INNOVATIVE: Journal Of Social Science Research
Volume 3 Nomor 5 Tahun 2023 Page 814-830
E-ISSN 2807-4238 and P-ISSN 2807-4246
Website: https://j-innovative.org/index.php/Innovative

Implementation of Machine Learning for Sharia financing Scoring in Indonesian MSME sectors

Gery Abdel Gafar[1✉], Gunawan Wang[2]
Information System Management Department, Binus Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia 11480
Email: gery.gafar@binus.ac.id[1✉]

Abstrak

Kemajuan teknologi melalui revolusi industri 4.0 diharapkan menjadi kunci dalam menyelesaikan permasalahan sulitnya usaha mikro menjangkau permodalan bank syariah, keterbatasan informasi dalam memenuhi standar analisis kelayakan pembiayaan dan kepemilikan aset sebagai agunan menjadi faktor penghambat utama dalam mencapai tujuan. sektor UMKM. Kehadiran teknologi AI – Machine learning diharapkan dapat membuat proses analisis pembiayaan menjadi lebih efektif dan efisien dalam mengolah informasi yang diperoleh. Pada PT Bank XYZ Syariah rasio pembiayaan terhadap sektor UMKM masih sebesar 20% dari total eksposur pembiayaan yang disalurkan (Laporan Keuangan PT Bank XYZ Syariah Tahun 2020) penetrasi yang masih cenderung kecil disebabkan oleh proses analisa yang panjang dengan tingkat keberhasilan yang kecil sehingga diperlukan suatu model yang dapat membantu perusahaan dalam menyaring pengajuan yang memiliki tingkat keberhasilan yang tinggi. Pada tulisan ini akan dibuat model credit scoring berdasarkan data historis di PT Bank XYZ Syariah menggunakan beberapa algoritma klasifikasi, dengan mengacu pada CRISP -DM framework data akan diproses sedemikian rupa sehingga dapat menghasilkan dataset berkualitas kuat. Model dibuat menggunakan tools RapidMiner dengan dua skenario pengujian sehingga menghasilkan beberapa model dengan akurasi yang baik, berdasarkan hasil pengujian dari dataset yang tersedia model klasifikasi hutan acak menghasilkan akurasi yang paling baik diantara yang lainnya. Credit scoring yang dimodelkan dapat membantu perusahaan untuk dapat melakukan screening awal terhadap setiap pengajuan yang ada.

Kata Kunci: *Credit Scoring, UMKM, Algoritma Klasifikasi, Machine Learning, Bank Umum Syariah*

Abstract

Technological progress through the industrial revolution 4.0 is expected to be the key in solving the problem of difficulties for micro-businesses to reach Islamic bankcapital, limited information in meeting the standards of analysis of financing feasibility and asset ownership as collateral becomes the main obstacle factor in the MSME sector. The presence of AI - Machine learning technology is expected to make the financing analysis process more effective and efficient in processing the information obtained. At PT Bank XYZ Syariah financing ratio to MSME sector is still 20% of the total financing exposure channelled (Financial Statement of PT Bank XYZ Syariah in 2020) penetration that still tends to be small is caused by a long analysis process with a small success rate, so a model is needed that can help companies to filter submissions that have a high success rate.In this paper, the credit scoring model will be created based on historical data in PT Bank XYZ Syariah uses several classification algorithms, with reference to the CRISP-DM framework the data will process in such a way that it can produce a robust quality dataset. The model was created using RapidMiner tools with two testing scenarios, resulting in several models with good accuracy, based on the results of testing from available datasets random forest classification model produces the best accuracy among others. The Modelled credit scoring can help the company to be able to conduct an initial screening of every existing submission.
Keywords: *Credit Scoring, MSMEs, Classification Algorithm, Machine Learning, Islamic Bank*

INTRODUCTION

Indonesia, a developing country with the world's top 5 economic growth and the largest Islamic population, has a significant number of MSMEs, which are pillars of the Indonesian economy. However, the distribution of sharia financing to these sectors is low, with the corporate business sector and individual mortgage sector receiving the largest financing distribution. This research aims to implement credit scoring based AI machine learning for Sharia financing in Indonesia, particularly for MSMEs, to help them simulate business conditions and assess if their businesses meet the standard requirements for applying for sharia financing at Islamic banks.

The authors used data mining with the CRISP-DM method to build credit scoring, using data from PT. Bank XYZ Syariah. The data was processed through six stages of CRISP-DM, including business understanding, data understanding, data preparation, building model, testing and evaluation. The Credit Scoring model was built using several classification methods in three stages: pre-processing data, developing the model(s), and extracting and deploying the model using a rapid miner machine learning tool.

The research contributes to solving the problem by providing entrepreneurs with an idea of how their business can meet the requirements set by Islamic banks, helping them become bankable.

LITERATURE REVIEW

To support this research, literature is needed as a reference and as a comparison with the results of the research. In this case, the literature used as a reference, Including Sharia Financing, Credit Scoring, data mining, Machine learning, and classification algorithms.

Sharia Financing

In a nutshell, Islamic banking could be a framework that takes after Islamic Laws (Sharia) standards and Islamic-based financial matters. Islamic standards direct that cash loaning (interest-based) as well as contributing to businesses that are considered haram (illegal) are prohibited. Conventional financing benefits by charging intrigued on the advance, meaning that borrowers will conclusion up paying back more than they borrow. Sharia financing is instep based on making a benefit through many commodities. The buyer inquires the bank to purchase the precise thing that they need to be financed, which is at that point sold to them at an extra markup. The buyer is at that point inquired to pay for the thing in installments or through a "rent-to-own" plot. All usually known as the rule of Murabaha and is the concept of giving financing to buyers based on pre-determined benefits instead of managing with intrigued(Hosen et al., 2019).

Credit Scoring

Credit scoring is a crucial factor in determining an individual's or small business's financial health, influencing various transactions such as contracts, autos, credit cards, and direct loans.

Over the years, the bank has gathered data on client default behavior, including birth date, sexual introduction, compensation, and trade status. All this data has been organized fun to the super huge database (eg, social) or data dissemination center (Maldonado et al., 2020).

Credit scoring, originating from credit reports, estimates credit events like borrowers being 90 days late in multiple trade lines for two times worn. Loan scores for data-mine credit reports to induce this assumption, which means they do almost all jobs to produce predictions(Onay & Öztürk, 2018).

The credit score evaluation model will be utilized to determine the necessity of determining the credit utility or predicting future default risk. To summarize, credit score scores are the senses of primary danger management for financial institutions for optimally manipulating, understanding, and modeling the risk of credit scores found(JunediHutagaol&Mauritsius, 2020)

Data Mining

Data mining according to Larose in the book 'Data Algorithms Mining, Data Mining is a field of several scientific fields that brings together patterns, statistics, databases, and visualization for problem solving retrieval of information from large databases (Dastile et al., 2020).

Information based society is defined as a society whose major part is involved in mental activities rather than physical ones. Such societies mostly focus on information activities including collecting, process, generation, registering, transmission and management of information and the large portion of the budget is spent on information procedures(Chacko et al., 2020). one of the most prominent method whereby useful pattern that can be recognized in the data is the data mining (Li & Chen, 2020).

Machine Learning

Machine learning is a mathematical technique that improves performance in tasks by training data, forming predictions and decisions. It uses sophisticated algorithms like structured division and neural networks. Automation in machine learning technology drives businesses, such as cargo transport calculations, without human intervention. This technology is promising for providing more effective solutions than working energy. On the other hand, machine learning nevertheless able to be a problem since the fall of automated tradingplatform in the United States stock market (Varol et al., 2017).

Machine learning can improve response time by analyzing and preprocessing data from circulation input, despite limitations in the IoT environment such as heart computing and storage. The term "computing fog" or "edge" was mistakenly used to refer to the function of the edge of the IoT network.(Ali & Ishak, 2020), This technology's design is more portable and practical for implementing in various devices that require a function in Machine Learning.

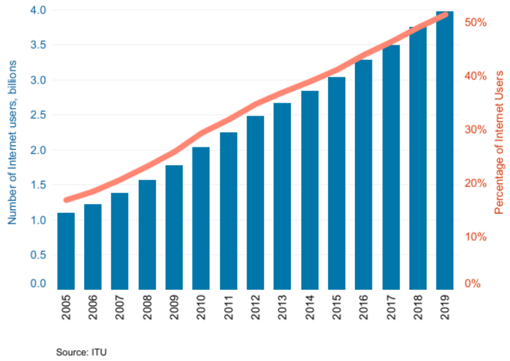

Figure 1 IoT Connected devices installed worldwide (ITU, 2020)

The picture above shows how IoT ecosystem are increasing from year to year, illustrating that the use of technology is now increasingly advanced and approaches almost every aspect of business processes and operations

Classification algorithm

Some algorithms that are often used in classifying data are Decision Tree, Naïve Bayes, Logistic regression, and Deep Learning (Neural Networks).

Decision Tree

The decision tree may be a well-known procedure and has had numerous fruitful applications to real-world issues (Tsai & Hung, 2014). The decision tree is a symbolical learning technique that regulates the news extracted from the training dataset in the hierarchical structure which consists of nodes and consequences. Because the results of the decision tree can be set in the form of a tree or law, simple to understand what will occur for decision trees(Machine Learning by Thomas Mitchell, 1997). Additionally, decision tree will be able to build models using datasets including numerical and categorical data (Lorena & de Carvalho, 2007).

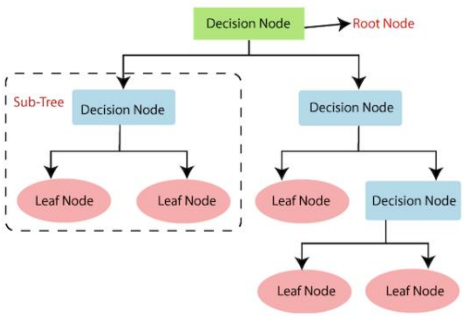

Figure 2 Decision Tree result model

A choice tree uses square and circular leaf hubs to represent choices made, with circular hubs indicating uncertain results. However, it's not commonly used to predict outcomes, except in information analysis.

Naïve Bayes

The Naïve Bayes Classifier, a classification strategy based on the Bayes hypothesis, predicts future opportunities based on past experiences, assuming the autonomy of each condition/event.

Naive Bayes Classifier works truly well in comparison to other classifier models. Xhemali, Hinde Stone proves in his journal says that "Naïve Bayes Classifier includes a way better exactness rate than other classifier models"(Xhemali et al., 2009). Naive Bayes has advantaged this method work for both quantitative and qualitative data, train the model to builddoesn't require a lot of data.

Random Forest

The first solving procedure for randomized forest decisions designed in 1995 by Tin Kam Ho using a random method of substrace, which, in Ho's formulations(Kam Ho, n.d.), is a way to apply the "subordinate stochastic" approach to the description proposed by Eugene Kleinberg(Kleinberg, 1990).

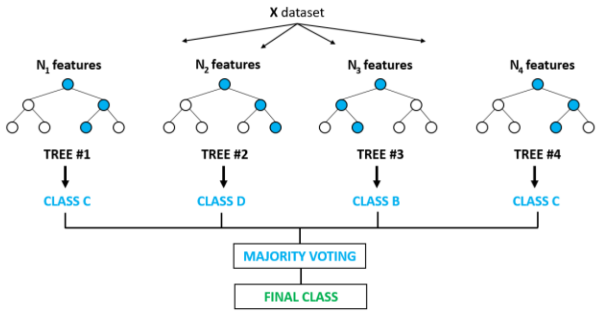

Figure 3 Random Forest Classifier model

The extension of the solving procedure was developed by the Leo Breiman and Adele Cutler, which registered "Random Forest" into a trademark in 2006 (in 2019, owned by Minitab, Inc.). The extension combines inspiration "bagging" Breiman and the choice of rando features, introduced first by Ho and then independently Amit and Geman to create a decision tree collection with controlled variance(Breiman et al., 1998).

Random forest is often used as a "Blackbox" model in business because they produce reasonable predictions in many data while needing a little configuration.

Logistic Regression

Logistic regression, which means that statistical modeling techniques that are widely used, can form models using dichotomes and has proven to be a strong solving procedure (Lee et al., 2006). The details of the logistics regression installation can be found in other studies (Hosmer & Lemeshow, 1989).

As mathematical theorem, a logistic regression model predicts P(Y=1) as a work of X. It is one of the only machine learning calculations that can be utilized for different classification issues such as phising detection, weather forecasting, illnesprediction, etc(Nie et al., 2011).

Based on number of categories, Calculated relapse can be separated into the taking after sorts:

- Binominal or binary
- Multinominal
- Ordinal

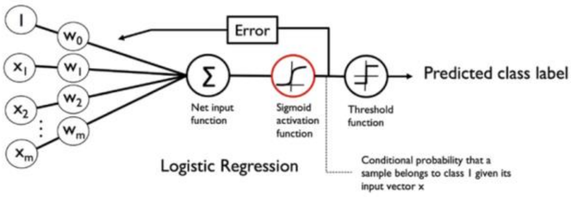

Figure 4 logistic regression classifier model

In simplest explanation of logistics regression is binary or binomial process of modeling the probability where the target or subordinate variable can only have two sorts which will be 1 or 0. This permits us to demonstrate the relationship between a few indicator factors and binary/binomial target factors. on logistic regression terms(Itoo et al., 2021)

Deep Learning

Deep learning is an advanced artificial neural network approach, featuring various architectures like multilayer, convolutional, and recurrent neural networks. Figure 5shows the two hidden layers neural network architecture(Ha & Nguyen, n.d.).

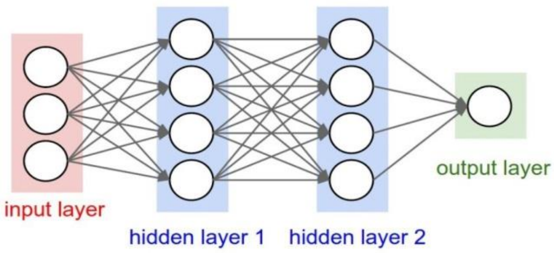

Figure 5 A deep multilayer perceptron neural network (Deep Learning Model)

Deep structured learning is a subset of machine learning that updates input data through layers. It can be used in image recognition software to encode features like nose and eyes, and assess optimal feature placement. It does not eliminate hand adjustment requirements; For example, various layers of layers and layers can deliver an incompatible level of abstraction(Dastile& Celik, 2021).

RESEARCH METODE

In this Research, the CRISP-DM method will be used as a reference. CRISP- DM represents Cross Industry Standard Process for Data Mining is a process model under the scope of information mining that is used by experts in businesses. CRISP-DM became one of the most popular research methodologies, especially in developing machine learning or deep learning models [7]. CRISP-DM is made up of six main processes, i.e., business understanding, data understanding, data preparation, modelling, evaluation, and deployment.

3.1 Business Understanding

Understanding existing business processes, AI Machine Learning credit scoring, and variables affecting sharia financing application approval or rejection is crucial. Knowledge about its benefits compared to conventional methods and its impact on business processes is also required.

3.2 Data Understanding

The author plans to use MSME applicant data for a machine learning-based Credit Scoring simulation process. The data will include Approval Status labels, integer, polynominal, and real variables for further analysis. There are 35variables used as training variables in the model. All these variables are described on the Table 1 below

Table 1 Variables of sharia fianancing dataset

No	Attribute name	Attribute Descripton
1	id	Identification for each data Row
2	first facility	Indicates whether the customer already has the first facility in the institution
3	second facility	indicate whether the customer already has a second facility in the institution
4	Group Facility	types of financing facilities taken by customers
5	Facility	financing facility products contained in the institution selected by the customer
6	financing purpose	the purpose of using financing funds submitted by customers to the institution
7	occupation	The type of work that the customer is engaged in
8	occupational type	specific type of work done by the customer (due to its focus on SME then there are only 2 types of specifications)
9	long established	the length of the customer's business or the age of the customer's business
10	company type	types of companies established by customers based on indonesian regulations
11	type business	business sector entered by customer company

No	Attribute name	Attribute Descripton
12	type Sub Sektor	specific business sectors engaged by customer companies
13	Maximum Facility	maximum ceiling financing that can be accessed by customers
14	installment	installments charged to customers every month
15	time period	the period of financing facility where the customer pays installments every month
16	age	age of the customer
17	educational	the level of education that has been taken by the customer
18	religion	religion embraced by the customer
19	count facilities	the number of financing facilities already owned by customers based on information from regulators
20	object type	objects that will be collateral and/or objects that will be the object of financing
21	property type	the type of property that is a guarantee and or property that becomes the object of financing

No	Attribute name	Attribute Descripton
22	collateral value	value of the collateral object
23	Margin	additional costs for providing financing facilities
24	marital status	customer's marital status
25	income	nominal income earned by customers each month
26	Join Income	income merger status registered in the financing agreement
27	Status PKS	the company's cooperation status with the institution
28	Payroll	whether the payroll system has cooperated with the institution
29	Dsr	value of debt-to-income ratio owned by customers
30	collateral certificate	certificate of collaterla that guaranteed in the institution
31	Program	the type of program offered by the institution to the customer
32	KTP	Proof of citizenship
33	NPWP	proof of registration in taxation
34	Approval	results of customer financing facility submission analysis
35	Ftv	nominal ratio value of financing to value of guaranteed collateral

3.3 Data Preparation

This research uses data from sharia financing application data at PT. Bank XYZ Syariah, mined using Export Wizard and imported into Microsoft Excel for data selection, cleansing, and formatting. Non-numeric training variables require data pre-processing in Microsoft Excel, Tableau, and Navicat for data transformation, visualization, and database conversion. The type of data from each attribute are represented as data variables on data structure described on theTable 2 below.

Table 2 type of format data for each variables

No	Attribute name	format variable
1	id	Integer
2	first facility	polynominal / Text
3	second facility	polynominal / Text
4	Group Facility	polynominal / Text
5	Facility	polynominal / Text
6	financing purpose	polynominal / Text
7	occupation	polynominal / Text
8	occupational type	polynominal / Text
9	long established	Integer
10	company type	polynominal / Text
11	type business	polynominal / Text
12	type Sub Sektor	polynominal / Text
13	Maximum Facility	Integer
14	installment	Integer
15	time period	Integer
16	age	Integer
17	educational	polynominal / Text
18	religion	polynominal / Text

No	Attribute name	format variable
19	count facilities	Integer
20	object type	polynominal / Text
21	property type	polynominal / Text
22	collateral value	Integer
23	Margin	Real
24	marital status	polynominal / Text
25	income	Integer
26	Join Income	polynominal / Text
27	Status PKS	polynominal / Text
28	Payroll	polynominal / Text
29	Dsr	Real
30	collateral certificate	polynominal / Text
31	Program	polynominal / Text
32	KTP	polynominal / Text
33	NPWP	polynominal / Text
34	Approval	polynominal / Text
35	Ftv	Real

The data sample was thoroughly examined and identified several variables that could be improved to create a more robust and accurate dataset, as the following:

firstfacility and *secondfacility*It is important whether the customer has or does not have facilities in the institution, but the input data is the account number of the customer's financing account that would be processed as integer type of data so would affect the accuracy of the learning model, therefore we group the variable into two categories of magnitudes, "Ada" or "Tdk" so that the varabel can be processed by the learning machine unbiasedly. The result of this pre-processing is illustrated using the following barchart;

Figure 6 Result of Data Pre-processing (First facility & Second facilitty)

The same scenario also applied to the *KTP* and *NPWP*. Where each represents the numberof customer identity for citizenship and for taxation, number identity has integer type of data so it would affect the quality of whole data. So we change and grouping the data to be "*ada*"&"*tdk*". The result of this pre-processing is illustrated using the following barchar;

Figure 7 Result of Data Pre-processing (KTP & NPWP)

The original dataset contained 197 thousand raw data points, with many missing labels, values, and randomized variables. To improve quality, we removed low-quality data and processed it using Microsoft Excel. We generated 40 thousand high-quality data, divided by 20,000 approved and 20,000 rejected data, for this research, the tabel below described the example of good variable of data:

Table 3 Example of eligible variable in final dataset

Type of Business	Approved	Reject	Total	% Approval
KONSTRUKSI	1326	1187	2513	3%
INDUSTRI PENGOLAHAN	938	981	1919	2%
JASA-JASA DUNIA USAHA	4971	5320	10291	12%
JASA-JASA SOSIAL MASYARAKAT	2376	2448	4824	6%
LISTRIK, GAS DAN AIR	224	167	391	1%
PENGANGKUTAN, PERGUDANGAN DAN KOMUNIKASI	726	844	1570	2%
PERDAGANGAN, RESTORAN DAN HOTEL	8902	8530	17432	22%
PERTAMBANGAN	93	141	234	0%
PERTANIAN, PERBURUAN DAN SARANA PERTANIAN	444	382	826	1%
Grand Total	20000	20000	40000	50%

Finally, the last pre-processing step was examined unnecessary variable such as *id* which represents the identification of each row for helping the authors cleaned and transformed the data in Microsoft excel, *id* variable has integer type of data, we noticed if don't remove the variable it would affect the performance of the model, so we removed *id* variable which described in figure 3

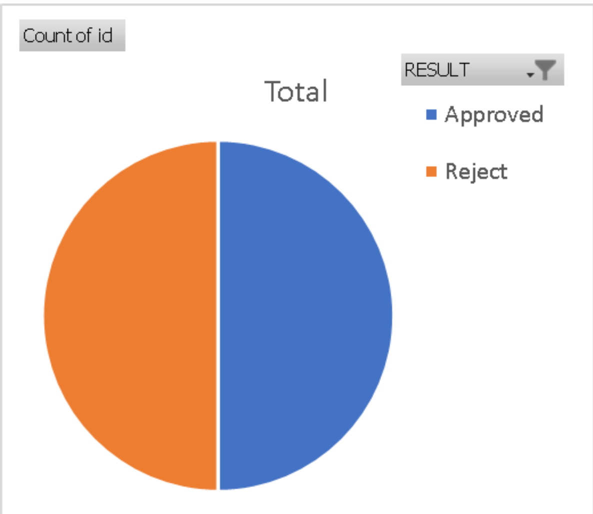

Figure 8 Example of unselected data from final dataset

3.4 Modelling

Feature Selection.

Feature selection is a crucial step in preprocessing, weighing variables towards the selected data label. It involves converting data into numeric using a convert operator in rapid miners.

Attributes are calculated using attribute selection techniques to improve model accuracy and accuracy. The threshold value is obtained from the highest Information Gain Value from each technique.

Modelling with Rapidminer.

RapidMiner used for generating the model with several classification methods for AI machine learning modelling, as shown on the Figure 9 below

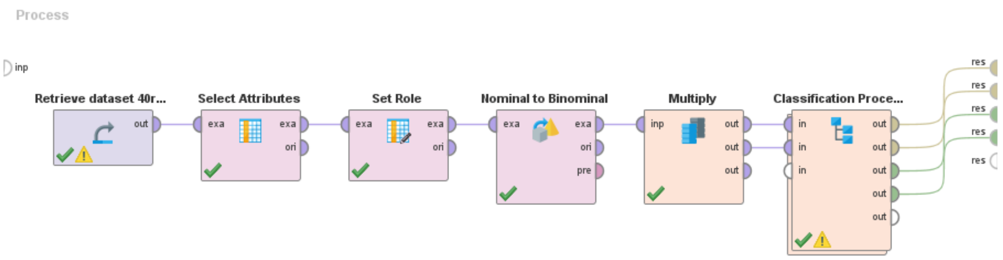

Figure 9Rapidminer pre-processing design

Figure10 shows the Rapidminer process flow in create process design. The data set was imported for process design, and to improve model performance, we used select attribute to eliminate unnecessary variables, set role for prediction label, and used nominal to binomial operator to convert attribute characters.

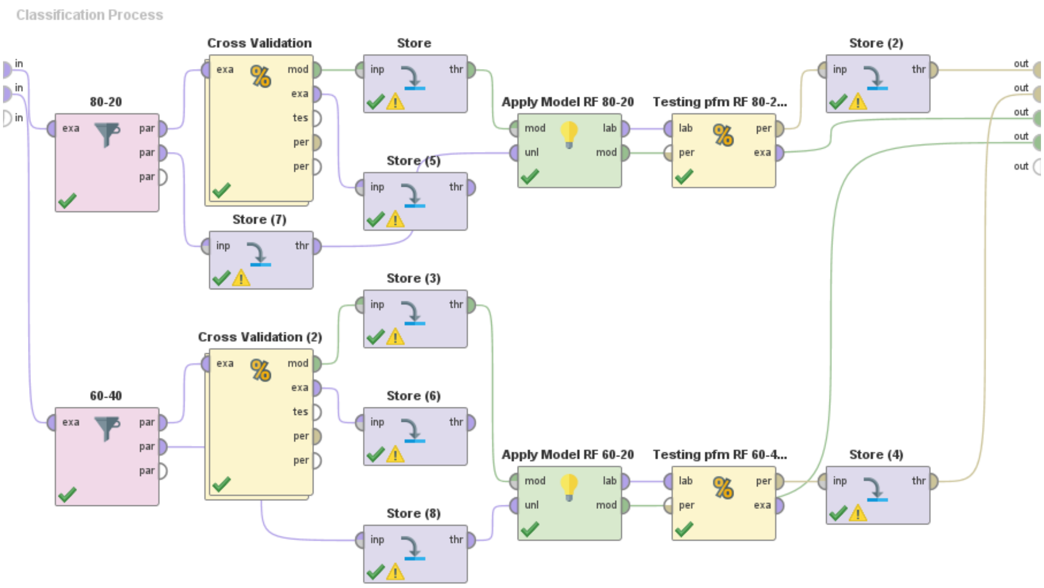

Figure 10 Process design modeling

After data transformation and cleaning, two data inputs were divided into two different modeling scenarios (80:20 and 60:40) using the same method. The original dataset was split into two ratios to ensure optimal separation of attributes. The larger ratio data set was used as the data train, while the smaller ratio (20&40 ratios) was used for testing the model. Data splitting was crucial for improving model performance and avoiding overfitting issues. The output from the split operator was sent to the Cross Validation operator to test the model's ability to predict output and identify problems like overfitting or selection bias.

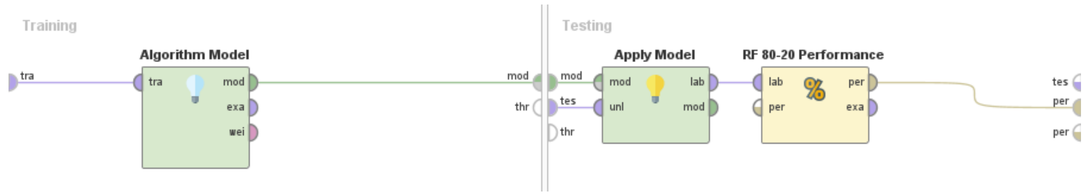

Figure 11 algorithm apply model and performance

The model's accuracy and sample number were determined using learning curves, with a goal of using 40,000 samples. Various predictive algorithms were used for data training and testing, including Naïve Bayes, random forest, logistic regression, decision tree, and deep learning. Data validation was used to determine the best credit scoring model. Multiple models were generated to compare performance results, and the model's performance was tested. The confusion matrix was created to predict true and false values. The model's accuracy and sample number were not linear.

RESULT AND DISCUSSION

4.1 Evaluation

After model is constructed, an indicator is requiredto show the model performance. In this research, performance indicators that are derived from the confusion matrix will be used. those are accuracy, Recall, Precision, Specificity and F-Measure. After the model performance generated, we could adjust the parameter data for the best performance from the model.

In this research we are generating 5 model using 5 different algorithm, those are Decision Tree, Deep Learning, Logistic Regression, Naïve bayes, and Random Forest. After we run the process for those algorithm we have the performance and confusion matrix for each algorithm as shown below:

Table 4 Confusion Matrix Decision Tree 60:40 (Performance Testing)

	true Approved	true Reject	class precision
pred. Approved	7978	1	99.99%
pred. Reject	22	7999	99.73%
class recall	99.72%	99.99%	

Table 5 Confusion Matrix Decision Tree 80:20 (Performance Testing)

	true Approved	true Reject	class precision
pred. Approved	3997	0	100.00%
pred. Reject	3	4000	99.93%
class recall	99.92%	100.00%	

Table 4 & 5 shows the example confusion matrix for on of the model (Decision Tree) using the test data in 2 scenarios. The prediction ability of the model is about 99 % to 100 %. From this Table, resultfrommodel performance metrics of 5 algorithm in 2 scenarios ratio can be calculated as shown in Table 7.

Table 6 Performance metrics on scenario 80:20

Model¥Criterion	Accuracy	Precision	Recall	Specificity	Class. Error	F-Measure
Decision Tree	99,96%	99,93%	100%	99,92%	0,04%	99,96%
Deep Learning	99,76%	99,90%	99,62%	99,90%	0,24%	99,76%
Logistic Regression	99,81%	99,85%	99,78%	99,85%	0,19%	99,81%
Naïve bayes	99,24%	98,98%	99,50%	98,98%	0,76%	99,24%
Random Forest	100%	100%	100%	100%	0,00%	100%

Table 7 Performance metrics on scenario 60:40

Model¥Criterion	Accuracy	Precision	Recall	Specificity	Class. Error	F-Measure
Decision Tree	99,86%	99,73%	99,99%	99,72%	0,14%	99,86%
Deep Learning	99,76%	99,85%	99,67%	99,85%	0,24%	99,76%
Logistic Regression	99,58%	99,67%	99,48%	99,67%	0,42%	99,57%
Naïve bayes	98,96%	98,53%	99,41%	98,51%	1,04%	98,97%
Random Forest	99,95%	100%	99,90%	100%	0,05%	99,95%

The Table shows the accuracy of the model almost reach 100% for all Algorithm model in both scenarios. This validates that the model is appropriate to be used as a tool to assess new Shariah financing application in MSME Sector.

Moreover, as shown in Figure 12 below, the AUC of Average 0,99 confirms the robustness of the model.

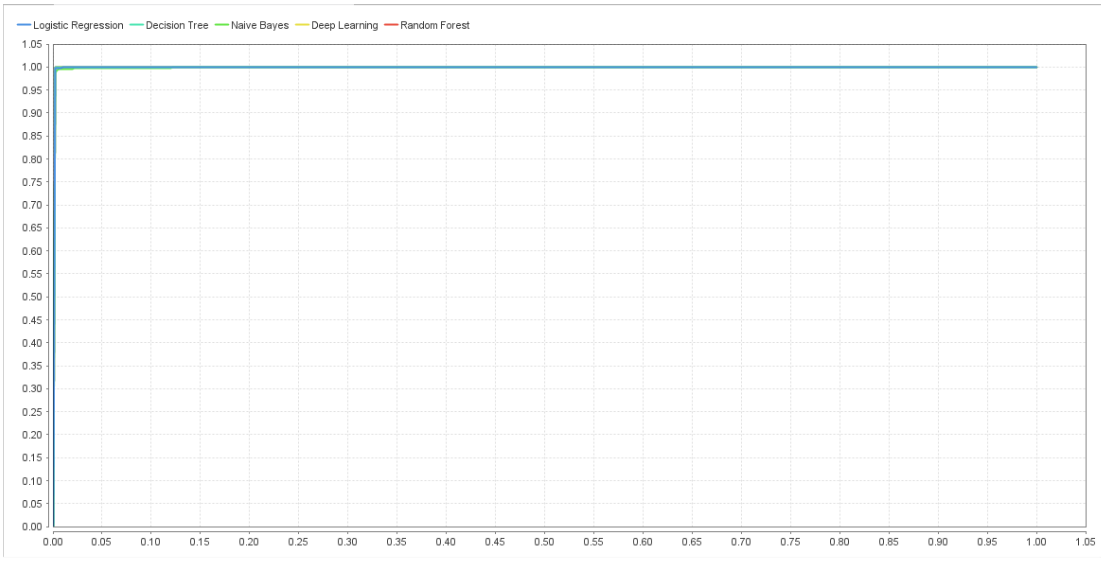

Figure 12ROC Comparison Graphics

CONCLUSION

The results of this research proofing that most classification Models could perform the training and testing process using the sharia financing dataset. In the training and testing process, all models have an accuracy average of 99% for the training testing process. In this research, the model gave the highest value on predicting rejected customers, as shown in Tables 4 & 5. From the 8000 reject labeled existing testing data, 7999 data were able to be recognized correctly and from 4000 reject labeled existing testing data, 4000 data were able to be recognized correctly. This also means that some data used for this research had strong characteristics and also had some variables that were too strongly influenced in decision making, so if we still want to make a model using this data, we need to separate variables that have too strong influence and add another variable has good relevancy but not too dominant for the decision. Including the strong influence of data on the decision will be bad for the development of the AI model because it will limit the learning process for future data.

REFERENCE

Ali, O., & Ishak, M. K. (2020). Bringing intelligence to IoT Edge: Machine Learning based Smart City Image Classification using Microsoft Azure IoT and Custom Vision. Journal of Physics: Conference Series, 1529(4), 0–10. https://doi.org/10.1088/1742-6596/1529/4/042076

Breiman, L., Friedman, J. H., Olhsen, R. A., & Stone, C. J. (1998). Classification and regression trees.

Chacko, A., Antonidoss, A., & Sebastain, A. (2020). Optimized algorithm for credit scoring. International Journal of Advanced Trends in Computer Science and Engineering, 9(1.3 Special Issue), 361–365. https://doi.org/10.30534/ijatcse/2020/5691.32020

Dastile, X., & Celik, T. (2021). Making Deep Learning-Based Predictions for Credit Scoring Explainable. IEEE Access, 9, 50426–50440. https://doi.org/10.1109/ACCESS.2021.3068854

Dastile, X., Celik, T., &Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing Journal, 91. https://doi.org/10.1016/j.asoc.2020.106263

Ferrari, S., & Stengel, R. F. (2005). Smooth function approximation using neural networks. IEEE Transactions on Neural Networks, 16(1), 24–38. https://doi.org/10.1109/TNN.2004.836233

Ha, V.-S., & Nguyen, H.-N. (n.d.). Credit scoring with a feature selection approach based deep learning. http://www.r-project.org

Hosen, M. N., Jie, F., Muhari, S., & Khairman, M. (2019). The Effect of Financial Ratios, Maqasid Sharia Index, and Index of Islamic Social Reporting to Profitability of Islamic Bank in Indonesia. Al-Iqtishad: JurnalIlmu Ekonomi Syariah, 11(2). https://doi.org/10.15408/aiq.v11i2.11588

Itoo, F., Meenakshi, & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. International Journal of Information Technology (Singapore), 13(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y

JunediHutagaol, B., &Mauritsius, T. (2020). Risk level prediction of life insurance applicant using machine learning. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 2213–2220. https://doi.org/10.30534/ijatcse/2020/199922020

Kam Ho, T. (n.d.). Random Decision Forests.

Kennedy, K., Namee, B. mac, & Delany, S. J. (2013). Using semi-supervised classifiers for credit scoring. Journal of the Operational Research Society, 64(4), 513–529. https://doi.org/10.1057/jors.2011.30

Khan, W. A., Chung, S. H., Awan, M. U., & Wen, X. (2020). Machine learning facilitated business intelligence (Part I): Neural networks learning algorithms and applications. Industrial Management and Data Systems, 120(1), 164–195. https://doi.org/10.1108/IMDS-07-2019-0361

Kleinberg, E. M. (1990). STOCHASTIC DISCRIMINATION. In Annals of Mathematics and Artificial Intelligence (Vol. 1).

Li, Y., & Chen, W. (2020). A comparative performance assessment of ensemble learning for credit scoring. Mathematics, 8(10), 1–19. https://doi.org/10.3390/math8101756

Lorena, A. C., & de Carvalho, A. C. P. L. F. (2007). Protein cellular localization prediction with Support Vector Machines and Decision Trees. Computers in Biology and Medicine, 37(2), 115–125. https://doi.org/10.1016/j.compbiomed.2006.01.003

Machine Learning by Thomas Mitchell (z-lib.org). (n.d.).

Maldonado, S., Peters, G., & Weber, R. (2020). Credit scoring using three-way decisions with probabilistic rough sets. Information Sciences, 507, 700–714. https://doi.org/10.1016/j.ins.2018.08.001

Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. Expert Systems with Applications, 38(12), 15273–15285. https://doi.org/10.1016/j.eswa.2011.06.028

Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. In Advanced Data Mining Techniques. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76917-0

Onay, C., & Öztürk, E. (2018). A review of credit scoring research in the age of Big Data. Journal of Financial Regulation and Compliance, 26(3), 382–405. https://doi.org/10.1108/JFRC-06-2017-0054

Tsai, C. F., & Hung, C. (2014). Modeling credit scoring using neural network ensembles. Kybernetes, 43(7), 1114–1123. https://doi.org/10.1108/K-01-2014-0016

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017, 280–289.

Xhemali, D., Hinde, C. J., & Stone, R. G. (2009). Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. IJCSI International Journal of Computer Science Issues, 4(1).