Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# IMPLEMENTATION OF MACHINE LEARNING TECHNOLOGY FOR CONSUMER CREDIT SCORING IN BANKING INDUSTRY: STUDY CASE OF PT BANK BNI SYARIAH

**FIRNANDO MUSLIMIN[1], TUGA MAURITSIUS[2]**

[1]Information System Management Department, Binus Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia 11480

[1]Information System Management Department, Binus Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia 11480

E-mail : [1]firnando.muslimin@binus.ac.id, [2]tmauritsus@binus.edu

**ABSTRACT**

Banking industry mainly runs their business on financing business. This business type currently still plays a role as a core business of PT Bank BNI Syariah among other business models in the company. Financing or credit business is not only provided by bank where we know that non-bank organization is also capable to provide similar services to customer which called as Financial Technology (Fintech). Fintech delivers its service to end-user through a portable application that can be accessed by end-user anytime and anywhere. Various automation is implemented in order to give excellent service level agreement (SLA) towards the product. Another high technology is implemented to obtain a very fast decision-making process for each loan request is powered by Artificial Intelligence (AI) technology. This technology is built on top of machine learning where loan requests can be determined just less than 10 minutes. The same service is mostly performed manually by a bank, where at this point there are a lot of manual processes that should be handled by a human. Physical interaction is needed to verify the customer as a set of activities of due diligence. By this condition, bank should be able to catch up to keep up with the direct or indirect competitors by the implementation of machine learning to perform credit approval.

**Keywords**: *Credit Scoring, Fintech, Classification Algorithm, Machine Learning, Consumer Banking*

## 1. INTRODUCTION

Economic growth in a country is very important. The development of the financial sector is expected to bring positive changes to the national economy. This is something that must be considered because the rotation of the wheels of the economy must continue under government supervision. In terms of realizing this, the banking sector acts as an intermediary between creditors and debtors which describes the ratio of the number of loans extended to third parties (LDR / Loan to Deposit Ratio).

The presence of technological innovations participates in businesses running in the financial industry, various types of financial services are developed in an easier, more compact form, and a much better user experience. This is made possible by the implementation of technology 4.0 in the financial industry. Some examples of companies engaged in the financial technology industry (fintech) can channel financing or credit to the public without the public having to go to the office or meet with the financial analyst officer. Fintech takes advantage of building web and mobile-based platforms to make it easier for customers to interact with the services provided by the fintech. Most fintech have used Machine Learning and Artificial Intelligent technology in providing credit decisions to the public. The system, which certainly involves the role of Big Data, provides a more accurate decision space for fintech to customers who apply for loans, the application of these 4.0 technologies makes fintech enough to divert people's attention to using fintech instead of banking due to Service Level Agreement of loan application until Disbursement only takes less than 1 hour.

Based on data submitted by the Indonesian Fintech Association (AFI), fintech has contributed to an increase in GDP of IDR 25.97 trillion and an

increase in household consumption of IDR 8.94 trillion in 2018. Two main factors are driving the evolution of financial technology innovation: the power of demand (demand side) and the power of supply (supply-side). Factors originating from the demand side include: First, changes in consumer preferences affect the consumer's need for innovation. The easy Internet access and real-time transaction capabilities of network users have set high expectations, especially in terms of convenience, speed, cost reduction and ease of use of financial services. main. This is very different from the services provided by banks where people must come and meet with bank officers to be able to apply for credit/financing plus the processing time is quite long. In addition, changes in preferences also occur due to the influence of demographic factors that drive demand, such as the growing acceptance of groups that have grown up with digital technology (digital natives) and millennials. Second, change technology. Technological innovation in financial services is evolving rapidly and in new ways and using different business models. With new business models and technology applications, new players can emerge in the financial services sector.
The process of granting credit/financing at Bank BNI Syariah is currently still carried out manually starting from the process of data collection, analysis and providing financing decisions. the process of providing financing which is carried out manually takes quite a long time, 1. the process of collecting data carried out by sales officers to customers requires at least 3 days; 2. after that the bank officer requires the customer's financial data which is also done manually and the customer must collect the data following the criteria required by the bank, 3. the appraisal team will conduct a guaranteed assessment which will take at least 3-4 days; 4. the analysis process is carried out manually by analyst officers and will verify data and site visits; 5. the financing breaker will give a financing decision referring to the level of authority to decide on each branch office leader, if the branch leader does not have the authority on the financing limit, it will be raised to the regional leader level and so on. This research will focus on the financing business that is channeled to customers with the consumer retail (mortgage) segment. This research will be executed with CRISP-DM methodology, where there will be six steps they are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment.

## 2. LITERATURE REVIEW

### 2.1 Consumer Credit / Financing Business

The consumer segment is a very attractive segment for the Bank due to the many business variants that can be a target for banks in distributing financing to customers. Both Islamic banks and conventional banks are increasingly focusing on the consumer retail business by creating various types of banking products that can meet the needs of retail customers. Some of the financing products offered by banks to consumer customers are as follows:

Financing Products

a. Homeownership financing/credit
b. Vehicle ownership financing/credit
c. Multipurpose financing/credit
d. Gold financing/credit
e. Home renovation financing/credit
f. Financing/credit for the purchase of plots of land
g. Financing/credit of tuition fees

The above products are accompanied by transactional features that vary based on the type of product used, provided by the bank. Based on Mckinsey projection data, the following is the consumer business climate in Indonesia.

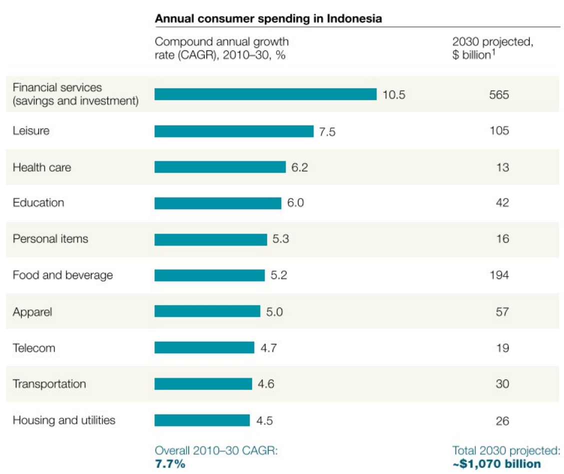


*Figure 1 Annual consumer spending in Indonesia [1]*

From the data above, most retail customers allocate their funds to have savings and investments. This can have a positive impact on the banking industry as a financial industry that can facilitate customer needs in placing funds and arranging their investments, especially in the form of fixed assets. Customers come to the bank to be able to have credit/financing facilities in the hope that the bank can assist customers in achieving their investment targets without having to wait for the customer's funds to be sufficient to purchase cash assets at once. The increasing growth of financing/credit makes banks must be able to provide convenience for customers to be able to access bank products and get services that are expected by customers. Process speed is very important for customers, considering the competition between banks is quite tight regarding the services provided to customers. In addition to competition between banks, the emergence of fintech (financial technology) has impact on banks. Some fintech mostly offer unsecured consumer lending where customers can receive loan facilities with a limit of up to IDR 2 billion. Fintech has a fairly with high transparency value so that public trust in fintech lending increases.

Seeing the above phenomenon, banks must be able to provide competitive services compared to fintech. Fintech can provide financing decisions in just a matter of minutes so that it significantly gives an attractive impression to customers even though the pricing offered is much higher than that offered by banks.

### 2.2 Credit / Financing Risk Exposure

Credit risk according to Bank Indonesia is the risk arising from the failure of the debtor or other party to fulfill their obligations to the Bank. This risk is a risk that is directly exposed to the financing business in various segments. Customer failure to pay can harm the bank, this can cause losses to the bank which must provide a higher Allowance for Impairment Loss (CKPN) as determined by the regulator. CKPN will take the bank's financial portion on the balance sheet so that it cannot be allocated as funds that can be used for banking business expansion.

### 2.3 Machine Learning in Credit Scoring

Machine Learning (ML) is a mathematical model used to improve performance for certain tasks. ML will generate a trained data model which is usually called "Training Data". It will make predictions and decisions. A set of data will be trained through ML and will act as a human brain that knows to recognize something from the input. ML is currently increasingly advanced, for example in the use of image classification. In this study, an artificial neural network is used for the ML algorithm. ML technology drives business automation in various areas, such as the calculation of cargo transportation into selected loan applications without human intervention, credit approval process, behavioral predictive analysis in financing activity and so on. This kind of technology is very promising because it can provide much more cost-effective than labor. On the other hand, ML can also be a problem due to the fall of automated trading platforms in the US stock market [2], and ML's misunderstanding of admitting a road position on a self-driving car that caused one pedestrian to die in the US. One of the most popular methods in Machine Learning is Feedforward Neural Network (FNN), FNN can solve problems in its universal approximation ability [3]. The ability of the FNN algorithm has succeeded in overcoming real-world problems in management, engineering and health science problems showing that the ability of this algorithm provides advantages in improving decision making for practical operations [4]. Other machine learning techniques are also used to achieve the best credit scoring performance such as Deep Learning (DL) that used by Zhang, Niu and Liu 2020 [5] on the research of using Deep Learning in Peer to Peer (P2P) business. This method is used to determine the appropriate borrower to be funded by lenders. On the other research K-Nearest Neighbor (KNN) is also one of the techniques used to handle credit scoring cases. In the Li research in 2009, he use KNN on the process of attribute selection and combined it with Linear discriminate analysis (LDA) and Decision Tree (DT) [6].

The popularity of using machine learning techniques in a financial institution is always developing to overcome their problem and fulfill the necessity of an effective and fast business process. Logistic Regression (LR) is widely used in health research, to assist the health industry able to identifying and generating diagnostic and prognostic [7]. On the other hand, LR is used for credit scoring cases by Bolton in 2009 on his research of the LR application for credit scoring [8]. As the growing the business needs and competition, a lot of research are being conducted to fit their business on the market competition through technology.

### 2.4 Classification Algorithm

Some algorithms that are often used in classifying data are SVM, Decision Tree, Naïve Bayes, k-NN and Neural Networks.

1. Decision Tree

Decision Tree is an approach that can solve problems in determining a decision, especially in multi-stage decision making [9]. If described, the Decision tree is a diagram that can help to choose one of several choices of actions or decisions. Generally, a decision tree starts with a single node

Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

or nodes. Then, the node branches to represent the available options. Furthermore, each of these branches will have new branches. Therefore, this method is called 'tree' because its shape resembles a tree with many branches. In a decision tree, the sequence and arrangement of rules can be carried out on various choices and investigate the possible outcomes of these choices. In addition, the possible risks and advantages of each available option can be seen from the shape of the tree. According to (venngage.com) Decision Trees usually consist of the following components:

- root node (root): the goal or major decision to be taken
- branches (twigs): various action options
- leaf node (leaf): possible outcomes for each action

Usually, there are two types of leaf nodes, which are square and circular. The square leaf nodes represent the decisions taken. Meanwhile, the circular leaf node represents an uncertain result. Decision tree is a commonly used method for making informal or simple decisions. However, according to Lucidchart, not a few also use it to predict results systematically. One example is in data analysis.

The Decision Tree model is a hierarchical model formed from the rules, discriminant functions that are applied through attributes/features in a feature space that exists in the dataset [10]. This model was developed by recursively partitioning the data on the feature space dataset. This is done to find the most optimal decision rules to be used in the model [10].

2. Naïve Bayes

The naive Bayes classifier is a classification method based on Bayes' theorem. The classification method proposed by the British scientist Thomas Bayes using probabilistic and statistical methods predicts future odds based on previous experience and is known as Bayes' theorem. The main feature of this Nave Bayes Classifier is a very strong (nave) assumption of the independence of each condition/event.

Olson & Delen (2008) Describe the naive Bayes of each decision class and calculate the probabilities given the object information vector, provided that the decision class is true. This algorithm assumes that the attributes of the object are independent. The probability of producing the final estimate is calculated as the sum of the frequencies from the "master" decision table. [11].

Naive Bayes Classifier works very well compared to other classifier models. This is evidenced by Xhemali and Hinde Stone in his journal "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages" says that "Naïve Bayes Classifier has a better accuracy rate than other classifier models". Naive Bayes has the advantage that this method can be used for both quantitative and qualitative data. we don't need a lot of data to train the model. [12].

3. K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is a method of classifying datasets based on learning from previously classified data. Included in supervised learning, the results of new query instances are categorized into ANNs based on most of the categorical proximity.

An example of a simple case study to describe k-NN is for example taking a decision (class) between attending or not attending a place. To support this decision-making, there is most of the decisions of friends or the environment (other instances). The friends are selected based on their proximity to the entity that is considering a decision. The measure of the closeness of this relationship can vary among neighbors, a hobby, a class, or other things. These measures can also be used together, for example, A is a neighbor, a hobby, and a class, while B is only one class and so on. The following figure shows the visualization of the KNN concept.

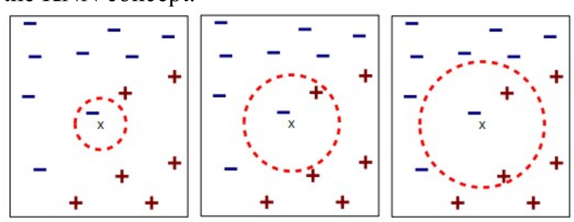

*Figure 2 KNN Concept* [13]

Near or far neighbors are usually calculated based on Euclidean distances, or other distance formulas can also be used. The close distance can be thought of as the inverse of distance, i.e. inversely proportional to distance. The smaller the distance between two instances, the greater the "closeness" between the two instances. Thus, the k nearest neighbors of an instance of x are defined as k instances that have the smallest distance (nearest, nearest) from x. In general, the steps taken to use k-NN are as follows:

a. Specifies the parameter k (number of closest neighbors).
b. Calculate the square of the object's Euclidean distance to the given training data.

c. Sort the results of no. 2 in ascending order (sequentially from high to low value)
d. Collect Y category (Nearest neighbor classification based on k-value)
e. By using the most majority nearest neighbor category, it can be predicted the object category

In KNN, there are four popular algorithms used by the researcher. Frist is a basic KNN algorithm that n-dimensional space $R^n$ becomes the correspondent point by all instances. The nearest neighbors of the instance are typically defined in terms of standard Euclidean distance. The second is Distance-Weighted Nearest Neighbor, this algorithm is the refined version of the basic nearest neighbor. The difference is this algorithm gives weight contribution on each k neighbor to the distance. At a closer neighbor, the weight will be much higher. The third is CHDM, HEOM and HVDM distance functions, this algorithm is useful when we know that Euclidean distance does not handle the qualitative attributes naturally. When the dataset contains qualitative and quantitative attributes, a heterogeneous distance function is required to handle it (HEOM). On the other hand, the overlap metric generated by HEOM fail to use the additional information in the dataset for qualitative attributes, an approach of Value Difference Metric (VDM) is then used to handle this condition. The dataset condition is matters whether to use what algorithm in this case, another algorithm Heterogeneous Distance Function (HVDM) also can be used which actually similar to HEOM, but this algorithm use VDM instead of an overlap metric for qualitative attributes.

## 3. RELATED RESEARCH

**Decision Tree -** In a study written by Gang Wang, Jian Ma, Lihua Huang, Kaiquan Xu in his research entitled "Two credit scoring models based on dual strategy ensemble trees" using the Decision Tree (DT) model to perform credit scoring modeling. The data set used is a dataset from the UCI Machine Learning Repository region of Germany and Australia. In this study, the author wants to see the performance of DT in the classification process because in doing classification, DT usually has a lower level of accuracy compared to other algorithms. The business domain targeted by the researcher is financial institutions that provide credit or financing to consumers. Here are the details of the data set used

*Table 1 Dataset for the research*

| | Total cases | Good/bad cases | No. of attributes |
|---|---|---|---|
| Australian credit | 690 | 307/382 | 14 |
| German credit | 1000 | 700/300 | 20 |

From the results of testing using these data, the average accuracy of DT is 84.39% for the Australian Credit dataset and 72.1% for the German Credit dataset [14]. Here's a comparison with other methods.

*Table 2 DT Credit Scoring Results*

| Methods | Australian credit dataset | | | German credit dataset | | |
|---|---|---|---|---|---|---|
| | Average accuracy (%) | Type I error (%) | Type II error (%) | Average accuracy (%) | Type I error (%) | Type II error (%) |
| LRA (West, 2000) | 87.25 | 11.07 | 14.09 | 76.30 | 11.86 | 51.33 |
| LDA (West, 2000) | 85.96 | 7.82 | 19.06 | 72.60 | 27.71 | 26.67 |
| MLP (West, 2000) | 85.84 | 15.40 | 13.26 | 73.28 | 13.52 | 57.53 |
| RBFN (West, 2000) | 87.14 | 13.15 | 12.74 | 74.6 | 13.47 | 52.99 |
| DT | 84.39 | 18.00 | 13.70 | 72.10 | 17.06 | 53.20 |
| Bagging DT | 86.38 | 14.15 | 13.19 | 76.45 | 11.85 | 50.83 |
| Random Subspace DT | 86.93 | 18.18 | 8.97 | 76.12 | 6.44 | 64.60 |
| Random Forest | 86.89 | 13.75 | 12.60 | 77.05 | 9.52 | 54.28 |
| Rotation Forest | 86.55 | 13.17 | 13.69 | 77.00 | 9.39 | 54.78 |
| RS-Bagging DT | 88.17 | 19.44 | 7.52 | 78.36 | 5.98 | 58.56 |
| Bagging-RS DT | 88.01 | 15.6 | 9.00 | 78.52 | 7.19 | 55.44 |

**Naïve Bayes -** Naïve Bayes is one of the algorithms that is also often used in the classification process, in the research conducted by Radha Vedala and Bandaru Rakesh Kumar in their research entitled "An Application of Naive Bayes Classification for Credit Scoring in E-Lending Platform". In this study, the intended business domain is P2P lending, on the P2P lending platform the author wants to apply ML with the Naïve Bayes method. The dataset contains two types of data, namely hard information (customer ratings, customer finances, repayments, etc.) and soft information (information extracted from social media to determine customer behavior). In this case study, the prediction results that will be carried out are the number of
people who default and who can fulfill their obligations. Here are the results of the classification using Naïve Bayes

*Table 3 Classification result using Naïve Bayes [15]*

| Actual/Predicted | Paid | Defaulted |
|---|---|---|
| Paid | 646 | 129 |
| Defaulted | 11 | 214 |

From the classification results, obtained an accuracy value of 86% and a sensitivity of 95.11% [15]. a good classification produces a high value of accuracy and sensitivity. In this study, sensitivity is defined as the portion of the prediction results of defaulters as defaulters.

**KNN** - KNN or K-Nearest Neighbor is a classification method that is often used for various case studies and one of them is credit scoring. Feng-Chia Li in his research "The Hybrid Credit Scoring Model Based on KNN Classifier" uses KNN as a classification algorithm with datasets from the UCI Machine Learning Repository region of Germany and Australia. The basic business domain is the consumer credit industry business. In this study, the authors combine the KNN method with several other methods such as DT, RoughSet, Linear Discriminant Analysis (LDA) and F-score with the main method base using KNN. From the results of the model evaluation, the average accuracy value is 88.64% and for the original KNN method itself, the accuracy value is 89.10% on the Australian dataset.

*Table 4 KNN results with Australian dataset*

| Combined approaches | Features selected | Accuracy rate Avg. (%) | Accuracy rate Std. (%) |
|---|---|---|---|
| Original+KNN | 14 | 89.10 | 11.98 |
| LDA+ KNN | 7 | 88.27 | 2.88 |
| DT+ KNN | 7 | 88.41 | 3.61 |
| RST + KNN | 7 | 88.54 | 6.34 |
| F-score+KNN | 7 | 90.40 | 10.36 |

In the German dataset, the original KNN accuracy value is 72.2% and the average accuracy with the overall combination method is 73.04% [6]

*Table 5 KNN result with German dataset*

| Combined approaches | Features selected | Accuracy rate Avg. (%) | Accuracy rate Std. (%) |
|---|---|---|---|
| Original+ KNN | 24 | 72.20 | 7.87 |
| LDA+ KNN | 12 | 74.50 | 5.48 |
| DT+ KNN | 12 | 71.90 | 5.22 |
| RST + KNN | 12 | 72.30 | 7.02 |
| F-score + KNN | 12 | 74.30 | 7.26 |

**Neural Network** - Neural Network (NN) has been widely used in various classification cases, in this case, one of the case studies that use it a lot is the financial services industry. In carrying out the analysis process that results in decisions, the financial services industry uses this algorithm to be able to provide accurate decisions on lending/financing to customers. As research conducted by the Department of Decision Science, East Carolina University that they use NN due to NN's ability to perform non-linear pattern recognition [16]. Desai et al. [17] in their research obtained an average accuracy of 83.56% for the Quinlan Credit Card database case study with 10 repetitions and a single data partition.

Tsai Chih-Fong and Hung Chihli [18] researched by implementing ANN in a credit card scoring case study using a dataset of Australian, German and Japanese credit card users and got the following results

*Table 6 Prediction accuracy with Neural Network*

| Data set | Single NN | NN Ensembles (voting) | NN Ensembles (weighted voting) | Hybrid NN |
|---|---|---|---|---|
| Australian | 0.8944 | 0.9017 | 0.9075 (0.8694) | 0.9161 (0.9014) |
| German | 0.8537 | 0.8711 | 0.8734 (0.7577) | 0.8745 (0.8283) |
| Japanese | 0.8708 | 0.8766 | 0.8797 (0.8647) | 0.8717 (0.8617) |
| Avg. | 0.8730 | 0.8831 | 0.8869 (0.8306) | 0.8874 (0.8638) |

From the results of the research above, it can be seen that NN has a fairly high level of accuracy with an average accuracy performance of above 85%, both single NN to hybrid NN. Referring to the results of the study, NN is one of the algorithms that has excellent accuracy performance to be implemented in consumer credit scoring research case studies.

## 4. LITERATURE DISCUSSION

Credit scoring has already been widely used by financial service company to leverage their business capable of handling large capacity of business process. The previous researcher Gang Wang, Jian Ma, Lihua Huang, Kaiquan Xu use Decision Tree with 86% of accuracy, while Radha Vedala and Bandaru Rakesh Kumar obtain 89,10% of accuracy on Australian dataset, Feng-Chia Li use KNN as the classifier algorithm and obtain accrucay score around 71,90% - 74.5% for german dataset, 88.27% - 90.40% accuracy on Australian dataset. Most of algorithm perform well, as proven by the high accuracy of every algorithm. In this case we also can see that different dataset can give different results with the same algorithm. By then these algorithms will be tested on authors dataset to see the best results for implementation.

## 5. RESEARCH METHODOLOGY

This research uses CRISP-DM as a methodology to conduct this research. CRISP-DM was evaluated as the methodology that fit the case that was being conducted by the author based on a survey performed by Daderman, et al 2018.

Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 7 Study cases with various methodology [19]*

| Studies | Industry | Framework | Application |
|---|---|---|---|
| **H.J Gómez Palacios et al. [9]** | Information technology | CRISP-DM<br>SEMMA | Studies of land use and cover change |
| **N. Caetano et al. [3]** | Medicine | CRISP-DM | Prediction of length of hospital stay |
| **R. Wirth, J. Hipp [26]** | Software | CRISP-DM | Response modeling. Improve the efficiency and effectiveness of mailing actions in marketing. It allows to increase the response rate while decreasing the costs of a campaign. Predict the likelihood of potential customers to reply to mailings. |
| **P. Kalgotra R. Sharda [23]** | Medicine | CRISP-DM | Progression analysis of signals. Converting streams of records to be able to detect useful signals for analysis, health care is used as an example. |
| **C. Zhang et al. [5]** | Transportation | KDD | Decision Making of Railway Traffic Safety in China |
| **F. Rebón et al. [7]** | Tourism | KDD | An antifraud system for operations with credit cards |
| **SAS Institute [2]** | Customer care | SEMMA | Prediction of the behavior of customers. |

Based on the table above, CRISP-DM is the most used methodology for data mining. This is the baseline of this research use CRISP-DM as the methodology to conduct credit scoring research.

## 6. EXPERIMENT SCENARIOS

This research is conducted with more than one scenarios as shown by the figure below.

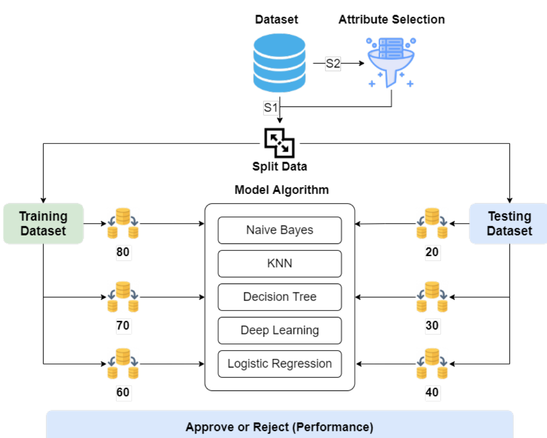


*Figure 3 Experiment Scenarios*

The figure above is the reference to how this research was conducted. There will be 5 algorithms that are used for building the model are Naïve Bayes, KNN, Decision Tree, Deep Learning and Logistic Regression. Each of the algorithms will generate three models based on data splitting scenarios 80-20, 70-30 and 60-40. Every generated model will be evaluated by each scenario.

## 7. RESEARCH STEPS

1. Business Understanding

The process of giving to customers is currently done completely manually, where when the customer has submitted an application and completes the data needed for analysis, there is a processing analyst who will analyze the prospective customer who will be given the financing facility. The following is a general illustration of the financing process.

- Customers apply for financing to sales
- Sales collect the necessary data
- Sales provide the necessary data to the processing analyst
- Processing analysts analyze data manually
- Processing analysts verify data for prospective customers
- Processing analysts make visits if needed
- The results of the analysis are submitted to the financing breaker
- Financing breaker provides financing decisions

In the analysis phase, the processing analyst has an important role to be able to determine that the prospective customer is eligible to receive financing facility. This process is carried out the same as what is done with the submission of existing customers at PT Bank BNI Syariah so that both new customers and existing customers will go through the same and quite a long process.

2. Data Understanding

This Credit Limit Automation System is intended for existing customers who have become customers of PT Bank BNI Syariah with a minimum period of 5 years. In the dataset, 30 attributes will be used to create the model. some highlights of the attribute contain information about the financing amount proposed by the customer, the maximum amount of installment that is afforded by the customer, ratio of given financing to the requested amount of financing, customer age, income, source of repayment, and financing purpose. The complete attributes are as follows

*Table 8 attributes on credit scoring dataset*

| | |
|---|---|
| Financing Amount | Sub Collateral Type |
| FTV | Facility Count |
| DSR | Land Certificate Status |
| Collateral Value | Company Type |
| Property Type | Join Income |
| Program | Building Area |
| Years of Services | Business Sector |
| Duration | Occupation |
| Income | Surface Area |

| | |
|---|---|
| Payroll | Facility Type |
| PKS | Facility Group |
| Education | Financing Purpose |
| Position | Installment |
| Age | Marital Status |
| Job Type | |

PT Bank BNI Syariah uses collateral as one of their financing analysis parameters, then inside the dataset the collateral value, sub collateral type, land certificate status, building area, and surface area are also included. On the side of repayment profile represented by position, job type, company type, and business sector attributes generate insight for bank to know the risk of a source of repayment would be sustainable until the due date of financing duration.

3. Data Preparation

48.000 datasets will be used in this research. The data used in this data will use existing data at PT Bank BNI Syariah. The data will be processed from raw data so that it can meet the needs of existing data to be further used in the process of modeling and training data. The tools used in the data preparation stage are Rapid Miner's TurboPrep. Concerning the data provided above, it is necessary to make the following preparations

a. Selecting data
From the total available data (48.000 data) then the data selection with adequate quality, in this case, is complete data on each attribute.

b. Cleaning data
For the data used, there are still errors and also a lack of data in some cells and at this stage the replacement of missing values will be carried out, adjusting the values that do not match the data format for each feature as it should.

c. Constructing data
At this stage, data aggregation or calculations will be carried out from several attributes that can produce information so that the data can be more relevant to use as a dataset and the construction of the data will generate a generated record.

d. Integrating data
The current dataset has been integrated into a single data query, so data integration is not required. However, if in the future other data sources will be linked to the existing data in this study, then the data integration process is very important and must be done.

e. Transforming data
In this phase, the data set can be reformatted in a form that is adapted to the tools used in the analysis process referring to each method or algorithm used. Some of the things that will be done in this stage are as follows:
- Assign role to "Approval" attribute as the label
- Convert label attribute to binominal
- Remap the binominal where "Approved" as positive class and "Reject" as negative class
- All attributes are then converted into numerical with "unique integer" on the parameter of coding type.

4. Modelling

The modeling process is the main basis for this research to get the performance results of each algorithm used. In this study, each model was built using the Cross Validation method to obtain optimal results. Each algorithm has its own configuration when the modeling process is carried out. The dataset used to build the model is divided into 3 scenarios, each scenario produces a model which will then be tested based on the test scenario of each data.

*Table 9 data partition scenarios*

| Scenario | Training | Testing | Total |
|---|---|---|---|
| Scene 80-20 | 38.400 | 9.600 | 48.000 |
| Scene 70-30 | 33.600 | 14.400 | 48.000 |
| Scene 60-40 | 28.800 | 19.200 | 48.000 |

Before all model generated from the dataset above, then attribute selection is applied to see if any attribute will be eliminated as the experiment scenario would like to compare the performance between with and without feature selection. The feature selection gives the following results.

*Table 10 feature selection results with WEKA*

| Feature Selection Algorithm | Attr. Count | Attribute |
|---|---|---|
| Cfs-SubsetEval | 4 | Collateral Value, Facility Count, Dsr, Ftv |
| ChiSquared-AttributeEval | 29 | All Default Attributes |
| Classifier-AttributeEval | 29 | All Default Attributes |
| Classifier-SubsetEval | 0 | None of Attribute is Selected |
| Correlation-AttributeEval | 29 | All Default Attributes |
| GainRatio-AttributeEval | 29 | All Default Attributes |
| InfoGainAttributeEval | 29 | All Default Attributes |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| Feature Selection Algorithm | Attr. Count | Attribute |
|---|---|---|
| OneR-AttributeEval | 29 | All Default Attributes |
| SymmetricalUncert-AttributeEval | 29 | All Default Attributes |
| Wrapper-SubsetEval | 0 | None of Attribute is Selected |

From the table above most feature selection algorithms prefer to use all default attributes as mentioned in table 8. CfsSubsetEval algorithm evaluates that Collateral Value, Facility Count, Dsr and Ftv as the attributes that have the biggest effect on the target label (Approval), while ClassifierSubsetEval and WrapperSubsetEva tend to evaluate that none of the attributes have a relationship to the target label. By the results of the feature selection process, authors will use the most dominant results from each feature selection algorithm which is 29 attributes (all default attributes). With this result, the scenario with and without feature selection will generate the same model. By then, both scenarios will be covered by the same modeling and evaluation process.
All 29 attributes will be used to generate each model and the modeling process will use the following configuration.

*Table 11 model configuration*

| Model | Parameter | Value |
|---|---|---|
| Naïve Bayes | Iaplace Correction | True |
| K-NN | K | 7 |
| | weighted vote | True |
| | measure types | MixedMeasures |
| | mixed measure | MixedEuclideanDistance |
| Decision Tree | criterion | gain_ratio |
| | maximal depth | 10 |
| | apply prunning | True |
| | confidence | 0,1 |
| | apply prepruning | true |
| | minimal gain | 0,02 |
| | minimal leaf size | 2 |
| | minimal size for split | 4 |
| | number of prepruning alternatives | 3 |
| Deep Learning | activation | Rectifier |
| | hidden layer | 50 : 50 |

| Model | Parameter | Value |
|---|---|---|
| | size | |
| | reproducible 1 thread | false |
| | epochs | 10 |
| | computable variable importances | False |
| | train samples per iteration | -2 |
| | adaptive rate | True |
| | epsilon | 1.0E-8 |
| | rho | 0.99 |
| | standardize | true |
| | L1 | 1.0E-5 |
| | L2 | 0 |
| | max w2 | 10 |
| | loss function | Automatic |
| | distribution function | Automatic |
| | early stopping | False |
| | missing values handling | Meanimputation |
| | max runtime seconds | 0 |
| | expert parameters | null |
| Logistic Regression | solver | Auto |
| | reproducible | False |
| | use regularization | False |
| | standardize | True |
| | non-negative coefficients | False |
| | add intercept | True |
| | compute p-values | True |
| | remove collinear collumns | true |
| | missing values handling | MeanImputation |

## 8. RESULT AND ANALYSIS

After the data is prepared and the specified model has been built, at this stage, an evaluation of the input data will be carried out to see the confidence level of the results of the machine learning analysis. To perform this measurement, the Accuracy, Precision, Recall and F-Measure indicators will be used. From the modeling process, the performance obtains the following results

1. Naïve Bayes (NB)

The following table are the results of Naïve Bayes training performance.

*Table 12 NB performance on training process*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 65,94% | 65,68% | 66,15% | 65,92% |

From the test results above, the third scenario with the composition of the training: testing 60:40 has a better accuracy value than other scenarios, which is 66,15%. Several things affect the accuracy from the dataset point of view, namely the presence of a local random seed variable when the data split process is carried out. But overall, the average performance accuracy of the three scenarios is 65,92%.

Respectively to the testing scenarios, each model was then tested using the dataset that was already split out of the dataset for model creation. The result of testing are as follows

*Table 13 NB models testing result*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 66,95% | 65,78% | 66,35% | 66,36% |
| Classification Error | 33,05% | 34,22% | 33,65% | 33,64% |
| Precision | 61,00% | 60,00% | 60,44% | 60,48% |
| Recall | 93,96% | 94,72% | 94,66% | 94,45% |
| F-Measure | 73,98% | 73,46% | 73,78% | 73,74% |

The results of the model test shown in the table above can be seen that the model that can provide the highest accuracy value is the Scene 80:20 model, which is 66.95%. From the test results above, the following is the resulting confusion matrix:

*Table 14 confusion matrix from NB test scenarios*

| NB 80-20 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 1917 | 290 |
| pred. Approved | 2883 | 4510 |

| NB 70-30 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 2653 | 380 |
| pred. Approved | 4547 | 6820 |

| NB 60-40 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 3653 | 513 |
| pred. Approved | 5947 | 9087 |

The precision value of each model in each scenario does not have a significant difference, resulting in an average precision of 60.48%. This illustrates that the level of precision of data that is correctly classified as "Approved" compared to other data that is incorrectly classified in the class "Approved". This illustrates that the precision of Naïve Bayes is quite low because of the many positive classes that are misclassified in the "Approved" class.

The results in the table above also show that the average recall in the three models is 94.45%. This shows that most of the correctly predicted data into the "Approved" class has been covered on average as much as 94.45%, while the average misclassification to the "Approved" class is in the negative "Reject" class. To describe the overall model performance representing recall and precision variables, the F-Measure measurement is used. In the results obtained in this model, the average F-Measure is 73.74% with the highest F-Measure value in the 80-20 model scenario.

2. KNN

After the model is built at the modeling stage, each model is then tested regarding each predetermined scenario. Each model produces a different performance on the modeling facet. At this stage, testing using test data is carried out to see the performance of each model against the dataset used to test the model.

*Table 15 KNN performance on training process*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 65,94% | 65,68% | 66,15% | 65,92% |

From the test results above, the third scenario with the composition of the training data: testing 60:40 has a better accuracy value than the accuracy in the other scenarios, which is 66.15%. Several things affect the accuracy from the dataset point of view, namely the presence of a local random seed variable when the data split process is carried out. But overall the average performance accuracy of the three scenarios is 65.92%. Then the model was tested using test data provided in experiment scenarios and resulting performance as follows.

*Table 16 KNN models testing result*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 95,20% | 95,18% | 95,24% | 95,21% |
| Classification Error | 4,80% | 4,82% | 4,76% | 4,79% |
| Precision | 93,12% | 93,07% | 93,26% | 93,15% |
| Recall | 97,60% | 97,62% | 97,53% | 97,58% |
| F-Measure | 95,31% | 95,30% | 95,35% | 95,32% |

Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 17 confusion matrix from KNN test scenarios*

| KNN 80-20 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 4454 | 115 |
| pred. Approved | 346 | 4685 |

| KNN 70-30 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 6677 | 171 |
| pred. Approved | 523 | 7029 |

| KNN 60-40 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 8923 | 237 |
| pred. Approved | 677 | 9363 |

From the table above, KNN provides excellent performance with an average accuracy of 95.21%. When compared from the three models, scenario 60-40 gives the highest accuracy, which is 95.24%, but has no significant difference to other models. Each class has an average precision level of 93.15%, which means the ability of KNN to distinguish data in positive (Approved) and negative (Reject) classes is very good. The recall rate on KNN also gives a very good output, with an average of 97.58%, this shows that of the overall data in the "Approved" class in the dataset, only 2.42% of the data in the positive class are not classified. Correctly. To see the performance that can describe the precision and recall of the KNN model, it is reflected in the F-Measure variable. The average F-Measure value in the three models is 95.32%, which means that the KNN model can be concluded to be able to provide optimal performance to be used in predicting the eligible customers.

3. Decision Tree (DT)

Modeling process on DT model resulting in the following performance on the training process.

*Table 18 training performance DT*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 93,81% | 93,35% | 93,81% | 93,66% |

Based on the table above, there is no significant effect on the test scenario carried out with a difference in the accuracy of less than 1%. The performance of the 80-20 and 60-40 scenarios produces performance with the same accuracy value of 93.81% while the 70-30 scenario produces 93.35% performance. From the three scenarios, the average accuracy value is 93.66%. the model is then tested using test data based on the scenarios, the test gives the following results.

*Table 19 DT models testing result*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 93,83% | 93,65% | 95,33% | 94,27% |
| Classification Error | 6,17% | 6,35% | 4,67% | 5,73% |
| Precision | 96,92% | 96,02% | 96,25% | 96,40% |
| Recall | 90,54% | 91,07% | 94,34% | 91,98% |
| F-Measure | 93,62% | 93,48% | 95,29% | 94,13% |

*Table 20 confusion matrix from DT test scenarios*

| DT 80-20 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 4662 | 454 |
| pred. Approved | 138 | 4346 |

| DT 70-30 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 6928 | 643 |
| pred. Approved | 272 | 6557 |

| DT 60-40 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 9247 | 543 |
| pred. Approved | 353 | 9057 |

From the table above, the highest accuracy is generated by the model in the 60-40 scenario with an accuracy value of 95.33%. The performance of each model does not have a large difference, which is less than 1%, from the three test scenarios the average model produces an accuracy value of 94.27%. This accuracy value shows that the Decision Tree's ability to classify each data is very good with a relatively small average error rate of 5.73%.

The Decision Tree model also produces a very high precision value, namely 96.92% in the 80-20 scenario with an average precision value of 96.40%. This shows that the Decision Tree's ability to make predictions in the positive class is very good when viewed from the data that is actually in the positive class (Approved). Besides that, the recall value was also calculated and got the highest results in the 60-40 scenario, which was 94.34% with an average of 91.98% for the three models. This shows that the Decision Tree can provide high performance in classifying data correctly. These two variables are then described in one variable,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

namely the F-Measure. The highest F-Measure value obtained is 95.29% in the 60-40 scenario.

4. Deep Learning (DL)

The modeling process of DL resulting the following performance on the training process.

*Table 21 training performance DL*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 91,85% | 91,24% | 89,60% | 90,90% |

From the table, the most optimal model is using the model with the first scenario, namely the model that uses the 80:20 training testing data partition. The model produces the highest accuracy of 91.85% and the lowest model in the third scenario is 60:40 with an accuracy of 89.60%. The average accuracy obtained from the three scenarios is 90.90%. the generated model on the modeling process is then tested with three scenarios as planned in this research. The three scenarios give the following result

*Table 22 DL models testing result*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 93,44% | 91,81% | 93,94% | 93,06% |
| Classification Error | 6,56% | 8,19% | 6,06% | 6,94% |
| Precision | 89,62% | 94,37% | 94,84% | 92,94% |
| Recall | 98,25% | 88,92% | 92,93% | 93,37% |
| F-Measure | 93,74% | 91,56% | 93,88% | 93,06% |

*Table 23 confusion matrix DL test scenarios*

| DL 80-20 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 4254 | 84 |
| pred. Approved | 546 | 4716 |

| DL 70-30 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 6818 | 798 |
| pred. Approved | 382 | 6402 |

| DL 60-40 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 9115 | 679 |
| pred. Approved | 485 | 8921 |

In the test results of the model above, the 60-40 scenario gives the highest accuracy value of 93.94% followed by the 80-20 and 70-30 scenarios. The performance of the three models shows optimal results, which are above 90% or the average accuracy in the three scenarios is 93.06%. When viewed from the precision values in the three models, although the 80-20 scenario provides a higher accuracy value than the 70-30 scenario, the 70-30 scenario provides a better level of precision, which is 94.37%. This shows that the 70-30 scenario has a better ability to classify the true positive class (Approved) against data that is in the true positive class. While the 80-20 model has a better recall value when compared to the 70-30 model, which is 98.25%. This shows that the 80-20 model has a better ability to identify customers who are eligible for financing facilities. To describe precision and recall in one variable, the indicator used is the F-Measure. The F-Measure value of the three models shows that the 60-40 model has the highest performance, which is 93.88%. Both recall and precision are important to be used as reference parameters for model performance, therefore F-Measure is used to see the balance of the model.

5. Logistic Regression (LR)

LR models are built based on experiment scenarios where each model represents the number of training data. These processes generate training performance as follows

*Table 24 training performance LR*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 95,34% | 95,50% | 95,53% | 95,46% |

The table above shows that the training performance of the model with the training:testing 60:40 data partition scenario has the highest accuracy value of 95.53%. While the accuracy in other models is lower but only with a difference of not more than 1%. Overall the three models above provide optimal performance with a fairly thin accuracy difference with an average accuracy of 95.46%. Three of the model then tested based on scenarios. Each scenario generate results like the following table.

*Table 25 LR models testing result*

| Measurement Variable | Scenario | | | Average |
|---|---|---|---|---|
| | 80-20 | 70-30 | 60-40 | |
| Accuracy | 95,39% | 95,22% | 95,29% | 95,30% |
| Classification Error | 4,61% | 4,78% | 4,71% | 4,70% |
| Precision | 95,28% | 95,35% | 95,43% | 95,35% |
| Recall | 95,50% | 95,08% | 95,12% | 95,23% |
| F-Measure | 95,39% | 95,22% | 95,28% | 95,30% |

*Table 26 confusion matrix LR test scenarios*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| LR 80-20 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 4573 | 216 |
| pred. Approved | 227 | 4584 |

| LR 70-30 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 6866 | 354 |
| pred. Approved | 334 | 6846 |

| LR 60-40 | true Reject | true Approved |
|---|---|---|
| pred. Reject | 9163 | 468 |
| pred. Approved | 437 | 9132 |

The performance of each LR model gives quite competitive results compared to the models produced by other algorithms. The average accuracy of the three models is 95.30% with the highest accuracy obtained from the 80-20 model. However, from the three models above, there is no significant difference in accuracy because each model produces an accuracy above 95%. On the other hand, the highest level of precision is obtained by the 60-40 model followed by the 70-30 and 80-20 models. The precision indicator explains that the model's ability to classify the true positive class (Approved) against all data that is actually in the positive class. Each model shows a performance that tends to be similar because the average precision value of the three models above is 95.35%. Besides that, the recall indicator also shows a very good value, with an average of 95.23%. This shows that each model can correctly classify 95.23% positive class. The performance of each model is then described as a whole through the F-Measure indicator, which is an indicator that can include precision and recall in one indicator. So that both precision and recall value can be represented by one variable. The highest F-Measure value was obtained from the 80-20 model with a value of 95.39%, however, there was no significant difference between the three models that were able to provide an average F-Measure value above 95%, namely 95.30%.

## 9. CONCLUSION

In this research, the authors can obtain that CRISP-DM is a research methodology that fits perfectly to study case of this research. The business understanding phase makes the further steps easier especially when data understanding is performed. After data is clearly understood, the data preparation will be more reasonable due to the understanding of data enabling authors to prepare data to meet the best structure without elimination the information in it. The modeling and evaluation process is performed after all preparation is set. There are five algorithms and three dataset partitions on the experimentation scenarios, the model was created on Naïve Bayes (NB), Deep Learning (DL), Decision Tree (DT), K-Nearest Neighbors (KNN), and Logistic Regression (LR) algorithm. Each algorithm generates three models based on the scenarios they are 80-20, 70-30 and 60-40 data partition. Each scenario performs with different results.

*Table 27 model performance summary*

| Algorithm | Average on Each Models | | | |
|---|---|---|---|---|
| | Accuracy | Precision | Recall | F-Measure |
| Deep Learning | 93,06% | 92,94% | 93,37% | 93,06% |
| Decision Tree | 94,27% | 96,40% | 91,98% | 94,13% |
| Naïve Bayes | 66,36% | 60,48% | 94,45% | 73,74% |
| K-Nearest Neighbor | 95,21% | 93,15% | 97,58% | 95,32% |
| Logistic Regression | 95,30% | 95,35% | 95,23% | 95,30% |

From the table above can be obtained that Logistic Regression stands out of all other models for accuracy. Logistic Regression performs with 95,30% accuracy, on the other hand, Naïve Bayes gives the lowest accuracy 66,36%. The accuracy differences between Logistic Regression and Naïve Bayes are quite big that is 28,94%. While other models from Deep Learning, Decision Tree and KNN perform with very close accuracy. Looking at the precision indicator, Decision Tree performs with 96,40% although it gives less accuracy than Logistic Regression. This means that the ratio of Decision Tree model to correctly classify the Approved customer towards all data that predicted into Approved class is very good. Another measured indicator is recall, where KNN obtain a higher recall than other models which is 97,58%. This indicator informs that KNN identifies the true positive class (Approved customer) better rather than any other model. In the case of banking business, either precision or recall is very important, where banks want to identify as much as customer that can be approved to expand the business, but bank also need to identify the correct customer precisely that can be given financing facility without worrying the customer won't default as long as the period is running. By then F-Measure covers precision and recall in one variable, to see both performances in a single variable. The results above show us that KNN model can perform better than any other model that its F-Measure reach 95,32%. While Naïve Bayes is still in the lowest position since it only gets 73,74% though it has a good recall at 94,45%.

As discussed in this paper, we can obviously see that financing act as the main fuel of a financial services company to their business with sustainability. In this case financing is the most contributing profit on overall profitability portfolio. The long financing process makes bank cannot expand their business with existing resources, instead they must add more resource to acquire more financing submission. As described in introduction section, overall financing process can take more than 14 days, which means that each resource capable of maximum 2 customer per month. Bank BNI Syariah need to add more resource to process more customer at the same time. It will force BNI Syariah to get more employee to increase the business capacity which will affect the overhead cost and there will be a systemic impact to the related working unit due to the increasing business load. This will be a complicated problem when bank need to remap their business goal and target, the growing business will be followed by high cost in it. On this research try to implement machine learning to overcome the time consuming of financing disbursement process. Compared to existing credit scoring process, machine learning will give efficiency more than 85%, bank will be able to boost their business acquisition with low cost. The use of attribute selection algorithm gives the same results as the original dataset attribute, which means that most of attribute selection algorithm consider all attribute should be used in the dataset for modeling process.

After all, can be concluded that the intervention of machine learning in the credit scoring study case is significantly affect the business process of financial service company in terms of financing business. The role of technology driving the business much faster without expelling the risk exposure that appear on its way. This is handled by every algorithm that perform with high accuracy as described on this section. Most of algorithm works perfectly and obtain F-Score above 90% except for Naïve Bayes. According to the author, Deep Learning, Decision Tree, KNN and Logistic Regression is reliable to be implemented for credit scoring in PT Bank BNI Syariah. These models can be utilized for a credit automation system which will affect the efficiency of PT Bank BNI Syariah to process a credit/financing submission.

## 10. DISCUSSION FOR FURTHER RESEARCH

This research try to achieve high acquisition with high cost efficiency by using machine learning algorithm to perform credit scoring. All financing proposal were treated as the same way, property financing proposal, multipurpose financing, vehicle purchase financing, and other financing product in Bank BNI Syariah. Authors create the model for all this financing process, where every financing type could have tricky part where analyst need to follow up more for certain points. At this stage Authors propose to the next researcher to enrich the dataset by financing type and category and create each model for each financing category to achieve more accurate and precise results.

## REFERENCES

[1] A. Budiman, H. Chhor, and R. Razdan, "Understanding the diversity of Indonesia's consumers," *McKinsey Q.*, no. 2, pp. 8–11, 2013.

[2] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," *Proc. 11th Int. Conf. Web Soc. Media, ICWSM 2017*, pp. 280–289, 2017.

[3] S. Ferrari and R. F. Stengel, "Smooth function approximation using neural networks," *IEEE Trans. Neural Networks*, vol. 16, no. 1, pp. 24–38, 2005, doi: 10.1109/TNN.2004.836233.

[4] W. A. Khan, S. H. Chung, M. U. Awan, and X. Wen, "Machine learning facilitated business intelligence (Part I): Neural networks learning algorithms and applications," *Ind. Manag. Data Syst.*, vol. 120, no. 1, pp. 164–195, 2019, doi: 10.1108/IMDS-07-2019-0361.

[5] Z. Zhang, K. Niu, and Y. Liu, "A deep learning based online credit scoring model for P2P lending," *IEEE Access*, vol. 8, pp. 177307–177317, 2020, doi: 10.1109/ACCESS.2020.3027337.

[6] F. C. Li, "The hybrid credit scoring model based on KNN classifier," *6th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2009*, vol. 1, pp. 330–334, 2009, doi: 10.1109/FSKD.2009.261.

[7] D. S. Courvoisier, C. Combescure, T. Agoritsas, A. Gayet-Ageron, and T. V. Perneger, "Performance of logistic regression modeling: Beyond the number of events per

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

variable, the role of data structure," *J. Clin. Epidemiol.*, vol. 64, no. 9, pp. 993–1000, 2011, doi: 10.1016/j.jclinepi.2010.11.012.

[8] C. Bolton, "Logistic regression and its application in credit scoring," p. 238, 2009.

[9] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 660–674, 1991, doi: 10.1109/21.97458.

[10] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *J. Chemom.*, vol. 18, no. 6, pp. 275–285, 2004, doi: 10.1002/cem.873.

[11] D. L. Olson and D. Delen, *Advanced data mining techniques*, no. January. 2008.

[12] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009, [Online]. Available: http://cogprints.org/6708/.

[13] R. Agrawal, "K-Nearest Neighbor for Uncertain Data," *Int. J. Comput. Appl.*, vol. 105, no. 11, pp. 13–16, 2014.

[14] G. Wang, J. Ma, L. Huang, and K. Xu, "Two credit scoring models based on dual strategy ensemble trees," *Knowledge-Based Syst.*, vol. 26, pp. 61–68, 2012, doi: 10.1016/j.knosys.2011.06.020.

[15] R. Vedala and B. R. Kumar, "An application of Naive Bayes classification for credit scoring in e-lending platform," *Proc. - 2012 Int. Conf. Data Sci. Eng. ICDSE 2012*, pp. 81–84, 2012, doi: 10.1109/ICDSE.2012.6282321.

[16] D. West, "Neural network credit scoring models," *Comput. Oper. Res.*, vol. 27, no. 11–12, pp. 1131–1152, 2000, doi: 10.1016/S0305-0548(99)00149-5.

[17] V. S. Desai, D. G. Conway, J. N. Crook, and G. A. Overstreet, "Credit-scoring models in the credit-union environment using neural networks and genetic algorithms," *IMA J. Manag. Math.*, vol. 8, no. 4, pp. 323–346, 1997, doi: 10.1093/imaman/8.4.323.

[18] C. F. Tsai and C. Hung, "Modeling credit scoring using neural network ensembles," *Kybernetes*, vol. 43, no. 7, pp. 1114–1123, 2014, doi: 10.1108/K-01-2014-0016.

[19] A. Daderman and S. Rosander, "Evaluating Frameworks for Implementing Machine Learning in Signal Processing," *Examensarbete Inom Tek.*, pp. 1–36, 2018.