Journal of Theoretical and Applied Information Technology
15th July 2018. Vol.96. No 13

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

ALGORITHM COMPARISON OF NAIVE BAYES CLASSIFIER AND PROBABILISTIC NEURAL NETWORK FOR WATER AREA CLASSIFICATION OF FISHING VESSEL IN INDONESIA

[1]MUSTAKIM, [2]ASSAD HIDAYAT, [3]ZULIAR EFENDI
[4]ASZANI, [5]RICE NOVITA, [6]EPLIA TRIWIRA LESTARI

[1,2,3,4,5]Puzzle Research Data Technology (Predatech), Faculty of Science and Technology
[1,2,3,4,5,6]Department of Information System, Faculty of Science and Technology
Universitas Islam Negeri Sultan Syarif Kasim Riau 28293, Pekanbaru, Riau, Indonesia
E-mail: mustakim@uin-suska.ac.id

ABSTRACT

Indonesia's maritime area is twice the size of its archipelago, with an area of 5.9 million km2. Based on the United Nations Convention on the Law of Sea (UNCLOS 1982). Indonesia is also the second largest fish producing country in the world with fish catch of 6 million tons in 2014 based on the latest data from the Food and Agriculture Organization (FAO). The fish catching process requires the role of vessels suited to the existing water conditions, one of which has robust resilience to the state of the Indonesia sea. Thus, it is necessary to study the classification of aquatic types on Indonesian fishing vessels to determine the impact that will occur on the vessel. This research performs classification process using Naïve Bayes Classifier and Probabilistic Neural Network (PNN) algorithm. Accuracy result got in Naïve Bayes Classifier algorithm using RapidMiner tool is equal to 48%. While for PNN algorithm, experiment with three different spread values yield an accuracy of 68% for spread value 0.1, 78% accuracy for spread value 0.01 and the last experiment is the value of spread of 0.001 produce 100% accuracy. Therefore, in this study it is known the classification using PNN algorithm is better than Naïve Bayes Classifier.

Keywords: *Accuracy, Perairan, Fishing Vessel, Classification, Naïve Bayes Classifier, Probabilistic Neural Network.*

1. INTRODUCTION

Indonesia has great natural potential from the sea for national development, not only that the sea has many benefits for human life and other living things, because in the sea the wealth of natural resources are very abundant. This shows that Indonesia is the largest Maritime nation in the world. The existence of the Indonesian sea becomes very familiar in the midst of society and the sea is also the source of food, rich in animal protein and as a source of income for fishermen [1]. Potential fishing can be utilized through responsible exploitation by means of fishing vessels and fishing gear. Fishing vessels are ships, or boats used for fishing to support fishing operations, and fisheries research/exploration [6]. There are several types of fishing vessels in Indonesia such as motor boats, boats without motors and outboard motors [1]. Each vessel has special needs in every existing water, such as in the sea, the ship used has resistance to the waves and big winds. In addition, the sea water content has a high degree of salinity so that the occurrence of rust on the body of the ship made of metal can cause the decreasing of ship speed, decreasing the resistance of the ship so that the longer the ship will be quickly damaged [9]. If this is ignored then it will be bad for the safety of fishermen, from these issues, the types of ships and types of waters that affect the safety of fishermen need to be noted. To know the type of waters that will be sailed, required a method of data mining such as classification of ship characteristics based on the type of marine waters in Indonesia.

Data mining is the process of finding unrelated data relationships o the users, and is presented in a way that is easily understood so that it can be a reference in decision making [7]. While Classification is a method used to define new data

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

records into categories which have been previously defined [8].

Research conducted by Rustiyan and Mustakim in 2016 staed that classification of type of waters at Indonesian fishing vessel using K-Nearest Neighbor method with microsoft excel and training data of 436 data and testing data of 50 data with valueof k = 1, k = 3, k = 5 and k = 7, the best accuracy value at k = 7 is 94%. Suggestions from the research is to conduct research with other classification algorithms to compare results and reproduce more knowledge in Machine Learning [1].

Two of the well-known algorithms often used in data mining classification are Backpropagation Neural Network and Naïve Bayesian Classifier (NBC) [9]. NBC is a statistical classification which can predict the probability of a class. NBC is based on the Bayes theorem which has similar classification capabilities to the Decision Tree and Neural Network. NBC proved to have a high accuracy and speed when applied into databases with large data [10]. Furthermore, Backpropagation algorithm is known to be very helpful in many applications because of its simplicity in the training process, but if applied to case examples that demand process speed, then it is less suitable since it will take a lot of time [11]. Probabilistic Neutral Network (PNN) solves the problem of prediction/ classification by considering each input in the input layer. The process undertaken by PNN can take place more quickly when compared with Backpropagation [12]. The advantage of the PNN algorithm is the ease with which the network is modified when used training data is added or reduced. Research using PNN has been widely used, one of them [21] on the diagnosis of internal faults of oil-induced power transformers based on PNN. In this study, comparisons were made using other algorithms such as backpropagation neural network, decision tree, bayes algorithm and fuzzy algorithm. Result of the research indicates that PNN get the highest accuracy position which reach 80%. Other research that is classification of data weight of infant using PNN and logistic regression, the accuracy of classification using spread value 0.2 to 1 yielded accuracy of 100% for training data while logistic regression method only reached 88.2% [12]. In this research, PNN method is better than logistic regression method. Therefore, in this research, developing the previous research to compare NBC and PNN Algorithm in classifying fishing vessels in Indonesian waters.

The novelty of this study compared to previous research is on the process of data distribution by K-Fold Cross Validation with the percentage of 70% and 30%. Previous experiment using random data, obtained a higher accuracy than the usual K-Fold Cross Validation technique. When further examined, both in terms of novelty and the application of comparison algorithms, the main contribution of this study is the refinement of previous studies which has small accuracy. Therefore, a K-Fold Cross Validation data-modification technique is required with the percentage distribution, which was attempted in this case. But do not rule out, the results of the experiment with a good accuracy, it all depends on the dataset used.

2. LITERATUR REVIEW

2.1. Data Mining

Data mining is a process of data analysis in different angles and the end result becomes a useful information. Technically, data mining is the process of finding correlations among the many fields in large datasets [2]. Knowledge Discovery in Data (KDD) is an activity that includes the collection, use of historical data to find regularities, patterns or relationships in large data sets [3].

In data mining, accuracy and error become major benchmarks in summarizing th results obtained from a series of experiments [33]. In addition to the accuracy of several classification algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Probabilistic Neural Network (PNN) and Support Vector Regression are more likely to use error approaches such as Means Squared Error (MSE) as a reference for the successful performance of the algorithm [34].

2.2. Classification

Classification is an act to look for images of class labels in such a way that can predict unknown class labels. Thus, the classification is usually used to predict unknown class labels. Classification implements several methods such as decision trees, Bayesian methods and induction rules. The classification process involves two steps. The first step is the learning step which involves the development of the model while the second step involves using the model to predict the class label [14]. The classification is identical to the training data and testing data, the most effective data divide using the clustering model [23].

2.3. Accuracy, Precision and Recall

Accuracy is calculated as the number of predicted events divided positively by the total number of

Journal of Theoretical and Applied Information Technology
15th July 2018. Vol.96. No 13

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

events. Means accuracy is an accurate percentage of the class predictions among the total class [13].

$$Accuracy = \frac{True\ Positive + True\ Negative}{P + N} * 100\% \quad (1)$$

Precision is a precision that is really classified the class, therefore known as a positive predictive value [14].

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} * 100\% \quad (2)$$

The most relevant level and part of a document [13].

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} * 100\% \quad (3)$$

2.4. Naïve Bayes Classifier

Naïve Bayes algorithm is one algorithm in the classification technique. The Naïve Bayes Method is a probability and statistic referred to the British scientist Thomas Bayes, who predicted future opportunities based on previous experience known as Bayes Theorem [3]. Naïve Bayes is a simple probabilistic class that calculates a set of probabilities by summing the frequency and value combinations of the got dataset [4]. The steps of the Naïve Bayes method are as follows [3]:

1. Read training data.
2. Calculate the number and the probability, if a data of a numeric type then:
   a. Find the mean and standard deviation of each parameter.
   b. Determine the probabilistic value by calculating the appropriate amount of data from the same category divided by the amount of data in the category.
3. Getting values in the standard deviation and probability tables [3].

Bayes's Theorem has the following general form of equation [5]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (4)$$

Note:
X = Data with unknown class
H = The hypothesis of X data is a class
P(H|X)= Probability of hypothesis H based on condition X (posteriori prob.)
P(H) = Probability of hypothesis H (prior prob.)
P(X|H)= Probability of X under these conditions
P(X) = Probability of X [5]

2.5. Artificial Neural Network

Artificial Neural Network (ANN) is a different approach from other artificial Intelligence science methods. ANN is a model of intelligence inspired from the human brain structure that is implemented using computer programs and computer applications, which are able to complete a number of calculations during the learning process [24] [26]. Neural network is a learning machine that is formed of a number of simple processing elements called neurons or nodes. Each neuron is connected to other neurons with direct communication links through the pattern of relationships called network architecture [25] [28]. Some types or groups of ANN are Perceptron, Backpropagation, Learning Vector Quantization (LVQ), Self-Organizing Map (SOM) and Probabilistic Neural Network (PNN) (SOURCE). Of the various types of the ANN algorithm that has the fastest level of accuracy is PNN [29].

2.6. Probabilistic Neural Netwok

The Probabilistic Neural Network (PNN) was first developed by Donald F. Specht in 1988 [15]. PNN is a method of artificial neural networks using the principle of statistical theory, Bayesian Classification to replace the heuristic principle used by the Backpropagation algorithm [16]. Probabilistic Neural Network is a method of a neural network that uses supervised training. Probabilistic artificial neural network belongs to the supervised learning because the expected outputs have been known and are models established by the estimation of probability functions [17]. The architecture of PNN comprises 4 layers, the input layer, pattern layer, summation layer and decision layer/output layer than shown below:

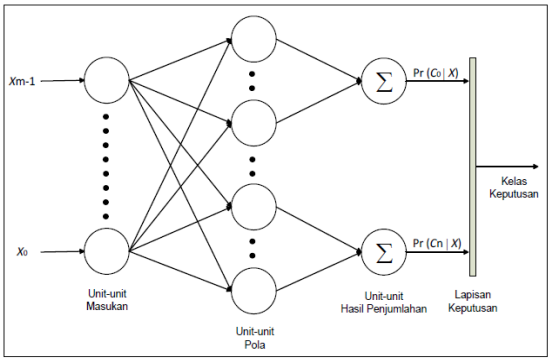

Figure 1. PNN Architecture [18].

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

2.7. K-Fold Cross Validation

K-Fold Cross Validation is done to divide the data into parts of training set and test set. The essence of this validation is to divide the data randomly into the subset of the required part irregularly. K-Fold Cross Validation repeats K as much as N to divide a set randomly into a free subset K, each loop leaving a subset for testing and another subset for training [30]. The use of cross validation attempted more practice data than test data.

The subset K is selected by one subset into test data and (k-1) is used as training data. However, in theory there is no definite benchmark for the value of K. The advantages of K-Fold Cross Validation compared with cross-validation variations such as Repeated random sub-sampling validation are all data used for test data and training data interchangeably [31]. Several studies related to data sharing are mostly done in this way.

2.8. National Fishing Vessel

Offshore fishing is a dangerous job in the world. The profession of seafaring fishing vessels has the characteristics of 3D work: dangerous, dirty and difficult (36). Accidents to fishing vessels are increasing every year. However, to this point there has been no standard regulation of vessels, manpower, endorsement and manning on fishing vessels [35]. Based on the results of research conducted by Suwardjo in 2010 stated that related to the characteristics of the work on fishing vessels are advised that the safety arrangements of the crew, the standards of fishing vessels, the manning, the working conditions on fishing vessels, education and training as well as the examination and certification are set apart [35].

3. RESEARCH METHODOLOGY

The methodology used in this study has several important aspects in general such as Literature Review, Preprocessing, Data Distribution based on K-Fold Cross Validation, Trial and comparison accuracy of Naïve Bayes Classifier and Probabilistic Neural Network algorithm, and final stage is the process of Evaluation of the accuracy of both algorithms used. In detail can be shown in Figure 2 .

The first stage was to conduct a literature study on the topic and the algorithm to be used. Further, data collection was done by using previous research data that is the number of fishing vessels from 2010 to 2012 in the field of marine downloaded from data.go.id [32]. At the pre-processing stage, a transformation was performed using a numerical scale based on each attribute, then normalized with

Min-max Normalization. Training data distribution and testing data were done after pre-processing stage is completed, resulting in 327 training data and testing data of 109 based on K-Fold Cross Validation. From the data distribution, then NBC and PNN algorithm will be implemented to know which algorithm better in classifying data. The K-Fold Cross Validation technique in this study will be modified into several sections in each K. For example K = 1 will be divided into 70% part of training data and 30% data testing. From the total K, part of the training and testing data will be combined based on the group. This technique was tested using highly efficient random data and has better accuracy than the usual K-Fold Cross Validation technique, with 2.5% difference. However, this technique has some disadvantages when applied using very large data, it requires many looping so that it affects the complexity of the algorithm.

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

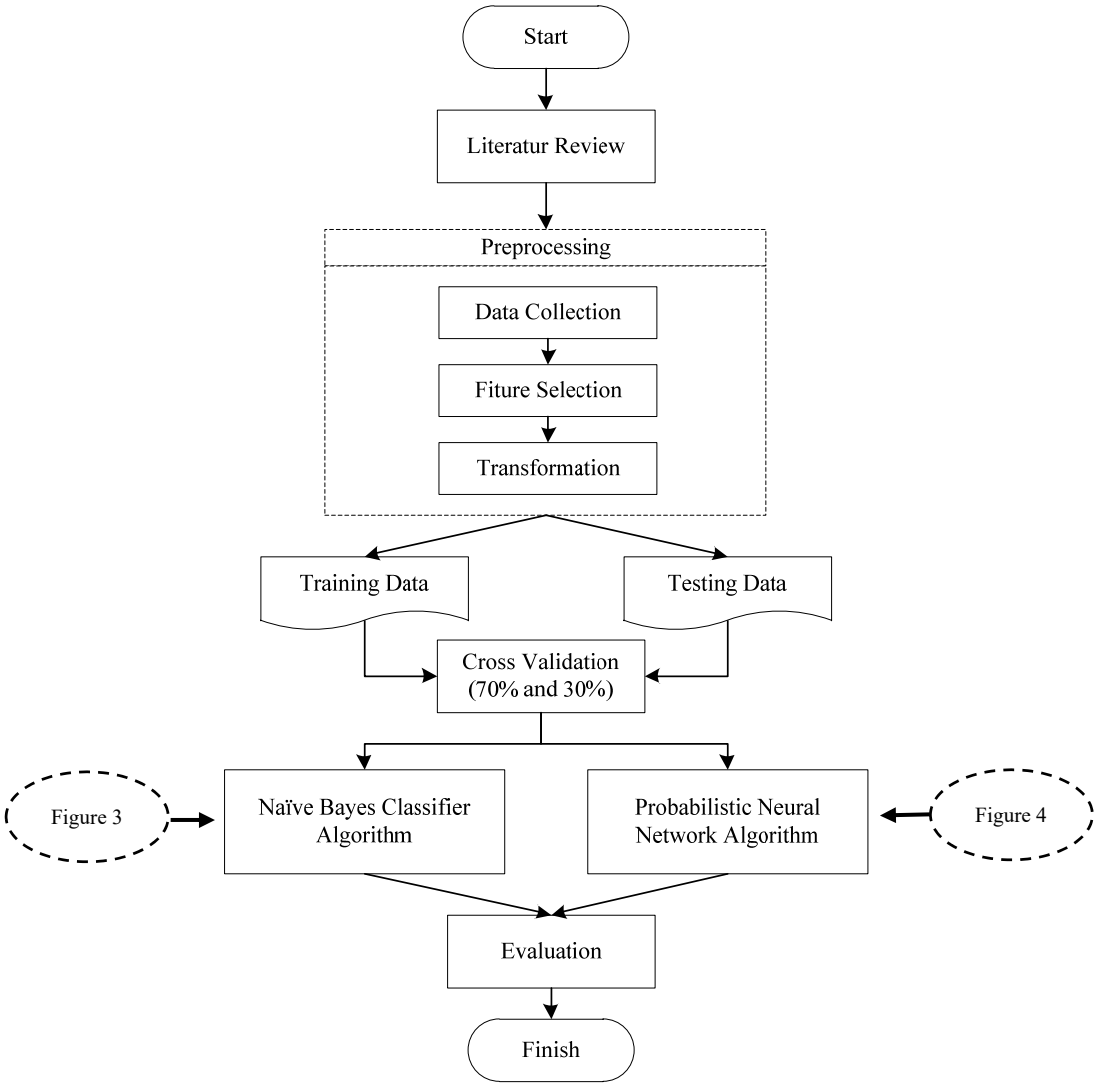

Figure 2. Research Methodology

4. RESULT AND ANALYSIS

Next is the analysis and results of classification using 2 algorithms, Naïve Bayes Classifier and Probabilistic Neural Network:

4.1. Data Used

The data used is data from the government site data.go.id in the field of marine. The distribution is 327 of training data and 109 of testing data with 436 data, same data based on previous research [1].

Table 1. Data Trining

No	Province	Province ID	Vessel Type	Category	Types of water	Years	Total	KapalID
1	Aceh	11	Kapal Motor 5 - 10 GT	Kapal Motor	Sea	2010	1155	KP030200
2	Aceh	11	Kapal Motor 5 - 10 GT	Kapal Motor	Sea	2011	1171	KP030200
3	Aceh	11	PTM - Perahu Papan - Total	Perahu Tanpa Motor	Sea	2011	1723	KP010200
4	Aceh	11	PTM - Perahu Papan - Total	Perahu Tanpa Motor	Sea	2010	1728	KP010200
…	…	…	…	…	…	…	…	…
327	North Sumatera	12	Kapal Motor Total	Sub Total	Sea	2011	18667	KP030000

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

Table 2. Data Testing

No	Province	Province ID	Vessel Type	Category	Types of water	Years	Total	KapalID
1	Aceh	11	Kapal Motor 5 - 10 GT	Kapal Motor	Sea	2012	1171	KP030200
2	Aceh	11	PTM - Perahu Papan - Total	Perahu Tanpa Motor	Sea	2012	1723	KP010200
3	Bengkulu	17	PTM - Perahu Papan - Total	Perahu Tanpa Motor	Sea	2012	1246	KP010200
4	Bengkulu	17	Perahu Tanpa Motor (PTM) - Total	Perahu Tanpa Motor	Sea	2012	1317	KP010000
…	…	…	…	…	…	…	…	…
109	South Sumatera	16	PTM - Perahu Papan - Total	Sub Total	Sea	2012	1734	KP010200

The table above consists of the provincial name, province id, vessel type, category, water type, year, total and shipid as the attributes used for the research. The explanation is as follows:

1. Province name: It is an attribute of provincial names located throughout Indonesia, which amounts to 33 provinces
2. ProvinceId: It is a unique code to distinguish provincial names across Indonesia
3. Type of ship: It is a ship that sailed in Indonesian waters that there are 19 types of ships
4. Category: These are the types of classes that sail in the waters of Indonesia are divided into 4 category.
5. Types of waters: It is where sailing vessels are sea and public. This attribute is the target class in this study.
6. Year: The year the ship operates from 2000 to 2012.
7. Total: The number of ships operating in the year that has been determined
8. KapalId: A unique code that distinguishes ship type:

4.2. Normalization and Class

Normalization was done by using min-max normalization with equation 5 below [22]:

$$d' = \frac{d - \min(p)}{\max(p) - \min(p)} \quad 5$$

The results of normalization can be seen in tables 3 below:

Table 3. Normalization of Data Trining

No	Province Name	N. Province ID	N. Vessel Type	N. Category	N. Water Type	N. Kapal ID	Water Type
1	Aceh	0.0000	0.4000	0.0000	0.0042	0.8182	Sea
2	Aceh	0.0000	0.4000	0.0000	0.0046	0.8182	Sea
3	Aceh	0.0000	1.0000	0.6667	0.0197	0.1818	Sea
4	Aceh	0.0000	1.0000	0.6667	0.0199	0.1818	Sea
…	…	…	…	…	…	…	…
436	North Sumatera	0.0120	0.5333	1.0000	0.4833	0.6364	Sea

4.3. Result of Calculation Using Naïve Bayes Classifier

The results obtained based on the calculation using naïve bayes algorithm are as follows:

Table 4. Precision, Recall and Accuracy

Attribute	True Laut	True Umum	Class Precision
Pred. Laut	24	0	100.00%
Pred. Umum	26	0	0.00%
Class Recall	48.00%	0.00%	

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

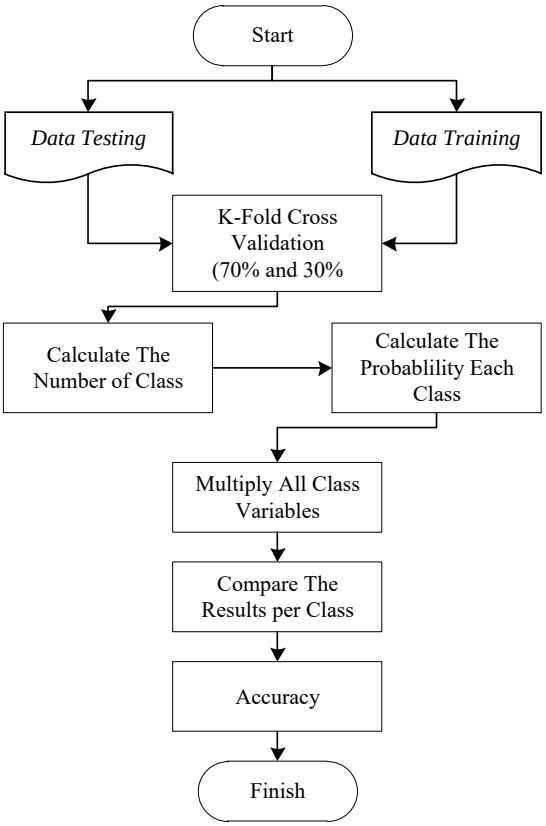

Figure 3. Naïve Bayes Classifier Process

Based on the results of calculations using Rapidminer, got the results on the prediction of sea class with an accuracy of 100%, whereas in the prediction of the general class of 0% accuracy. Accuracy got on Rapidminer using Naïve Bayes Classifier algorithm to the type of water data is by 48%.

4.4. Result of Calculation Using Probabilistic Neural Network

There are four layers in PNN i.e:

a. Input layer

Input layer is to accommodate the input data. In the input layer does not occur the computation process. The input layer serves to distribute the inputs to the neurons in the pattern layer. The input layer is fully connected to the next layer.

b. Pattern Layer

Each neuron in the pattern layer will perform the probability calculation (distance) between the inputs and the data stored on the neuron pattern layer. Pattern layer serves to compute the distance between input data and training data presented by neurons. The layer Pattern contains training data, which is represented as a neuron and denoted as xij. On the second layer computing is conducted with the formula [20]:

$$p(x|Ck) = \frac{1}{(2\pi)^{\frac{m}{2}}\sigma_k^{m|Ck|}} \sum_{pt \in Ck} \exp[-||x - w_i||^2 / (2\sigma_k^2)] \quad 6$$

Where:

|Ck| : number of training patterns on class Ck.
m : Is the input pattern vector dimension.
w_i : Is the weight vector in the pattern of the i-th training.
σk : Spread, Smoothing Parameters [20].

The value of the smoothing parameter can not be determined directly, but can be obtained through optimization methods or from experimental results [12]. Therefore, in this experiment the value of spread (σ) used, started from the default value of (0.1) followed by experiment of spread value 0.01 and 0.001.

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

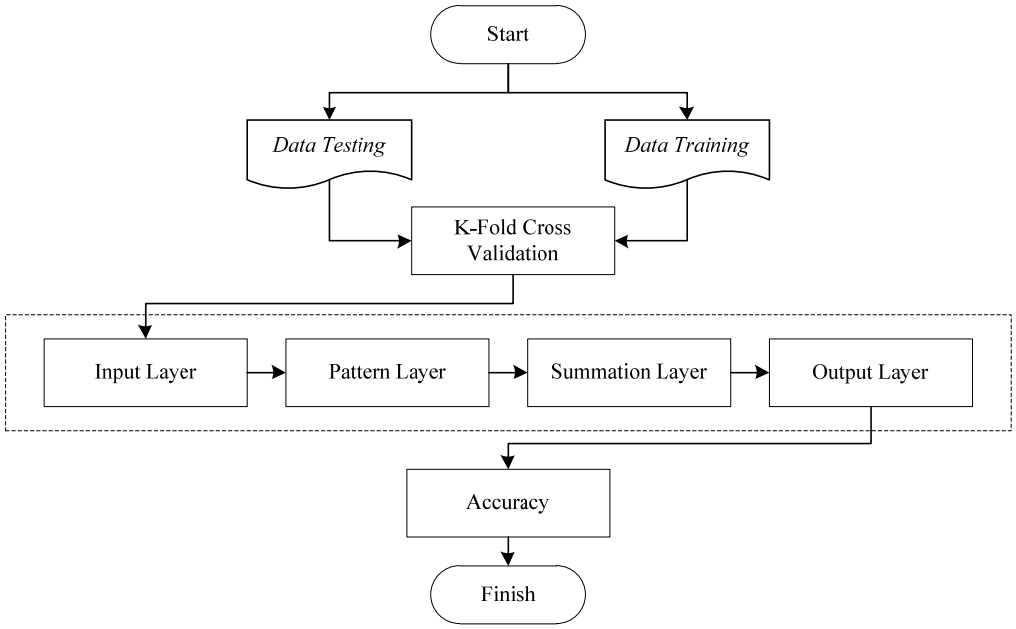

Figure 4. Probabilistic Neural Network Process

c. Summation Layer

Furthermore, the summation layer will receive input from each neuron pattern layer and will be summed so that it obtains how likely an input of x enters into a group.

d. Output Layer

In the output layer is taken the maximum value of the output vector then generate the class decision value. Decisions of classification can be formulated in a general way to minimize the risk by minimizing opportunities. The bayer decision rule in this case is simple enough to determine the Ck class, ie by choosing the largest Pr(x|Ck), this means:

$$d(x) = Ck \quad jika \quad p(x|Ck) > p(x|Ci)\,\Pr(Cj) \qquad 7$$

Where the output layer will produce classification results based on the results of summation neurons which has the greatest value. Here is the result of calculation using PNN algorithm:

Table 5. Random Data

No	Class Testing Data	Result of P1	Result of P2	Result of P3
1	Sea	Sea	Sea	Sea
2	Sea	General	General	Sea
3	Sea	Sea	Sea	Sea
4	Sea	Sea	Sea	Sea
5	Sea	Sea	Sea	Sea
6	Sea	Sea	Sea	Sea
7	Sea	Sea	Sea	Sea
8	Sea	Sea	Sea	Sea
9	Sea	Sea	Sea	Sea
10	Sea	Sea	Sea	Sea
....	....	....	....	....
50	Sea	General	General	Sea

Explanation of table 5 above is as follows:

1. Result of P1= 1st experiment using the value of spread 0.1
2. Result of P2= 2nd experiment using the value of spread 0.01
3. Result of P3= 3rd experiment using spread value 0.001

As for the results of accuracy on each experiment can be seen in the following figure:

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

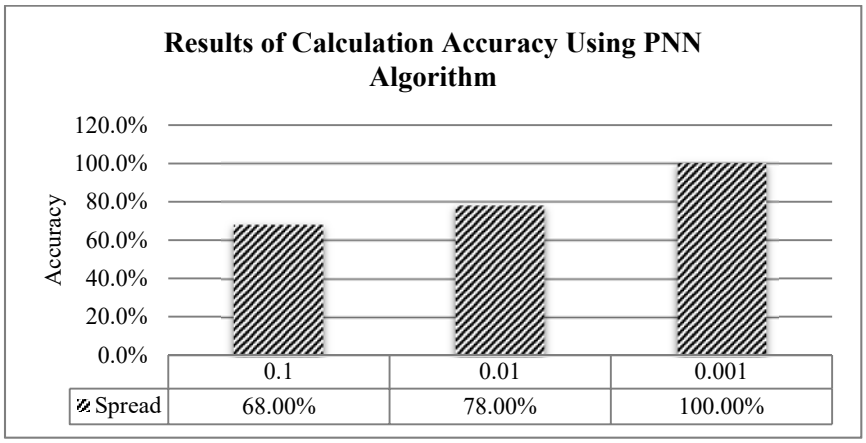

Figure 5. Results of Calculation Accuracy Using PNN Algorithm Fold 2

Figure 5 above is the result of PNN calculation based on the best fold set in this research. There are 4 folds used in performing experiments on this PNN algorithm with each training data is divided into 109 data. The weakness in this study are the size of the dataset used and therefore can not prove the research conducted by Franti [29] which states that PNN algorithm is a faster type of algorithm compared to other algorithms. The overall accuracy of each PNN fold in the study can be shown in table 7 below:

Tabel 7. Comparison of Accuracy on Each Fold

Spreat	Fold 1	Ford 2	Ford 3	Fold 4
0,1	55,00	68,00	61,00	65,00
0,01	77,00	78,00	77,00	74,00
0,001	100,00	100,00	100,00	100,00

Based on the results of calculations using Probabilistic Neural Network algorithm obtained the accuracy of 68% from spread 0.1, an accuracy of 78.00% from spread 0.01, and an accuracy of 100% from spread 0.001. In the picture above can be concluded that the calculation of PNN using the smaller spread resulted in higher accuracy.

Thus, the accuracy results from the classification using Naïve Bayes Classifier algorithm is 48% and with the highest Probabilistic Neural Network accuracy on the 0.001 spread is 100%. Meanwhile, in Rustiyan and Mustakim research in 2016 using k-Nearest Neighbor obtained the highest accuracy on k = 7 by 94% [1]. So the performance of Probabilistic Neural Network is better than Naïve Bayes Classifier and k-Nearest Neighbor.

Accuracy resulting from both algorithms has a close relationship between the dataset used and data distribution techniques. In reality, the concept of K-Fold Cross Validation applied to the 70% and 30% is not as good as data distribution in clustered data [23], where in the study, the more the data used, it will produce a higher accuracy. Data distribution technique in this study can be simulated as shown in Figure 6 below:

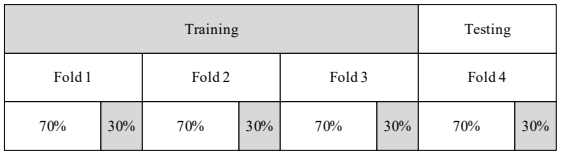

Figure 6. K-Fold Cross Validation (70% and 30%)

The data distribution with the model in figure 6 above can not be applied by using large data, so it can be concluded that the data sharing as mentioned above can only be applied to data with smaller dimension, the smaller the data the higher the accuracy.

9. CONCLUSION

The results show the accuracy of Naïve Bayes Classifier algorithm using RapidMiner tool is equal to 48%. While for PNN algorithm the three different spread values of 0.1 yield accuracy by 68%, 78% for spread value 0.01 and last experiment that spread 0.001 yield 100% accuracy were experimented. Therefore, in this study it can be concluded that the classification using PNN algorithm is better than Naïve Bayes Classifier and k-Nearest Neighbor. Data sharing technique using K-Fold Cross Validation with 70% and 30% model can only be applied by using small data, if applied to large data, either record or attribute, will experience higher complexity and drastic decrease in accuracy. In this research some experiments have been conducted using multiples of 100 data or 43.600 record with 5 multiplication attributes or 25 attributes used in this

research which only yield accuracy of 47%. Thus increasingly the amount of data and attributes will decrease its accuracy as well. Therefore, it can be concluded that the less data and attributes used then the accuracy will be higher.

The disadvantage of this research is that the amount of data is only 436 data and only has 5 attributes, for further research other neural network methods are proposed such as LVQ, Back-propagation and Perceptron to get better results.

ACKNOWLEDGEMENT

A biggest thanks to Faculty of Science and Technology UIN Sultan Syarif Kasim Riau on the financial support for this research and publication, the facilities and mental support from the leaders. And also thanks to Puzzle Reseach Data Technology (Predatech) Team Faculty of Science and Technology UIN Sultan Syarif Kasim Riau for their support to research can be done well.

REFRENCES:

[1] R Rustiyan and M Mustakim, "Klasifikasi Jenis Perairan pada Kapal Perikanan di Indonesia Menggunakan K-Nearest Neightbor", *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, pp. 178-185

[2] K. Sumathi, S. Kannan, K. Nagajaran, "Data Mining: Analysis of Student Database using Classification Techniques", *International Journal of Computer Applications*, Vol. 141, Iss. 8, 2016. pp. 22-27.

[3] L. Marlina, Muslim, APU Siahaan "Data Mining Classification Comparison(Naive Bayes and C4.5 Algorithms", *International Journal of Emerging Trends & Technology in Computer Science (IJETT)*, Vol. 38, Iss. 7, 2016. pp: 380-383.

[4] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga", *Citec Journal.* Vol. 2, Iss. 3, 2015. pp. 207-217.

[5] J Arief, "Algoritma Naïve Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa", *Jurnal Tekhnologi Informasi Dinamik*, Vol. 18, Iss.1, 2013. pp. 9 - 16.

[6] N. Nahdyah, F Aisjah and J Ilham, "Keragaman Jenis Kapal Perikanan di Kabupaten Takalar", *Jurnal IPTEKS PSP*, Vol. 1, Iss. 1, 2014, pp.81-94.

[7] Ridwan M, S Hadi and Sarosa, M , "Penerapan Data Mining untuk evaluasi kinerja Akademik Mahasiswa menggunakan Algoritma Naive Bayes Classifier", *Jurnal Electrics, Electronics, Communication, Controls, Informatics, Systems (EECCIS)* Vol. 7, Iss. 1, 2013. pp. 59-64.

[8] W Muslehatin, I Muhammad and M Mustakim, "Penerapan Naïve Bayes Classification Untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN SUSKA RIAU", *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI),* 2017. pp. 250-256.

[9] R Bhuvaneswari, and K. Kalaiselvi. "Naive Bayesian Classification Approach in Healthcare Applications", *International Journal of Computer Science and Telecommunications*, Vol. 3. Iss. 1. 2012. pp.106-112.

[10] Tsangaratos, Paraskevas, and I Ilia, "Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size", *Catena*, Vol. 145. Iss, 2016. pp.164-179.

[11] A Aszani and M Mustakim. Estimasi Pola Cuaca Wilayah Pekanbaru menggunakan Probabilistic Neural Network, *Prosiding: Seminar Nasional Teknologi Informasi Komunikasi dan Industri (SNTIKI) 8*, 2016, pp. 78-86.

[12] S Erfan, H Yasin, R Rahmawati, Klasifikasi Data Berat Bayi Lahir Menggunakan Probabilistic Neural network dan Regresi Logistik, *Jurnal Gaussian*, 2015, Vol. 4, Iss. 4, pp. 815-824.

[13] American Welding Society."Specification for Underwater Welding", *AWS D3. 6M. Miami, Florida: ANSI*, 1999.

[14] A Singh and S Sathyaraj, "A Comparison Between Classification Algorithms on Different Datasets Methodologies using Rapidminer", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 5, Iss. 5, 2016. pp. 560-563.

[15] B Santi, M Tripathy, JK Satapathy, "A novel approach for voltage secure operation using Probabilistic Neural Network in transmission network", *Journal of Electrical Systems and Information Technology*, Vol. 3, Iss. 1. 2016, pp. 141–150.

[16] FT Anggraeny, "Probabilistic Fuzzy Neural Network untuk Deteksi Dini Penyakit Jantung Koroner", *Jurnal Informatika Mulawarman*, Vol. 6, Iss. 2, 2011, pp. 75-79.

[17] A Lotfi, A Benyettou, "A Reduced Probabilistic Neural Network for The Classi Cation Of Large Databases", *Turkish Journal Of Electrical Engineering & Computer Sciences*, Vol. 22, 2014, pp. 979-989.

[18] D Fattah, "Identifikasi Citra Daging Sapi dan Babi menggunakan Ekstaksi Fitur HSV dan Fitur Gabor dengan Klasifikasi Probabilistic Neural Network", *Master Tesis. Pekanbaru: Universitas Islam Negeri Syarif Kasim Riau*. 2011.

[19] K Abhishek, "Comparative Study of Data Classifiers Using Rapidminer", *International Journal of Engineering Development and Research*, Vol. 5, Iss. 2, 2017, pp. 1041-1043.

[20] R Saputra, "Implementasi Metode Wavelet Haar dan Probabilistic Neural Network untuk Pengenalan Citra Daging Babi dan Daging Sapi", *Master Tesis: Universitas Islam Negeri Syarif Kasim Riau*, 2015.

[21] Y Shenghao, Z Dongming, C Wei and H Hui. "Oil-immersed power transformer internal fault diagnosis research based on probabilistic neural network." *The second International Workshop on Wireless Technology Innovations in Smart Grid (WTISG), Procedia Computer Science 83*, Vol. 83, 2016, pp.1327 – 1331.

[22] Jain and Bhandare. "Min Max Normalization Based Data Perturbation Method for Privacy Protection", *International Journal of Computer & Communication Technology*, Vol. 2 Iss. 8, 2011, pp. 45-50.

[23] M Mustakim, "Effectiveness of K-means Clustering to Distribute Training Data and Testing Data on K-Nearest Neighbor Classification". *Journal of Theoretical and Applied Information Technology (JATIT),* Vol. 95, Iss. 21, 2017, pp. 5693-5700.

[24] K Saravanan and S Sasithra, "Review On Classification Based On Artificial Neural Networks", *International Journal of Ambient Systems and Applications (IJASA),* Vol. 2, Iss. 4, 2014, pp. 11-18.

[25] DK Marcomini, AAO Carneiro, H Schiabel, "Application of Artificial Neural Network Models in Segmentation and Classification of Nodules in Breast Ultrasound Digital Images", *International Journal of Biomedical Imaging*, 2016, pp. 1-13.

[26] A Yasar, E Kaya, I Saritas, "Classification of Wheat Types by Artificial Neural Network", *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 4, Iss. 1, 2016, pp. 12-15.

[27] B Tian, MR Azimi-Sadjadi, "Comparison of two different PNN training approaches for satellite cloud data classification", *IEEE Transactions on Neural Networks*, Vol. 12, Iss. 1, 2001, pp. 164-168.

[28] M Mustakim, A Buono, and I Hermadi, "Performance Comparison Between Support Vector Regression and Artificial Neural Network for Prediction of Oil Palm Production". *Journal of Computer Science and Information*, Vol. 9, Iss. 1, 2016, pp. 1-8.

[29] P Franti, T Kaukoranta, DF Shen, KS Chang, "Fast and memory efficient implementation of the exact PNN", IEEE Transactions on Image Processing, Vol. 9, Iss. 5, 2000, pp. 773-777.

[30] SS Rao, "Engineering Optimization: Theory and Practice", John Wiley and Sons, Newyork, 2006.

[31] G Adhani, A Buono, A Faqih, "Support Vector Regression modelling for rainfall prediction in dry season based on Southern Oscillation Index and NINO3.4" *International Conference on Advanced Computer Science and Information Systems,* 2013, pp. 315-320.

[32] http://data.go.id/dataset/jumlah-kapal-penangkap-ikan-nasional/resource/f04b1f58-7326-4679-8d0c-87f96b5ac594 Acces Date: November 2017.

[33] N Gupta, A Rawal, VL Narasimhan, Shiwani S, "Accuracy, Sensitivity and Specificity Measurement of Various Classification

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

Techniques on Healthcare Data", *IOSR Journal of Computer Engineering (IOSR-JCE).* 3013, Vol. 11, Iss. 5, pp: 70-73.

[34] M Behzad, K Asghari, M Eazi, M Palhang, "Generalization Performance of Support Vector Machines and Neural Networks in Runoff Modeling", *Elsevier Expert Systems with Applications*, 2009, Vol. 36, Iss. 4, pp: 7624-7629.

[35] D Suwardjo, J Haluan, I Jaya and SH Poernomo, "Keselamatan Kapal Penangkap Ikan, Tinjauan dari Aspek Regulasi Nasional dan Internasional", *Jurnal Teknologi Perikanan dan Kelautan*, Vol. 1, Iss. 2, 2010, pp. 1-13.

[36] International Maritime Organization, "Document for Guidance on Training and Certification of Fishing Vessel Personnel", *FAO of United Nations, ILO and IMO, International Labor Organization, and Food Agriculture Organization*, 2006, Code of Safety for Fishermen and Fishing Vessels 2005.