Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 4, No. 2, September 2021, pp. 108 – 116
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Optimization of the Naïve Bayes Classifier (NBC) Algorithm Using the Sparrow Search (SSA) Algorithm to Predict the Distribution of Goods Receipts

[1]Rachma Oktari, [2]Tjong Wan Sen
[2]Information Technology, Faculty of Computing, President University
Email: [1]rachma.oktari@student.president.ac.id ,[2]wansen@president.ac.id

Article Info

Article history:
Received Sept 05th, 2021
Revised Sept 30st, 2021
Accepted Oct 10th, 2021

Keyword:
Data Mining
Enterprise Resource Planning (ERP)
Naïve Bayes Classifier (NBC)
Sparrow Search Algorithm (SSA)

ABSTRACT

Distribution must be able to meet all needs based on sales orders from consumers, be responsible for the delivery order process running optimally, and ensure the good receipt process is in accordance with consumer sales order requests. PT. Diamond Cold Storage currently uses Enterprise Resource Planning (ERP) to record all reports from production to sales. But in reality there are still some obstacles in the distribution section. In the good receipt process, several items were found that did not match the sales order, such as: the item did not match the order request or the item did not match the order request. The process of mismatching the good receipt with the sales order will be met with the completion of the good receipt process or the bad thing is that there is a cancellation, so this causes a loss for the company. This study uses data mining techniques with the Naïve Bayes Classifier algorithm to predict the distribution of goods receipts based on distribution data, and uses the Sparrow Search Algorithm (SSA) algorithm to optimize the Nave Bayes Classifier by selecting features to improve accuracy. In this study, the results obtained that the SSA algorithm can improve the performance of NBC from 95.05% to 97.95%.

Corresponding Author:
Rachma Oktari,
Information Technology, Faculty of Computing
President University,
Jababeka Education Park, Jl. Ki Hajar Dewantara, North Cikarang, Bekasi, West Java, Indonesia.
Email: rachma.oktari@student.president.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v2i2.15339

1. INTRODUCTION

The aspect of product distribution is a crucial spotlight because most of the production costs are spent by the distribution process to ordering agents. Distribution must be able to meet all needs based on sales orders from consumers, be responsible for the delivery order process running optimally, and ensure the good receipt process is in accordance with consumer sales order requests. PT. Diamond Cold Storage currently uses Enterprise Resource Planning (ERP) to record all reports from production to sales. But in reality there are still some obstacles in the distribution section. In the good receipt process, several items were found that did not match the sales order, such as: the item did not match the request order or the item did not match the request order.

The settlement process takes quite a long time, between the distribution party and the customer, while in the case of cancellation it is influenced by several factors, namely: from internal (company) and external (customer) parties. Internal factors, including: deliveries that exceed the rules or time limits from the customer, items that are rejected by the customer on the grounds that they do not match the quality desired by the customer. External factors, among others: a customer warehouse that is already full, input errors in the business to business customer system, and this causes losses for the company.

Based on the description above, this study uses data mining techniques to find the accuracy value which is used as a reference for predicting the distribution of goods receipts by looking at customer data, items,

item quality, delivery distance, vehicle temperature, vehicle conditions, weather, traffic conditions, delivery time, sales orders, and delivery orders. The basic approach in data mining is to summarize data and extract useful issues that were previously unknown. Data Mining can find hidden trends and patterns that don't arise in simple query analysis and as a result can have a crucial part in finding knowledge and making decisions. Such tasks can be predictive such as classification and regression or descriptive such as clustering and association [1].

Several studies have been carried out using data mining techniques to explore various issues from a database, such as the research conducted by Erwina Nurul Azizah, et al using Web History data and the number of interactions of students' web pages with the Naïve Bayes algorithm and the C4.5 algorithm to predict academic performance. students in the Learning Environment. the origin of the data that has been processed using two algorithms. The results obtained, both algorithms have almost the same level of accuracy. Naive Bayes accuracy is superior to 63.8% of C4.5 only 0.2% different from Naive Bayes accuracy [2]. Subsequent research conducted by Meylan Wongkar, et al in analyzing sentiment with twitter data regarding the 2019 presidential candidate of the Republic of Indonesia using the python programming language. In this study, a comparison was made using NBC, SVM and K-NN methods using RapidMiner, and resulted in an NBC accuracy value of 75.58%, an SVM accuracy value of 63.99% and a K-NN accuracy value of 73.34%. NBC outperforms other methods [3].

Naïve Bayes has several advantages, namely fast in calculation, simple algorithm and high accuracy. Naïve Bayes Classifier is better applied in large data and can handle incomplete data (missing values). and can handle irrelevant attributes and noise data. However, the Naïve Bayes Classifier also has drawbacks, namely the selection of attributes that affect the accuracy value. So the Nave Bayes Classifier needs to be optimized by weighting the attributes so that the Nave Bayes Classifier can work more effectively [4].

In solving these problems, this study uses one of the meta-heuristic optimizations, namely the Sparrow Search algorithm proposed by Xue and Shen in 2020 to increase the accuracy of the Naïve Bayes Classifier [5]. Meta-heuristic optimization techniques have become very popular over the last 2 periods, such as Particle Swarm Optimization (PSO) which was first developed by Kennedy and Eberhart in 1995, Ant-based techniques were first developed by Dorigo in 1996 using Ant Colony Optimization (ACO) to complete the Traveling Salesman, Genetic Algorithm was first developed by John Holland in the 1970s, Harmony Search Algorithm (HSA) was introduced by Zong Woo Geem, Joon Hoon Kim, and GV Loganathan in 2001, Bee Algorithm was developed by Pham,

There are several reasons why meta-heuristics are so commonly used, namely: simplicity, flexibility, derivation-free procedures, and avoidance of local optima. so that many theoretical works use optimization techniques like this in various fields of learning [6]. In the Sparrow Search Algorithm (SSA) technique, there are several ways to optimize, including reducing the error rate and feature selection. The data used in this study is the distribution of goods which aims to predict the distribution of goods receipts. With the Sparrow Search Algorithm (SSA) algorithm can optimize the accuracy value of the Naïve Bayes Classifier by means of feature selection.

2. RESEARCH METHOD

The following is a schematic of the research stages regarding the prediction of the distribution of goods receipts, which can be seen in Figure 1.

2.1. Dataset Collection

The process of collecting data, the data used is distribution data originating from delivery order and good receipt data from 2020 to 2021.

2.2. Dataset Preprocessing

The process of converting raw data or known as raw data, such as: merging data (data from several files such as data per month of delivery orders and good can be combined into one distribution file), then discarding noise data, duplicate data, empty value data, then the transformation process data from nominal to numeric, and then perform the feature selection process with the SSA algorithm for SSA-NBC testing. The model is built using the NBC algorithm and the feature selection method using the SSA algorithm.

In this stage, the distribution data will be reprocessed using data preprocessing techniques, then the training process will be carried out using the NBC and SSA-NBC algorithms and then a comparison will be made based on the values of accuracy, precision, and recall. The training model is shown in Figure 4. After preprocessing the data, then the data is divided into training data and testing data which will be entered into the NBC model architecture.

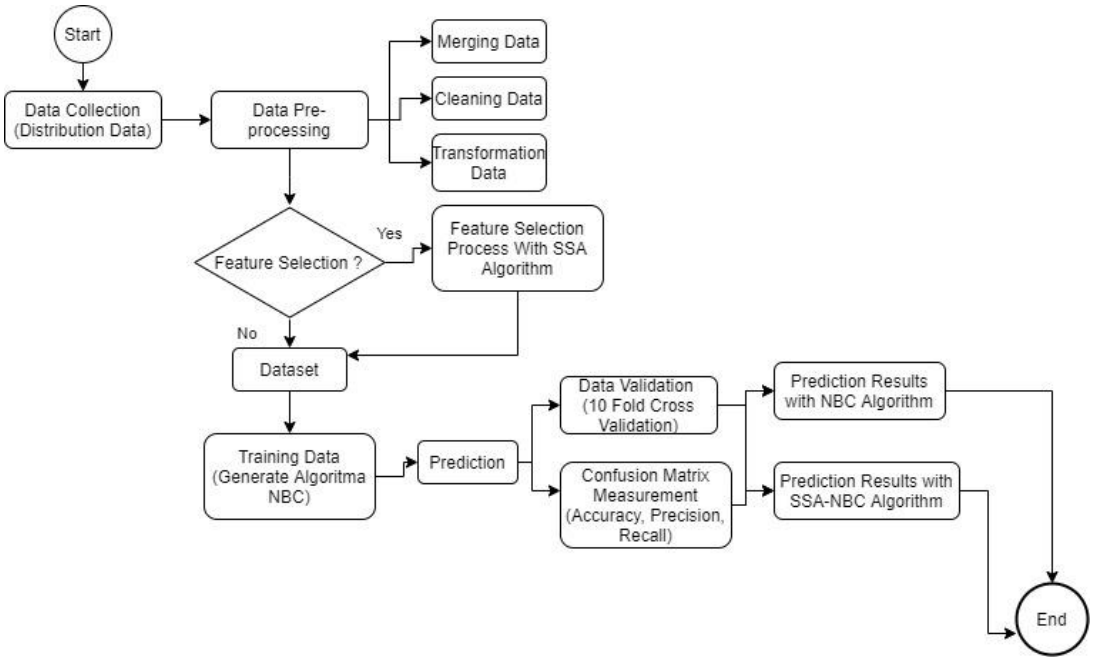

Figure 1. Research Stage

2.3. Training Data

Data training will be carried out using the NBC and SSA algorithms. The training conducted using google colab platform. The experiment in the first step will be trained using the NBC algorithm and then using NBC-SSA to get a comparison of the results of accuracy, precision and recall values. The following is an architectural model using a comparison between the NBC and SSA-NBC algorithms in finding the best accuracy, precision and recall values from the two models presented in Figure 2, Figure 3 and training model in Figur 4.

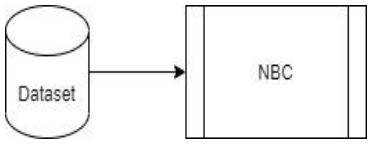

Figure 2. Model Architecture Without SSA

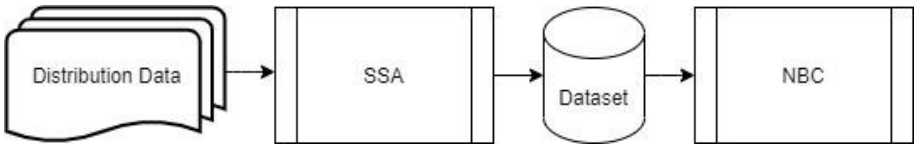

Figure 3. Model Architecture with SSA

2.4. Naïve Bayes Classifier

Naive Bayes classifier is a classification method with simple probability that applies Bayes theorem independently as a classifier in various real world problems such as sales prediction based on priority, and document categorization. Naive Bayes classification is built by training data to estimate the probability of each category contained in the characteristics of the document being tested [7]. The system will be trained using new data (training data and test data) and then given the task of guessing the value of the target function from the data.

NBC Method Flow:

1. Read training data
2. Calculate the number and probability:
   a) Find the mean and standard deviation of each parameter. To calculate the average value (mean) can use equation (1):

$$\mu = \frac{x1+x2+x3+\cdots+xn}{n} \quad (1)$$

μ: Average (mean), xi: Value of the 1st sample, n: Number of samples, and equation (2) to calculate the standard deviation (standard deviation) can be seen as follows:

$$\sigma = \sqrt{\frac{\sum_1^n=1\ (xi-\mu)^2}{n-1}} \quad (2)$$

σ: Standard deviation, xi: 1-th value of x, μ: Average count, n: Number of samples

b) Find the probabilistic value by calculating the number of appropriate data from the same category divided by the number of data in that category.

3. Get the values in the table mean, standard deviation and probability.
4. The solution is then generated.

For classification with continuous data, the Gaussian Density formula is used, in equation (3)

$$P = (Xi = xi|Yi = yi) = \frac{1}{\sqrt{2\pi\sigma ij}} e^{-\frac{(xi-\mu ij)2}{2\sigma^2 ij}} \quad (3)$$

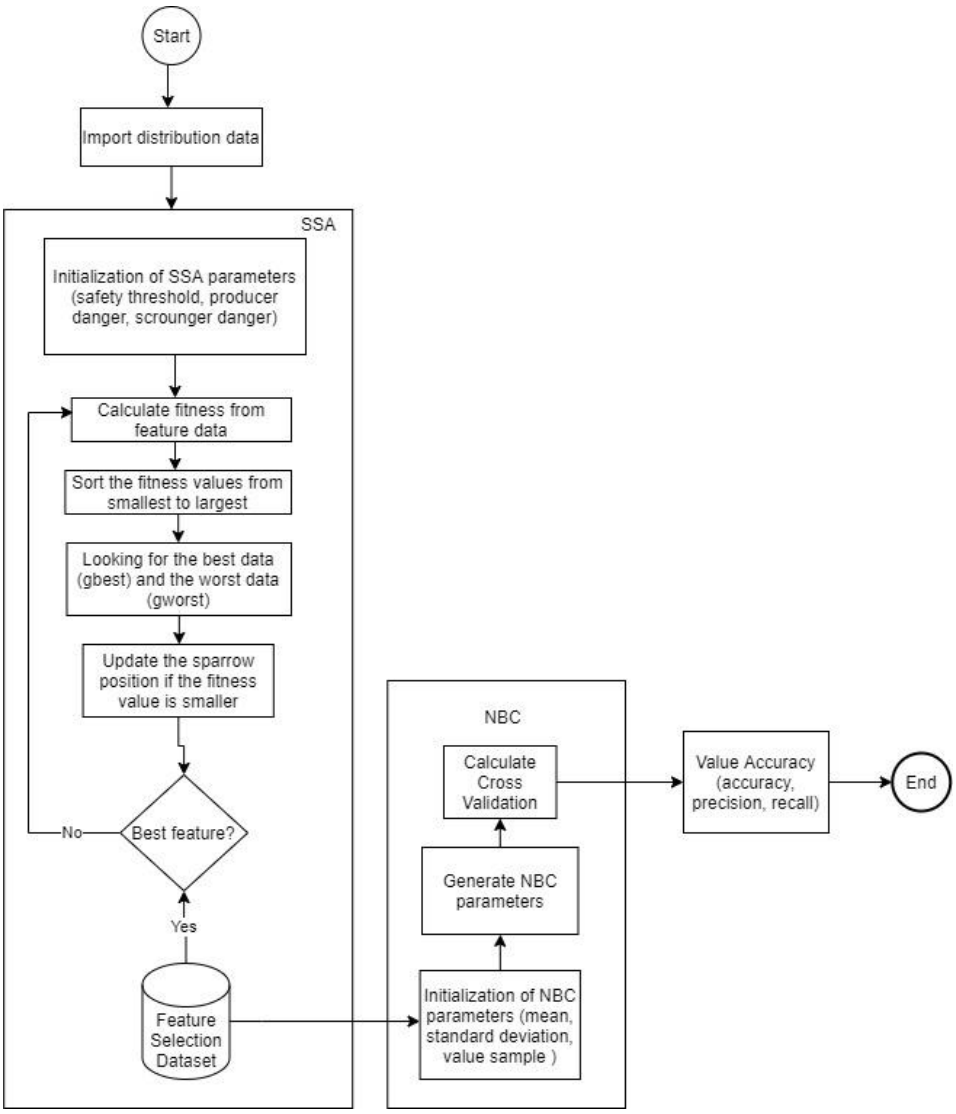

Figure 4. Training Model

2.5. Algoritma Sparrow Search (SSA)

The sparrow search algorithm Sparrow Search (SSA) is an effective optimization technique, which simulates the behavior of a group of sparrows in foraging. SSA observed that sparrows show high fidelity to their producer-scrounger (feeding) or scrounger (feeding) roles, leading them to speculate that the producer-scrounger (PS) role could be improved individually [8].

SSA algorithm steps [8]

Step 1. Input variables

T : maximum iteration
PD : number of producers
SD : number of sparrows that sense danger
R2 : mark alarm (random number)
ST : safety threshold
n : number of sparrows

Initialize a population of n sparrows and define the relevant parameters. Determine the population size $T \geq n$. Sample three variables x, y and z. To find the best set of values for x, y and z so that the total value is equal to the value of t (target), use equation (4).

$$x + y + z = t \tag{4}$$

Step 2. Create a random grouping of sparrows in the current population. The position of the sparrow can be represented in the following matrix:

$$X = \begin{bmatrix} x1,1 & x1,2 & \cdots & \cdots & x1,d \\ x2,1 & x2,2 & \cdots & \cdots & x2,d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ xn,1 & xn,2 & \cdots & \cdots & xn,d \end{bmatrix}$$

where *n* is the number of sparrows and *d* indicates the dimension of the variable to be optimized.

Step 3. Compare the fitness values of the n sparrows listed in the grouping, and copy the best ones to the next generation. Discard the string being compared. If n is the number of sparrows, x and y indicate the dimensions of the variable to be optimized. This fitness function automatically eliminates the grouping of points that are not dominated [5]. Then Δx is the number of features calculated into, function the fitness of all the sparrows can be calculated using an equation (5).

$$f(x) = t - \Delta x \tag{5}$$

Step 4. Calculate the best and worst scores

$$x_{worst}^{best}$$

After getting the fitness value from all the data, then sorting is done from the smallest to the largest fitness. Shows current worst and best solutions in global optimal locations.

$$X_{ij}^{t+1} = \begin{cases} X_{ij}^{t} \cdot \exp\left(\frac{-i}{\alpha \cdot itermax}\right) & if\ R2 < ST \\ X_{ij}^{t} + Q \cdot L\ \ if\ R2 \geq ST \end{cases} \tag{6}$$

Producer: where *t* represents the current iteration, *j = 1, 2, ... , d. Xt i,j* represents the sparrow dimension value in row I , column j, in iteration t. itermax is the constant with the largest number of iterations $\alpha \in (0, 1]$ which is a random number. *R2 (R2 ∈ [0, 1])* and *ST (ST ∈ [0.5, 1.0])* each represents an alarm value and a safety threshold. *Q* is a random number that follows a normal distribution, *L* denotes a matrix of $1 \times d$ (dimensions) in which every element in it is 1. When $R2 < ST$, it means that there are no predators around, so producers stay safe. If $R2 \geq ST$, it means that there are predators around, and all sparrows immediately fly to a safer area, it can be calculated using equation (6).

$$X_{ij}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^{t} - X_{ij}^{t}}{i^2}\right) & if\ i > n/2 \\ X_P^{t+1} + \left|X_{ij}^{t} - X_P^{t+1}\right| \cdot A^{+} \cdot L & otherwise \end{cases} \quad (7)$$

Scrounger: where XP is the optimal position occupied by the producer. Xworst shows the current global worst location. A represents a *1 × d* matrix in which each element in it is randomly assigned a value of 1 or 1, and *A+ = AT (AAT) 1*. When *i > n/2*, it indicates that the ith scrounger with a worse fitness value, is most likely to starve, can be calculated using equation (7).

$$X = \begin{cases} X_{best}^{t} + \beta \cdot \left|X_{ij}^{t} - x_{best}^{t}\right| & if\ fi > fg \\ X_{ij}^{t} + K \cdot \left(\frac{\left|X_{ij}^{t} - X_{worst}^{t}\right|}{(fi - fw) + \varepsilon}\right) & if\ fi = fg \end{cases} \quad (8)$$

Emergency: assume that these sparrows, aware of the danger, make up 10% to 20% of the total population. The initial position of sparrows is represented randomly in the population. In the mathematical model, where *Xbest* is the current global optimal location. β as a sparrow parameter to move from random numbers with mean values 0 and 1. $K \in [-1, 1]$ which is a random number. Here *fi* is the current fitness value of the sparrow. *fg* and *fw* are the current global best and worst fitness values, respectively. ε is the smallest constant to avoid zero division errors.

For convenience, when *fi > fg* indicates that the sparrows are far from each other. *Xbest* represents a population center location and its surroundings are safe. If *fi = fg* indicates that the sparrows in the middle of the population are aware of the danger and need to congregate with other sparrows. *K* indicates the direction in which the sparrow is moving and is also the coefficient of control of the sparrow's step, which can be calculated using equation (8).

Step 5. Repeat Steps 3 and 4 until no more selections are required for the next generation.

2.6. Feature Selection

Feature selection is an optimization problem that plays an important role in dealing with classification problems. It is the process of selecting an optimal subset of features from a data set so that the classifier can obtain better accuracy and/or reduce computational load. However, removing irrelevant features is challenging and time-consuming because of the large search space and the relationship between features [9]. The traditional FS method has the disadvantages of nesting effects and computational costs. To solve this problem, population-based optimization algorithms are used, such as gray wolf optimization (GWO), particle swarm optimization (PSO), genetic algorithm (GA), genetic programming (GP), ant colony optimization (ACO). Each algorithm has its advantages and disadvantages. For example, The ACO algorithm has a weakness that is slow search, and the PSO algorithm has a weakness that is easy premature convergence. Therefore, it is very important to improve the optimization algorithm. The proposed SSA algorithm is superior to other algorithms in terms of search precision, gence level conversion, stability and avoidance of optimal local values [10].

The feature selection process in this study is to find data attributes that have an effect on improving NBC performance. The feature selection process in this study is to determine the safety threshold value, producer danger, scrounger danger, lower limit value, upper limit value and features compared to other features through a 5 epoch process, the results obtained are the accuracy results from the comparison of each feature, can be seen in Table 1. Furthermore, the selected feature is the feature that has the highest accuracy from each epoch. The features before being selected were 11 features, after going through the feature selection process, the number of features became 6. The data from the feature selection are presented in Table 2.

Table 1. Best Feature Results in Process 5 Epoch

Epoch	Maximum Accuracy	Maximum Accuracy Column	Maximum Precision	Maximum Precision Column	Maximum Recall	Maximum Recall Column
0	95.04%	[Full Data]	82.82%	[Full Data]	73.97%	[Full Data]
1	96.51%	[X3, X7, X9 and X11]	84.09%	[X3, X7, X10 and X11]	85.44%	[X3, X7, X9, X10 and X11]
2	96.51%	[X3, X7, X9 and X11, X1, X3, X5, X7, X9, X10 and X11]	85.27%	[X3, X7, X10 and X11]	85.87%	[X2, X3, X7, X9, X10 and X11]
3	96.56%	[X3, X7, X9 and X11]	84.86%	[X3, X7, X9, X10 and X11]	85.41%	[X3, X7, X9, X10 and X11]
4	97.00%	[X3, X5, X7, X9, X10 and X11]	85.68%	[X3, X7, X10 and X11]	85.23%	[X3, X7, X9, X10 and X11, X3, X5,

Epoch	Maximum Accuracy	Maximum Accuracy Column	Maximum Precision	Maximum Precision Column	Maximum Recall	Maximum Recall Column
5	96.78%	[X3, X5, X7, X9, X10 and X11]	84.95%	[X3, X7, X9, X10 and X11]	84.67%	X7, X9, X10 and X11] [X2, X3, X7, X9, X10 and X11]

Table 2. Feature Selection Results

Actual string data with the highest accuracy after optimization and feature selection processing							
['X3', 'X5', 'X7', 'X9', 'X10', 'X11']							
	Quantity Do	Temperature	Weather	Item Quality	Vehicle	Other Problems	Good Receipt
0	1.0	FROZEN	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
1	1.0	FROZEN	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
2	1.0	FROZEN	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
3	1.0	FROZEN	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
4	1.0	FROZEN	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
...	...	...	...	...	...	...	...
7996	1.0	DRY	BRIGHT	GOOD	NOT OK	THERE IS NOT ANY	FULL
7997	1.0	DRY	BRIGHT	GOOD	NOT OK	THERE IS NOT ANY	FULL
7998	1.0	DRY	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
7999	1.0	DRY	BRIGHT	GOOD	OK	THERE IS NOT ANY	FULL
8000 rows × 7 columns							

2.7. Validation

Using "train and test" with K-Fold Cross Validation, namely by testing the amount of error in the test data [11]. In 10 fold Cross Validation, the amount of data used is 8000 data, the data is divided into 10 folds of the same size, so it has 10 subsets of data, 9 folds (7200 data) for training, 1 fold (800 data) for testing. Cross Validation K-fold is used because it can reduce computational time while maintaining the accuracy of the estimate [11]. Model architecture for the use of the NBC algorithm without SSA and using NBC with SSA.

2.8. Measurement Stage

Confusion matrix is a dataset that only has two classes, one class is positive and the other class is negative [11]. This method uses Table 3.

Table 3. Confusion Matrix Model [12]

		Actual	
		+1 (Positive)	-1(Negative)
Prediction	+1 (Positive)	TP	FP
	-1 (Negative)	FN	TN

Description:

a. *TP* is the amount of data that has a positive value and is predicted to be true as positive
b. *TN* is the amount of data that is negative and is predicted to be false as negative
c. *FN* is the amount of data that is positive and is predicted to be false as negative
d. *FP* is the amount of data that has a negative value and is predicted to be true as positive

The following is the equation of the confusion matrix model:

a. The accuracy value (acc) is the proportion of the number of correct predictions. Can be calculated using equation (9):

$$acc = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (9)$$

b. Recall or the true positive rate (tp) is the proportion of correctly classified positive cases, which is calculated using equation (10):

$$recall = \frac{TP}{TP+FN} \quad (10)$$

c. *Precision* (p) is the proportion of predicted positive positive cases that are correct, which is calculated using equation (11):

$$precision = \frac{TP}{TP+FP} \quad (11)$$

The model architecture for using the NBC algorithm without SSA and NBC using SSA can be seen in Figure 2 and Figure 3. The 10 cross validation process before and after feature selection with 3x3 confusion matrix measurements, because there are 3 good receipt targets or labels used, namely: cancel, partial, and full.

3. RESULTS AND ANALYSIS

After the data training process with the NBC algorithm and the SSA-NBC algorithm, using 10 fold cross validation data validation and measurement of the confusion matrix, the following are the experimental results of the model that has been trained, which can be seen in Table 4.

The NBC model without SSA uses 11 data attributes, including customer, item, quantity do (delivery order), distance, vehicle temperature, traffic, weather, delivery time, item quality, vehicle condition, other problems, and targets. The results of the average value of 95.05% accuracy, 81.01% average precision and 74.27% recall average. For the NBC model using SSA, it uses features that have been selected into 6 features, namely quantity do, vehicle temperature, weather, item quality, vehicle condition, other problems, and targets. The results of the average value of accuracy 97.95%, the average precision 84.40% and the average recall 83.32%. The comparison graph of training results is available in Figure 6.

Table 3. Training Results

CV K-Fold	*NBC*			*SSA-NBC*		
	Accuracy	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
1	96,13%	81,87%	74,17%	97,63%	95,09%	78,89%
2	96,75%	83,65%	76,58%	97,75%	93,55%	77,78%
3	97,63%	84,17%	81,77%	99,88%	70,37%	94,27%
4	95,63%	84,85%	77,12%	98,75%	85,19%	74,83%
5	95,50%	88,99%	69,17%	99,00%	85,19%	83,16%
6	96,13%	86,58%	72,07%	97,10%	93,18%	68,42%
7	94,88%	85,76%	72,67%	97,88%	77,69%	66,67%
8	93,50%	82,67%	71,58%	96,38%	78,63%	93,85%
9	95,88%	80,70%	73,81%	98,75%	82,55%	97,70%
10	88,50%	60,88%	73,72%	96,38%	82,55%	97,67%
Average	95,05%	82,01%	74,27%	97,95%	84,40%	83,32%

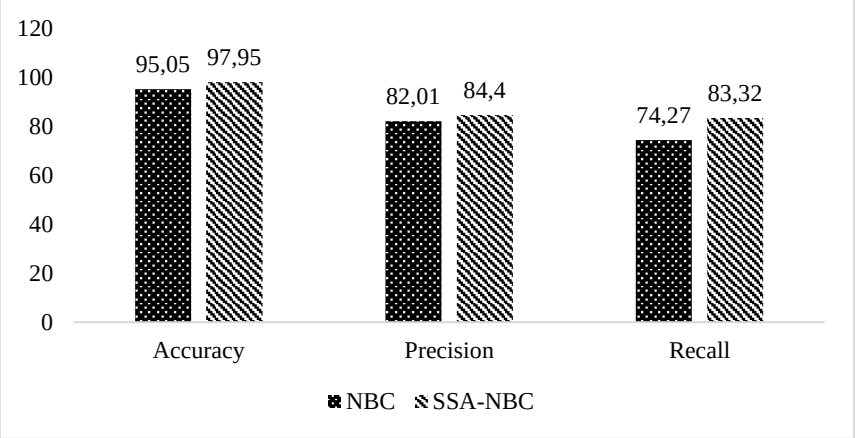

Figure 5. Comparison of Accuracy, Precision and Recall Values

Based on the research that has been done, SSA has increased the performance of NBC by 2.90%, from 95.05% to 97.95%. Based on the selection of features, SSA produces features that can increase the accuracy of the prediction of the distribution of goods received, namely: quantity do, vehicle temperature, weather, item quality, vehicle condition, other problems, and targets.

4. CONCLUSION

Based on research that has been done in predicting the distribution of goods receipts by selecting features using SSA to improve NBC performance, it can be concluded that SSA can improve NBC performance. So that the SSA-NBC method can predict the distribution of goods receipts more precisely on the distribution dataset of PT. Diamond Cold Storage in 2020 to 2021.

REFERENCES

[1]. Wanto, Data Mining Algoritma Dan Implementasi. Medan: Yayasan Kita Menulis, 2020.
[2]. E. N. Azizah, U. Pujianto, E. Nugraha, and Darusalam, "Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment," 2018 4th Int. Conf. Educ. Technol. ICET 2018, no. 1, pp. 18–22, 2018, doi: 10.1109/ICEAT.2018.8693928.
[3]. M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019, pp. 1–5, 2019, doi: 10.1109/ICIC47613.2019.8985884.
[4]. H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," J. Teknol. Inf. dan Ilmu Komput., vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.
[5]. Ouyang, D. Zhu, and Y. Qiu, "Lens Learning Sparrow Search Algorithm," Math. Probl. Eng., vol. 2021, 2021, doi: 10.1155/2021/9935090.
[6]. Pangestu et al., "Optimizing Neural Network Classifier for Diabetes Data Using Metaheuristic Algorithms," vol. 6, no. 2, pp. 85–91, 2017.
[7]. Lingga Aji Andika, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," Indones. J. Appl. Stat., 2019.
[8]. J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," Syst. Sci. Control Eng., vol. 8, no. 1, pp. 22–34, 2020, doi: 10.1080/21642583.2019.1708830.
[9]. J. Gholami, F. Pourpanah, and X. Wang, "Feature selection based on improved binary global harmony search for data classification," Appl. Soft Comput. J., vol. 93, p. 106402, 2020, doi: 10.1016/j.asoc.2020.106402.
[10]. S. Karthick and N. Gomathi, "EAI Endorsed Transactions Re search Article Sparrow Search Algorithm-based Resource Management in Internet of Things (IoT)," pp. 1–11.
[11]. M. Bramer, Principles of Data Mining. Springer London, 2007.
[12]. Vercellis, Business intelligence: Data Mining and Optimization for Decision Making. Chichester: John Wiley & Sons, 2009.

BIBLIOGRAPHY OF AUTHORS

Rachma Oktari is a master's degree student in computer science at the President University.

Mr. Tjong Wan Sen is a master's degree computer science lecturer at the President University.