampling Technique) to overcome the imbalance issue.
The remainder of the paper is organized as follows. Section 2 describes the data and the research steps, also introduced to the SMVs and SMOTE. Part 3 consists of two main topics, namely building blockchain prototypes and making financial distress predictions. Section 4 presents concluding remarks.

2. MATERIAL AND METHODS

2.1. Material

This study uses secondary data, namely data of Budget Realization Reports in district/city in Aceh Province 2014-2017. The data's source is the publication of the Central Bureau of Statistics of Aceh Province's on its official website, aceh.bps.go.id. The following is the research flowchart.

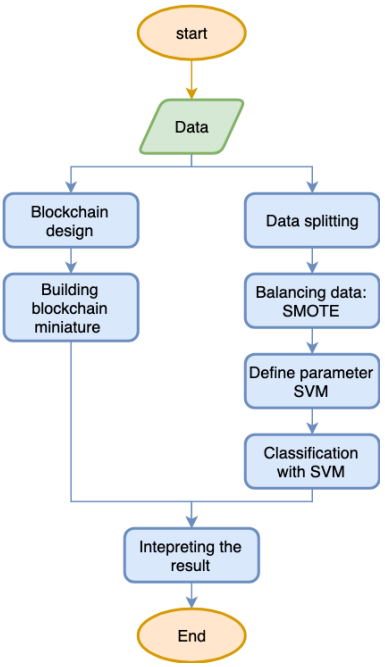

Figure 1. Research Flowchart

Bellow describe each steps:

1. The data of budget realization report consists regional revenue budget, realization of regional revenue, original regional income, realization of original regional income, total regional expenditure budget, realization of total regional expenditure, personnel expenditure budget, realization of personnel expenditure budget, goods and services expenditure budget, realization of goods and services expenditure, budget capital expenditure, capital expenditure realization, and financial condition. There are total 12 variables.
2. The steps are divided into two main cores:
    a. The building of blockchain prototype/miniature
        i. Designing the blockchain's user interface and the database includes the variable to be inputted on the system, the system's interface, and the order of reporting.
        ii. The building of the blockchain system uses python programming in the computer's localhost. It means both the user interface and the database utilizing browser interface (such as Chrome, Mozilla Firefox, and so all) running in localhost. Apart from budget realization, we need to input the district's identity and information on whether the district is experiencing financial difficulties.
    b. The classification analysis
        i. Before splitting, we calculate all the needed financial ratios from the data to predict financial distress. Its process also determines the dependent and independent variables. The dependent variable is the districts' financial condition (financial

distress or not). In contrast, the independent variable is financial ratios that are fiscal decentralization (X1, nominal), regional original income effectiveness ratio (X2, nominal), shopping efficiency ratio (X3, nominal), the regional financial efficiency ratio (X4, nominal), and capital expenditure efficiency ratio (X5, ratio). Most of the ratios become a nominal scale because we match the ratio value to the government regulation category.

ii. Data splitting process splits the data into training set and testing set with ratio 0.8:0.2.

iii. Because of the imbalance issue, the training set needs to be balanced using Synthetic Minority Over-Sampling Technique (SMOTE).

iv. SVMs uses cost and gamma as their parameters. The combined parameters are optimized to get the best classification.

v. After setting the model with the training set, its performance is evaluated using the testing set. The confusion matrix describes the performance of the model to classify new data.

2.2. Support Vector Machines (SVMs)

As the most popular kernel-based approach, SVMs have been proved to perform well in various application. The whole theories can be described as follows: searching an optimal hyperplane to satisfy the classification request, then using a particular algorithm to maximize the separation's margin beside the optimal hyperplane while ensuring the accuracy of correct classification. According to the theories, we can effectively classify the separable data into different classes [0][0][0][0][0][0].

Suppose that we have training set consists of N samples $\{x_i, y_i\}_{i=1}^{N}$, where $x_i \in \Re^d (i = 1, 2, \dots, N)$ is the $i-$th input sample and $y_i \in \{1, -1\}$ is the class label. Assume that the data is linearly separable in the original space, so that the decision function define as follow

$$y_i(\langle w, x_i\rangle + b) \geq 1 \quad (1)$$

where w is the vector weight, b is bias term, and $\langle w, x\rangle$ is the dot product of vector w and x.

There exist more than one hyperplane that capable to separate the two classes. However, only one of them define as best separator that maximizing the margin between two classes. Therefore, the SVMs training is to obtain w and b which maximize the separation.

$$\min \mathcal{J}(w) = \frac{1}{2}\|w\|^2 \text{ so that } y_i(\langle w, x_i\rangle + b) \geq 1, \forall_i \in [1, 2, \dots, N] \quad (2)$$

The assumption of using the parameter optimization in equation (2) is training data separable linearly. However, most of problem are linearly non-separable cases. In attempt to accommodating misclassification, the formulation introduce potsitive slack variable. Rewriting the equation (1) as follow

$$y_i(\langle w, x_i\rangle + b) + \xi_i \geq 1 \quad (3)$$

where $\xi_i \geq 0$. The data points placed on the wrong side of the margin boundary have a penalty. As we reduce the number of misclassification, the objective function from previously become this equation.

$$\min \mathcal{J}(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \text{ so that } y_i(\langle w, x_i\rangle + b) + \xi_i \geq 1, \forall_i \in [1, 2, \dots, N] \quad (4)$$

where the parameter C controls the trade-off between the slack variable and the size of the margin. Reformulating as a Lagrange, we have to minimize parameter w, b and ξ_i and to maximize α (where $\alpha_i \geq 0, u_i \geq 0$).

$$L_P \equiv \frac{1}{2}\|w\|^2 + +C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i(y_i(\langle w, x_i\rangle + b) + \xi_i) - \sum_{i=1}^{N}\mu_i\xi_i \quad (5)$$

2.3. Synthetic Minority Over-Sampling Technique (SMOTE)

Synthetic Minority Over-Sampling Technique (SMOTE) is one method in handling cases of imbalance datasets [0][**Error! Reference source not found.**]. The SMOTE method works by adding or generating artificial data or synthetic data. The purpose of adding synthetic data is that the number of minor data is equivalent to the major data. The synthetic data is generated based on the nearest neighbor.

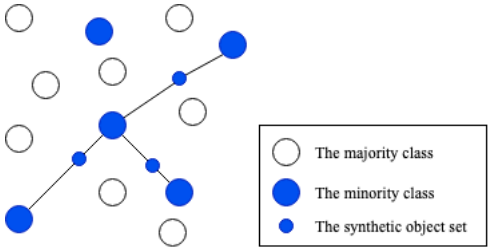

Figure 2. The Illustration of SMOTE

The K-nearest neighbors are the K elements whose Euclidean distance between itself and have the smallest weight along with the n-dimensional variable space of X. Randomly select one of the K-nearest neighbors and multiply corresponding Euclidean distance with a random

number between [0,1]. Finally, we add this value to the original value instance. The formula is written below [0][0].

$$New(x_i) = x_i + (\hat{x}_i - x_i) * \delta \quad (6)$$

where x_i is the sample from minority class used to generate synthetic data, $\hat{x}_i$ is the nearest neighbor, and δ is a random number between $[0, 1]$.

2.4. Blockchain

Blockchain is a distributed database storing growing records, controlled by several entities. It is a peer-to-peer (P2P) distributed ledger technology that establishes transparency and trust [0][0]. Generally, the blockchain acts as a trusted and reliable third party to maintain a shared state, mediate exchanges, and provide a secure computing engine. Blockchain is the underlying fabric behind bitcoin, which has three main components: the distributed network, a shared ledger, and digital transaction.

Blockchain contains a sequence of blocks holding transaction records [0]. A block consists of the block header and the block body. It has only one parent block since it has a previous block's hash in its header.

A user owns a pair of the private key and public key for doing some transaction using blockchain [0]. The private key, which shall be kept in confidentiality, is used to sign the transaction. The digitally signed transaction are broadcasted throughout the whole network. The transaction involves two phases that are signing phase and the verification phase. For instance, Alice wants to send a message to Bob; first the signing phase, Alice encrypts her message (data) with her private key, and then sends Bob the encrypted result and the original data. Second, the verification phase, Bob validates the value with Alice's public key.

3. RESULTS AND DISCUSSION

3.1. Blockchain Prototype

The prototype of smart contract of financial reporting system using blockchain describe as follows.

Table 1. The Description of The System's Prototype

Name	Description	Input Variables	Output Variables	Triggering Events

Financial report transfer	Send a report of budgeting and realization	• Address of the institution
		• The variable of report:
		  • Identity of financial report
		  • Regional income
		  • Regional original income (Pendapatan Asli Daerah: PAD)
		  • Total regional Expenditure
		  • Expenditure for employee expenses
		  • Expenditure for goods and service
		  • Capital expenditure
		  • Financial condition (whether or not it run into financial distress)
			The finansial report sent from A to B (the both are institutions)	When A wants to report the financial to B

There are some parties involves in financial reporting of Indonesia's regional. The process is describe bellow [0].

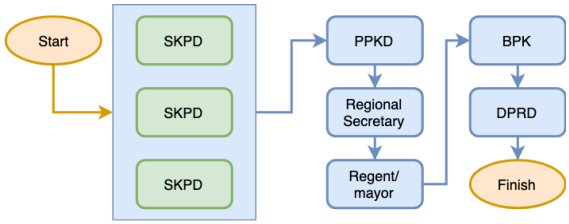

Figure 3. Financial Reporting Flow

Herewith is the description flow (Hariadi, Restianto, & Bawono, 2010):

1. S*atuan Kerja Perangkat Daerah* (SKPD/ Regional Work Unit) preparing the financial report from their region called LK-SKPD (*Laporan Keuangan-SKPD/* Financial Report-SKPD)
2. The LK-SKPD is submitted to *Pejabat Pengelola Keuangan Daerah* (PPKD/Regional Financial Management Officer). PPKD together with general treasurer make correction to the report, and after refinement called *Laporan Keuangan Pemerintah Daerah* (LKPD/ Regional Government Financial Reports).
3. The LKPD is turned over to regional secretary, then being submitted to regent/mayor.

4. The audit process of The LKPD is carried out by *Badan Pemeriksa Keuangan* (BPK/Audit Board of the Republic of Indonesia).
5. After being audited, the LKPD is turned back to regent and *Dewan Perwakilan Rakyat Daerah* (DPRD/ Regional People's Representative Assembly). At this state, the LKPD is valid to be the financial report of a region.

All the parties involved in financial report become nodes in the network. The process of recording financial report in the blockchain system is carried out when the SKPD's financial statements are ready to be submitted to PPKD. PPKD, which acts as one of the nodes, will create a block that will be sent to the next node. Before the block is sent, the block is verified to determine whether the block is valid. If the block is valid, the database will record the data. The next node needs to mining the blockchain in order for it to get the information contained in the block. The following is the main diagram for the process of inputting data and updating database.

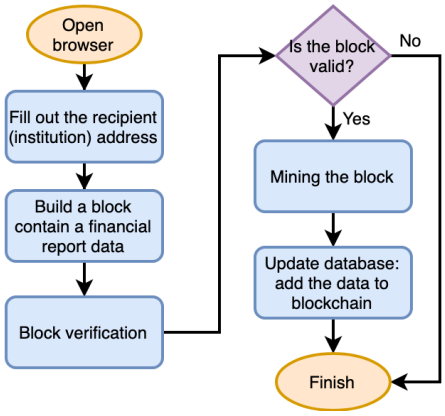

Figure 4. Main diagram of initiation of the blockchain system

The rule of financial report:

1. The block is built when financial reports are completed, such as monthly, semester, or yearly.
2. Each node have both a public key and private key. There are total six nodes (SKPD, PPKD, regional secretary, mayor, BPK, and DPRD). The order of building block for financial reporting is determined based on the Government Regulation Number 8 of 2006.
3. Each block that will be entered into the database contains the identity (such as sender public key, recipient address, name of district, year, and period) and financial data.

Figure 5. The User Interface for Building A Block

4. The verification is done by checking the digital signature's validity by decrypting it using the public key of each node listed in the system.

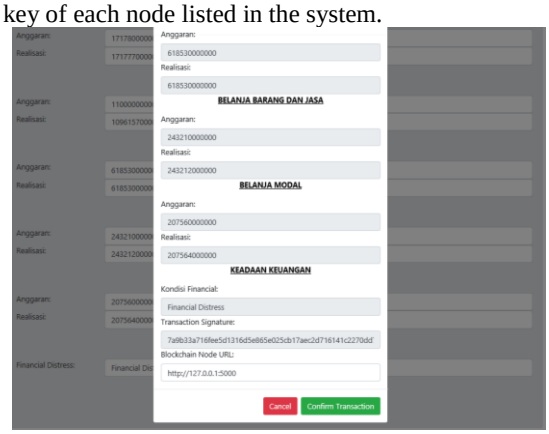

Figure 6. The Verification Process

5. Each node registered in the system keeps a copy of financial statement data in the database.

Figure 7. The Example of Database

3.2. Classification Analysis Using Support Vector Machine (SVM) with SMOTE

We divide the financial ratios data from the LKPD of each district in Aceh into two parts: a training set and a testing set. We get the best proportion between the training set and testing set through several trial errors, which result in 0.8:

0.2. The whole training set is 75, and the entire testing set is 17. Because of the imbalance of the response class in the training set, we apply the SMOTE technique. The proportion of districts that experienced financial distress and those that did not was 76%: 24%. After SMOTE is applied, the proportion of response classes become 50%: 50%.

SVM classification uses the kernel function. This function takes data as input and then transforms it into the required form. Considering the data's non-linear nature, we choose to use the Radial Basis Function (RBF). The experiment's error rates to find the values for the C and gamma parameters are presented in the following table.

Table 2. The Results of Error Rate

Nilai C *(cost)*	Gamma		
	0.09	**0.9**	**1**
0.01	0.520	0.537	0.537
0.1	0.340	0.477	0.517
1	0.237	0.257	0.257
10	0.277	0.240	0.24

Based on the smallest error value, SVM uses a C value of 1 and a gamma value of 0.09. The tuning parameter employs 10-fold cross validation. The measurement of accuracy and error of testing set is provided in the following table.

Table 3. The Accuracy of the SVM Classification with SMOTE

Accuracy	Precision	Recall	Error Rate	Kappa
94.44%	100%	75%	0.0556	82.34%

Based on **Table 3**, the SVM model's accuracy is 94.44%, which means the model is suitable for predicting the sample class. The precision value, also known as a positive predictive value, refers to the percentage of relevant results. Recall, known as sensitivity, refers to the percentage of total relevant results correctly classified by the algorithm. There is 75% of the financial distress case that is correctly predicted. Cohen's kappa statistic measures interrater reliability. It measures the agreement between two raters who each classify N sample into two mutually exclusive categories (financial distress or not).

4. CONCLUSION

Regional financial reporting systems in Indonesia can adopt blockchain decentralization technology. The distributed ledger accommodates the transparency and reliable system. All the parties involved in financial reporting become nodes that able to make a block contained data. The two main elements in the system are the user interface and the database. A party wishing to make a report creates a block. After the verification process, which is done by checking the digital signature's validity, the block is broadcasted in the network. The mining process by a node makes sure the data has been entered into the database. The weakness in the reporting system with blockchain is that correcting errors in data is not easy. A party wishing to correct a data error must create a new block, which means repeat the whole process.

The imbalance issue in the response training set is handled by applying SMOTE. It makes some synthetics set over minority class. Therefore the proportion of response class of experiencing financial distress or not is 50%:50%. SVMs' parameter (C and gamma) is obtained from trial and error. The kernel function applied in the SVMs model is RBF (Radial Basis Function).

The measure of the accuracy of financial distress classification using SVM with SMOTE is excellent. Likewise, the kappa value and error measurement. Even with the synthetic training set, the model has a good performance on the testing set.

ACKNOWLEDGMENT

Thanks to the Department of Statistics UII for supporting the fund of this research publication.

REFERENCES

[1] Bhagat, R. C., Patil, S. S. (2015). Enhanced SMOTE Algorithm for Classification of Imbalanced Big-Data using Random Forest. International Advance Computing Conference (IACC). 403-408. IEEE.

[2] Cao, Y., Wan, G., Wang, F. (2011). Predicting Financial Distress of Chinese Listed Companies Using Rough Set Theory and Support Vector Machine. Asia-Pasific Journal of Operational Research. 28(1): 95-109.

[3] Chawla, N., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. 16: 321-357.

[4] Demidova, L., & Klyueva, I. (2017). Improving the Classification Quality of the SVM Classifier for the Imbalanced Datasets on the Base of Ideas the SMOTE Algorithm. Seminar on Systems Analysis. 1-4. EDP Sciences.

[5]Demidova, L., Klyueva, I. (2017). SVM Classification: Optimization with the SMOTE Algorithm for the Class Imbalance Problem. 6th Mediterranean Conference On Embedded Computing(MECO). Bar. Montenegro: IEEE.

[6] Hamalik, O. (2001). Perencanaan Pengajaran Berdasarka Pendekatan Sistem. Bandung: Bumi Akasara.

[7] Hanifatunnisa, R. (2017). Perancangan Dan Implementasi Sistem. Bandung: Institut Teknologi Bandung.

[8] Hariadi, P., Restianto, Y. E., Bawono, I. R. (2010). Pengelolaah Keuangan Daerah. Jakarta: Salemba Empat.

[9] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., Xu, W. (2018). Application of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics & Proteomics 15: 41-51.

[10] Karamitos, I., Papadaki, M., & Barghuti, N. (2018). Design of the Blockchain Smart Contract: A Use Case for Real Estate. Journal of Information Security. 177-190.

[11] Lee, M.-C., To, C. (2010). Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress. International Journal of Artificial Intelligence & Applications (IJAIA). 1(3): 31-43.

[12] Liu, X. (2010). A Modified Support Vector Machine model for Credit Scoring. International Journal of Computational Intelligence Systems. 3(6): 797-803.

[13] Mathew, J., Luo, M., Pang, C. K., Chan, H. L. (2015). Kernel-Based SMOTE for SVM Classification of Imbalanced Datasets. IECON2015-Yokohama. 9-12.

[14] Nakamoto, S. (2019). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved from www.Bitcoin.Org: https://bitcoin.org/bitcoin.pdf. (Accessed March-23-2019).

[15] Nugroho, A. S. (2007). Pengantar Support Vector Machine.

[16] Scholkopf, B., Smola, A. J. (2001). Learning with Kernels. London: The MIT Press Cambridge. Massachusetts.

[17] Sun, J., Fujita, H., Chen, P., Li, H. (2016). Dynamic financial distress prediction with concept drift based on time weighting combined with adaboost support vector machine ensemble. Knowledge-Based Systems. 1-11.

[18] Watanabe, H., Fujimura, S., Atsushi, & Kishigami, J. (2015). Blockchain Contract: A Complete Consensus using Blockchain. 4th Global Conference on Consumer Electronics (GCCE). 577-578. IEEE.

[19] Wijaya, D. A. (2016). Bitcoin Tingkat Lanjut. Medan: Puspantara.

[20] Witten, H., Frank, E. (2005). Data Mining: Praticial Machine Learning Tools and Techniques Second Edition. San Fransisco: Morgan Kauffman.

[21] Yang, C., Yang, J., Ma, J. (2020). Sparse Least Square Support Vector Machine With Adaptive Kernel Parameters. International Journal of Computational Intelligence Systems. 13(1): 212-222.

[22] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. 6th International Congress on Big Data. 557-564. IEEE.