KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
VOL. 20 (1) (2023), 24-34 p-ISSN: 1693-7554, e-ISSN:2654-3990

Application of Naïve Bayes Algorithm to Analysis of Free Fatty Acid (FFA) Production Based on Fruit Freshness Level

Wahyu Supriyatin[1], **Yasman Riyanto**[2]

[1]Department of Information Systems, Faculty of Computer Science and Information Technology, Gunadarma University, West Java, 16424, Indonesia
[2]Department of Computer Systems, Faculty of Computer Science and Information Technology, Gunadarma University, West Java, 16424, Indonesia

Abstract

Cooking oil is a basic need for everyone who is used to process food ingredients. The use of cooking oil repeatedly and continuously by heating at high temperatures can increase the free fatty acid levels in the oil. The more the oil is reused, the higher the free fatty acid content. Testing the levels of FFA in oil can be done using the FFA test, because FFA can affect the selling price of CPO when it is marketed. In addition, FFA affects the levels of free fatty acids of CPO. This study aims to determine the analysis of FFA production in palm oil products based on the level of freshness of the fruit. The research was conducted by classifying data mining using the Naïve Bayes Algorithm. The Naïve Bayes algorithm was used to determine whether FFA production had an effect on fruit freshness, fruit quality and fruit soiling. The research was conducted using RapidMiner Studio 9.10 tools. The results of the research from the distribution table show that the value of the FFA attribute obtained 2 conditions, namely super conditions and normal conditions. Where each of these attributes is influenced by the variables of fruit freshness and fruit quality. Probability accuracy results from 60 training data and 40 testing data used are 92.50 percent for super FFA conditions.

Keywords: *Classification; Data Mining; Free Fatty Acid (FFA); Naive Bayes*

1. Introduction

Indonesia as one of the largest producers of palm oil in the world [1], [2]. Where the palm oil industry is one of Indonesia's mainstay products in the economic sector because it is a source of foreign exchange for exports in agriculture [1]. Oil palm plantations in Indonesia are spread over several areas such as Sumatra, Sulawesi and Kalimantan. Oil palm plantations in Indonesia are owned and managed by the state, private or foreign. The oil palm plantation sector in Indonesia in 2015 could produce more than 31 million tons of Crude Palm Oil (CPO) [1]. Oil palm is one of the industrial plants that can survive amid the threat of uncertain economic conditions. This is because the prospect of the palm oil market itself still has a high interest in the demand for oil products and other processed palm oil products [3]. Consumption of palm oil (CPO) from year to year for each country has increased, one of which is Indonesia. Indonesia as the largest CPO producing country but also as a country that absorbs a lot of the need for CPO in the world [4].

*Corresponding author. *E-mail address:* $yu_ws@staff.gunadarma.ac.id$
Received : 12*October*2022, *Accepted* : 09*December*2023*andavailableonline*31*January*2023
https : *//doi.org/*10.33751/*komputasi.v*20*i*1.6293

One of the palm oil processing companies located in Kalimantan, is one of the suppliers of cooking oil in various sectors both in Indonesia and abroad. Where in the process and processing of cooking oil comes from Crude Palm Oil (CPO) which goes through the process of degumming, bleaching, filtration and deodorization at the refining plant [3] [5]. To ensure the quality of the palm oil produced, Free Fatty Acid (FFA) testing is carried out. The Free Fatty Acid (FFA) test is carried out to find out how much free fatty acid content is contained in the resulting CPO [3]. Free Fatty Acids (FFA) is one of the main ingredients in palm oil [5]. High percentage of FFA in palm oil content can affect palm oil quality and human health [5]. The use and use of CPO, especially palm cooking oil, repeatedly and at high temperatures can result in an increase in free fatty acid levels [6], The using and use of CPO, especially on palm cooking oil, is repeated, so it is necessary to control the quality of cooking oil, especially the characteristics of FFA. So that the quality of the cooking oil produced is maintained and in accordance with the company's quality standards. The amount of FFA in palm oil will describe the level of damage from cooking oil caused by the breakdown of tryacilglycerol and fatty acid oxidation [3] , [5]. In addition, the amount of FFA in palm oil can affect the selling price of CPO in the market.

Oil palm said to be superior if it has a planting age of more than 3.5 years and has red fruit with a lot of oil content [2]. Ripe fruit will have an increased content of Free Fatty Acids (FFA) and the fruit will fall on its own according to the age of the fruit [2]. Good fruit conditions will affect the level of FFA production produced. FFA production in a level of palm oil will be influenced by several factors, including the level of freshness of the fruit (FFB Freshnees) produced, the level of dirtiness of the fruit, the quality of Fresh Fruit Bunches (FFB), age of the fruit, condition of the fruit and Fruit Stalk condition.

Knowledge Discovery in Database (KDD) is a search method in obtaining knowledge in a collection of databases in a series of processes[7] [8]. The existing knowledge in the database can be used as a knowledge base to obtain a decision[8]. The processes that occur in KDD are in the form of data selection, pre-processing/cleaning, transformation, data mining and interpretation/evaluation [8]. So it can be conveyed if data mining is part of the stages of the KDD process. Data mining is in the form of classifying, predicting, estimating and obtaining information from large data sets [9]. Data mining is a process of finding new data, new patterns and new trends by selecting some data in big data stored with statistical patterns and techniques [7] [9]. Classification is one of the data analysis used to predict samples where classification techniques have begun to be applied in various fields such as machine learning, expert systems and statistics [10]. Classification is one type of data analysis that is used to help determine the class label of the classified sample [2]. The purpose of self-classification is to find a different training set model in the appropriate class [2].

Dataset is a collection of objects and attributes in set of data [7]. The dataset consists of training data in the form of pre-existing data based on facts and testing data in the form of data that will be used as test material obtained from the training data process [7]. One of the classification algorithms in data mining is the Naïve Bayes classification algorithm. The Naïve Bayes algorithm is a method used to find probability values by adding up the frequency and combination of dataset values in an event [11]. The Naïve Bayes algorithm can predict future events with previous events already happening. The Naïve Bayes algorithm is similar to the decision tree classification algorithm and neural network, only Naïve Bayes has a higher level of accuracy and speed when applied in the database [11]. The Naïve Bayes algorithm calculates the probability of each class of existing attributes and sees which class is the most [11]. The advantage of using Naïve Bayes in the classification process is that algorithm requires two pieces of data, namely training data (training set) and testing data (testing set) in testing the data you want to find.

Research related to the Naïve Bayes Classification Algorithm was carried out by Syaputri, Irwandi and Mustakim [9]. The use of the Naïve Bayes Algorithm in research to classify the selection of majors in SMA Negeri 1 Kampar Timur. To find out the classification of majors at SMA Negeri 1 Kampar Timur, training data and test data are carried out with a ratio of 70% and 30%, respectively. From the research conducted, the results obtained with an accuracy value of 96.19%, where the Naïve Bayes Algorithm is suitable to be used to determine the direction of interest of students in SMA Negeri 1 Kampar Timur.

Research related to the Naïve Bayes algorithm was carried out by Awaludin, Yasin and Wahyuningsih [12] at the Information Systems Study Program, Air Marshal Suryadarma University. The study was conducted to predict the graduation of students in the department so that more graduates can be obtained than the previous generation. By using RapiMiner Studio 9.8 processing is carried out using several data attributes from 41 training data and 25 test data. Where the results of the study show an accuracy value of 96%, a recall value of 90.91% and a precision value of 100%.

The research related to the classification of data mining, among others, by Abdillah, Defit and Sumijan in the Proceedings of SemNas related to the performance analysis of the data mining classification method using the Naïve Bayes algorithm for the case of predicting the accuracy of student graduation at STIKES Syedza Saintika [13]. In his research that the graduation rate is a parameter to assess the effectiveness of an educational institution. There are many university databases where students' academic data is stored and managed using data mining techniques for easy decision making. The Naïve Bayes algorithm is one of the classification algorithms used to obtain probabilities related to future opportunities with previous experiences. The research was conducted at the Informatics Engineering Study Program, Padang State University in 2011 and the results of the test showed that Naïve Bayes has a good accuracy value of 93.48%. So it can be concluded that the Naïve Bayes classification algorithm has a good performance in predicting the punctuality of graduation.

Research conducted by Fitriani related to the application of data mining to predict participation in the gubernatorial election using the Naïve Bayes classification method was conducted in 2019 [14]. The research was conducted by examining election participation in the Jabon sub-district using the Naïve Bayes classification algorithm determined 10 variables as attributes. From the data obtained, it is divided into training data and testing data, so that the prediction results in the form of attribute attendance sets reach 97% correct predictions with 3% incorrect predictions.

Anggraini, Widagdo, Budi and Qomaruddin in their research in 2019 examined the application of calcification data mining on blogger data with Naïve Bayes[7]. Research related to data classification to determine whether blogger users are included in the category of professional blogger users or not. The classification technique uses descriptive and predictive models with Naïve Bayes data mining algorithms. The calculation of performance vector shows that the classification accuracy is 86.67% where the class precision and class recall are 91.30% for yes and 71.43% for no.

The purpose of this study was to analyze the production of FFA production based on the level of freshness of the fruit using the Knowledge Discovery in Database (KDD) technique. The level of fruit freshness here is Fresh Fruit Brunch (FFB) Freshness or Fresh Fruit Bunches (FFB) from an oil palm. Production analysis is carried out through a data mining classification process using the Naïve Bayes classification algorithm. Production FFA analysis in this study was carried out by looking at the attributes of FFB Freshnees, the level of dirtiness of the fruit and the quality of the fresh fruit bunches (fruit condition). FFB Freshnees used in this research are fresh condition and old bunch condition. The level of dirtiness of the fruit using the condition of the fruit of dirt and the condition of the fruit of no dirt. The quality of fresh fruit bunches (fruit condition) in this study used ripe conditions, unripe conditions, empty bunch conditions, cut/long stalk conditions, old crop conditions and overripe conditions. Implementation of the Naïve Bayes Classification Algorithm using RapidMiner Studio 9.10 tools. This research will produce a decision (goal) in the form of a probability accuracy value that affects the results of the production FFA analysis.

2. Methods

The data mining classification method used in this study is the Naïve Bayes Algorithm. It was used in this study to analyze the FFA production based on the level of fruit freshness. So it will be seen what kind of fruit conditions most affect FFA production in palm oil. Fruit conditions in this study will use three variables with each condition consisting of two to five attributes. The rules of the Naïve Bayes algorithm can be seen in the flowchart of Figure 1. The research flow used in the classification of data mining is shown in Figure 2.

The pseudocode of the Naïve Bayes Algorithm in Figure 1 which is used in obtaining fruit quality conditions that affect FFA production are:

1. Start.
2. Reading the training data to be classified.
3. Counting the number of classes or labels in the training data used. The number of classes or labels is calculated and grouped according to the number of samples used and the goal that will be generated and stated in the record.
4. Calculating the probability value of each class or label obtained using Equation 1 .

$$P(X) = \frac{H}{n} \tag{1}$$

Where X is the data class that you want to find the probability of, P(X) is the probability of class X, n is the number of all classes and H is the number of classes in the probability X.

5. Counting the number of cases with the same class (counting the number of cases per class). Do this by determining the 5 conditions of each attribute by adjusting the goals to be obtained. Equation 2 is used to count the number of cases with the same class.

$$P(H|X) = \frac{p(X|H)_{p(H)}}{p(x)} = P(X|H)xP(H)|P(X) \tag{2}$$

Where X is the data with an unknown class, H is the hypothesis of X data which is a specific class, P(H|X) is the probability of the hypothesis H based on the condition X, P(X|H) is the probability of X based on the condition of the hypothesis H, P (H) is the probability of the hypothesis H and P(X) is the probability of X.

6. Obtained the results of the calculation of each of each attribute. Next, multiply all the result values according to the X data that the class is looking for (multiplying all class variables). So that the value will be obtained according to the conditions of the requested goal.
7. Comparing the results of the class where the value with the largest class probability will show that the class is selected as the goal.
8. Finished.

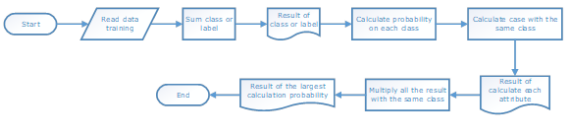


Figure 1. Flowchart of the Naïve Bayes Algorithm

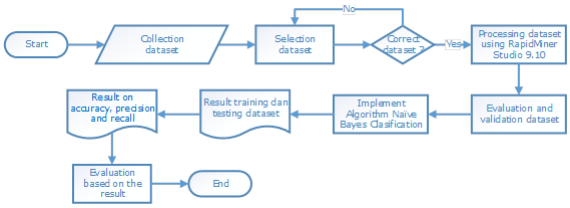
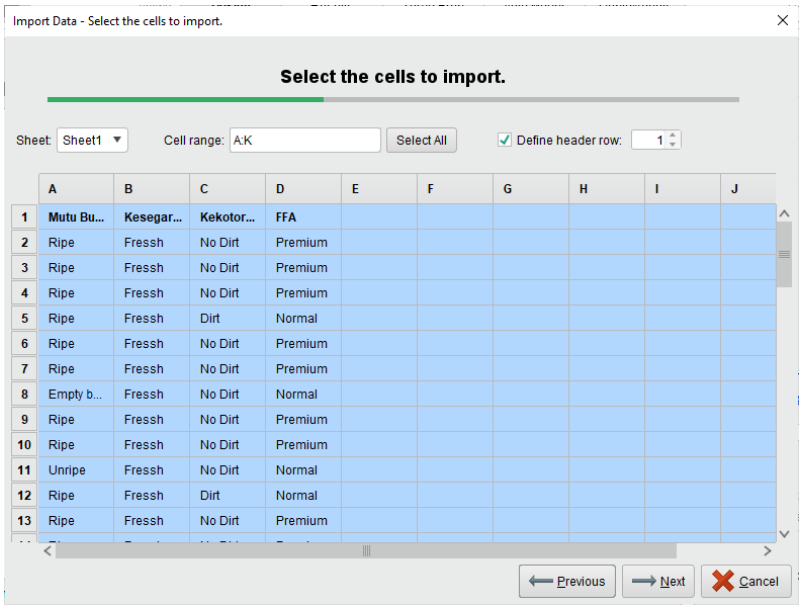


Figure 2. Research Flow

The steps in the classification of data mining using the Naïve Bayes Algorithm in this study are:

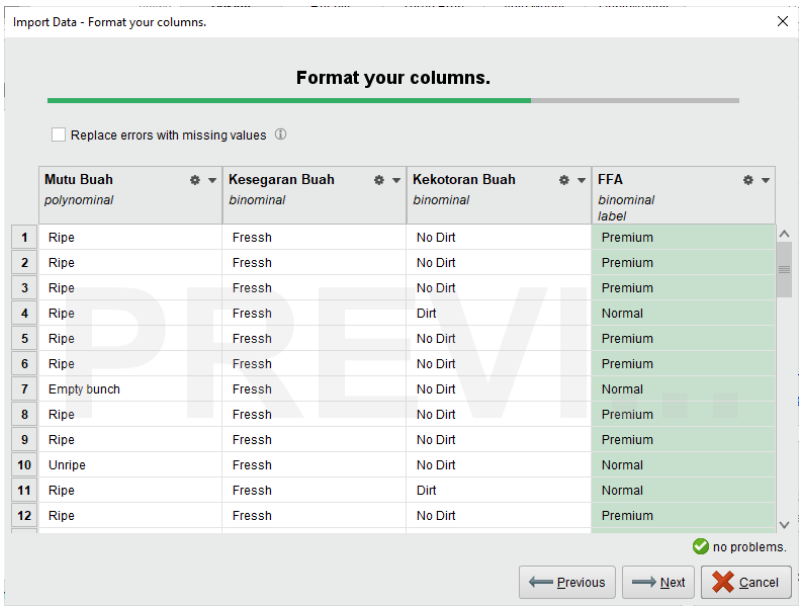
a Data Collection Process The data collection stage is the initial stage in the research before processing the data using the Naïve Bayes Algorithm. The data collection stage is carried out by taking a dataset that will be used as a sample for training data and testing data from oil palm plantation data.

b Data Processing The data processing stage is the next stage after obtaining training data and testing data. In the preprocessing stage, the dataset will go through three processes, namely removing duplication of data, checking for inconsistent data and correcting errors in the data. Where the final result of the data processing is a clean data collection or no missing value.



	A	B	C	D	E	F	G	H	I	J
1	Mutu Bu...	Kesegar...	Kekotor...	FFA						
2	Ripe	Fressh	No Dirt	Premium						
3	Ripe	Fressh	No Dirt	Premium						
4	Ripe	Fressh	No Dirt	Premium						
5	Ripe	Fressh	Dirt	Normal						
6	Ripe	Fressh	No Dirt	Premium						
7	Ripe	Fressh	No Dirt	Premium						
8	Empty b...	Fressh	No Dirt	Normal						
9	Ripe	Fressh	No Dirt	Premium						
10	Ripe	Fressh	No Dirt	Premium						
11	Unripe	Fressh	No Dirt	Normal						
12	Ripe	Fressh	Dirt	Normal						
13	Ripe	Fressh	No Dirt	Premium						

Figure 3. Import Dataset Training



	Mutu Buah *polynominal*	Kesegaran Buah *binominal*	Kekotoran Buah *binominal*	FFA *binominal label*
1	Ripe	Fressh	No Dirt	Premium
2	Ripe	Fressh	No Dirt	Premium
3	Ripe	Fressh	No Dirt	Premium
4	Ripe	Fressh	Dirt	Normal
5	Ripe	Fressh	No Dirt	Premium
6	Ripe	Fressh	No Dirt	Premium
7	Empty bunch	Fressh	No Dirt	Normal
8	Ripe	Fressh	No Dirt	Premium
9	Ripe	Fressh	No Dirt	Premium
10	Unripe	Fressh	No Dirt	Normal
11	Ripe	Fressh	Dirt	Normal
12	Ripe	Fressh	No Dirt	Premium

Figure 4. Dataset Training Preprocessing Process

c Data Evaluation and Validation Process Validation is an analysis of various models and se-

lecting a model with good predictive performance. The validation process is carried out using cross validation and the data mining stages are carried out in the cross validation process. From the cross validation process, it will then pass through the stages of the training and testing process.

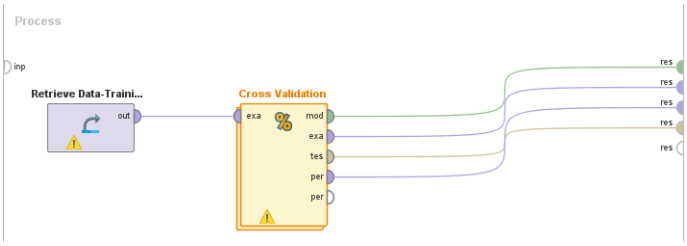


Figure 5. Stages of Dataset Validation

d Training and Testing Process Training and testing is the final stage in processing the dataset used. The training data and testing data used are processed using the nave Bayes algorithm by looking at their performance so that the accuracy value is obtained for either the recall class or the prediction class according to the specified label.

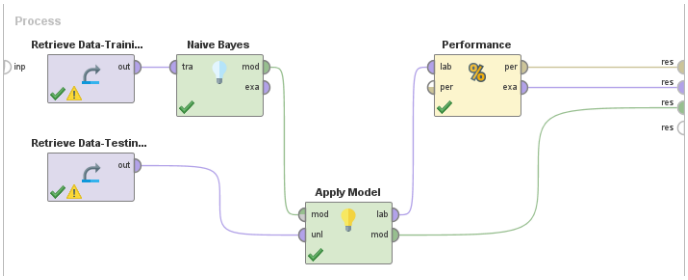


Figure 6. Training and Testing Process

3. Result and Discussion

Implementation of the application of data mining Classification Algorithms using Naïve Bayes processed with tools RapidMiner Studio 9.10. So that the probability value of each attribute with many conditions in the class/label will be obtained. The results of the probability value of the class attribute are used to see which attributes most influence the FFA production analysis based on the level of freshness of the fruit produced in oil palm plantations. The steps taken in the data collection stage until the results are obtained in the form of accuracy, precision and recall values are:

a. Data Collection Stages The stage where the dataset is collected that will be used as training data and testing data. The dataset used in this study was obtained from oil palm plantation data taken randomly for three months. The dataset used is 60 training data and 40 testing data. Table 1 is a dataset in the form of 60 training data samples that will be used. The dataset in the study resulted in a classification in the form of probability values that could affect FFA levels. The variables in the dataset consist of four attributes, namely:

   (a) Fruit Quality Variable, with attributes contents: Ripe, Unripe, Overripe, Empty Bunch, Old Crop and Under Ripe

   (b) Fruit Freshness Variable, with attribute contents: Fresh and Old

   (c) Fruit Impurity Variable, with the contents of the attributes: Dirt and No Dirt

   (d) FFA variable, with attribute contents: Premium and Normal

Table 1. Training Dataset

Mutu Buah	Kesegaran Buah	Kekotoran Buah	FFA
Ripe	Fressh	No Dirt	Premium
Ripe	Fressh	Dirt	Normal
Ripe	Fressh	No Dirt	Premium
Ripe	Fressh	No Dirt	Premium
Empty bunch	Fressh	No Dirt	Normal
Ripe	Fressh	No Dirt	Premium
Ripe	Fressh	No Dirt	Premium
Unripe	Fressh	No Dirt	Normal
.....	.....	.....	.....
.....	.....	.....	.....
.....	.....	.....	.....
Under ripe	Fressh	No Dirt	Normal
Ripe	Fressh	No Dirt	Premium
Ripe	Fressh	No Dirt	Premium
Under ripe	Fressh	No Dirt	Normal
Ripe	Fressh	No Dirt	Premium
Overripe	Fressh	No Dirt	Normal

b. Data Processing Stages The dataset in Table 1 will undergo a preprocessing stage by going through three processes, by removing duplicate data, checking for inconsistent and correcting errors. This stage is carried out to obtain a clean data set and there is no missing value for the validation and evaluation process to be carried out on the excel dataset, both training data and testing data. Figure 3 is the step of importing the training dataset Table 1 into RapidMiner Studio 9.10. From Figure 3, the preprocessing process is then carried out to check whether there is data that is problematic or not. Figure 4 shows the training dataset in Figure 3 that is appropriate and not problematic.

c. Stages of Data Evaluation and Validation The training dataset that has been preprocessed, before being used, the validation process is carried out using cross validation. The validation process is carried out by connecting the excel training data block with the cross validation block. Validation is carried out using 10 fold cross validation, which will be divided randomly into each section consisting of 10 sections that will be classified first. Measurement of accuracy using confusion matrix and Receiver Operating Characteristics (ROC) curve to measure the AUC value. Figure 5 shows the initial stages of data validation before training and testing the data. An optimistic AUC value at 0.945 and a pessimistic AUC at 0.927.

d. Stages of Training and Testing Data Data training and testing is the stage for processing 60 training data and 40 testing data that have been previously validated for the classification process with the Naïve Bayes Algorithm. The results of this testing and training will obtain the values of accuracy, precision and recall through performance vectors. Figure 6 is the stages of the process and training carried out on the dataset Table 1. The training data will be entered into the Naïve Bayes model to study the data pattern so that a model with probability values can be generated. The performance operator is used to see how smart the Naïve Bayes model that has been previously created from the apply model is and see the shape of the Naïve Bayes.

e. Performance Vector Results Naïve Bayes Classification using RapidMiner Studio 9.10 in Figure 6will produce accuracy, precision and recall values. The calculation of the performance vector in Table 2 shows the value of accuracy, precision and recall of the Classification using Naïve Bayes and obtained an accuracy value of 92.50%. It is known that from 40 testing data,

18 data were predicted to be "Premium" and actually turned out to be "Premium", 3 data were predicted to be "Normal" and in fact "Super", 19 data were predicted to be "Normal" and actually turned out to be "Normal". The results of class precision and class recall for premium predictions show a class precision value of 100% and for premium predictions with a class recall value of 85.71%.

Table 2. Training Dataset

	true Premium	**true Normal**	**class precision**
Rpred. Premium	18	0	100.00%
pred. Normal	3	19	86.36%
class recall	85.71%	100.00%	

f. Class Classification Results with Naïve Bayes Model

(a) Simple Distribution Models Figure 7 is the result of FFA data classification using the nave Bayes classification algorithm. The algorithm is divided into two classes, namely premium class and normal class. The premium class is a class with ripe, fresh and no dirt attribute values of 0.550, while the normal class is a class outside the premium class attribute with a value of 0.450.

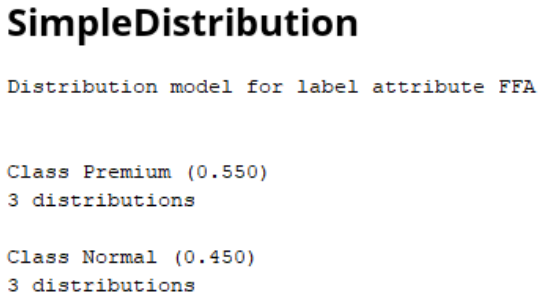


Figure 7. Simple Distribution

(b) Distribution Table Table 3 is the result of the distribution of tables generated by processing data using RapidMiner Studio 9.10 tools. Distribution in Table 3 shows the classification results in the form of Premium FFA and Normal FFA. Where the variables that affect the two classification results are fruit quality and fruit freshness.

4. Discussion

This study discusses the analysis of FFA production based on the level of freshness of the fruit by applying the principles of the Naïve Bayes Classification Algorithm. From the research conducted, the results obtained in the form of values of accuracy, precision and recall of the dataset used as the research sample. In accordance with the research flow in Figure 2, which was carried out on 60 training datasets and 40 testing datasets, an accuracy value of 92.50% on vector performance was obtained. From the results of the vector performance, it shows that the FFA analysis obtained data results in accordance with the predictions given, namely Premium predictions and Normal predictions. The results in the distribution table in Table 3 after training and testing as shown in Figure 6 show that the analysis of the production FFA is influenced by two attributes, namely the fruit quality attribute and the fruit freshness attribute.

Table 3. Distribution Table

Attribute	Parameter	Premium	Normal
Mutu Buah	value=Ripe	0.997	0.296
Mutu Buah	value=Empty bunch	0.001	0.074
Mutu Buah	value=Unripe	0.001	0.296
Mutu Buah	value=Old crop	0.001	0.074
Mutu Buah	value=Overripe	0.001	0.111
Mutu Buah	value=Under ripe	0.001	0.148
Mutu Buah	value=unknown	0.001	0.001
Kesegaran Buah	value=Fressh	0.999	0.777
Kesegaran Buah	value=Old	0.001	0.222
Kesegaran Buah	value=unknown	0.001	0.001
Kesegaran Buah	value=No Dirt	0.999	0.814
Kesegaran Buah	value=Dirt	0.001	0.185

In the processing of FFB at the Palm Oil Mill, there are many factors that affect FFA CPO Production. However, one of the factors that most influence FFA CPO production from FFB processing is fruit quality and fruit freshness. Where the factors of fruit quality and fruit freshness are the main factors that must be considered for an estate in making deliveries to the mill. If the quality of the fruit and the freshness of the fruit that was sent arrived and was well received by the mill, but the results of the FFA CPO Production did not match the results of the analysis, then there are other external factors that affect the quality of the fruit and freshness of the fruit. So the mill must check to see what other things affect the unsuitable FFA CPO production. The mill can check every tool used and other materials used during the FFB process.

5. Conclusion

Based on the research conducted, it can be concluded that the classification of the FFA dataset using the Naïve Bayes algorithm obtains appropriate results to determine how much influence it has on FFA for Premium and Normal conditions. Data processing is carried out using RapidMiner Studio 9.10 tools and from the results obtained a classification accuracy value of 92.50%. Dataset processing was carried out on 60 training data samples with 40 testing data samples. From the performance results, the accuracy shows that from 40 testing data, 18 data were obtained which were predicted to be FFA "Premium" and actually obtained FFA "Premium" results, 3 data were predicted to be FFA "Normal" and turned out to be FFA "Premium", 19 data were predicted to be FFA “Normal” and turned out to be a “Normal” FFA. The results of the distribution table show that the results of the Premium FFA and Normal FFA classifications are influenced by the variables of fruit quality and fruit freshness.

The next development related to this research is that other data mining classification methods can be used according to the needs at the location of oil palm plantations, such as the C5.0 Algorithm, C4.5 Algorithm, ID3 Algorithm, Classification and Regreesion Trees Algorithm or KNN Algorithm. The development of the use of other algorithms is used to see how the relationship between the influencing variables is so that it can see the probability value of accuracy or the value of the decision tree or the value of the gain that affects it. In addition, in terms of algorithms, the attributes used can be developed to determine the selling price of CPO and improve harvest quality based on plant age and harvesting round.

References

[1] Sudrajat ,Kepala Sawit : Prosperk Pengembangan dan Peningkatan Produktivitas, Bogor: Penerbit IPB Press, 2020.

[2] W.SUpriyatin , "Palm Oil Extraction Rate Prediction Based on the Fruit Ripeness Level Using C4.5 Algorithm," ILKOM Jurnal Ilmiah, vol. 13, no. 2, pp. 92-100, 2021. ,

[3] R. L. R. Silalahi, D. P. Sari and I. A. Dewi, "Testing of Free Fatty Acid (FFA) and Colour for Controlling the Quality of Cooking Oil Produced by PT. XYZ," Industria : Jurnal Teknologi dan Manajemen Argoindustri, vol. 6, no. 1, pp. 41-50, 2017. "

[4] Masykur, "Pengembangan Industri Kelapa Sawit Sebagai Penghasil Energi Bahan Bakar Alternatif dan Mengurangi Pemanasan Global (Studi di Riau sebagai Penghasil Kelapa Sawit Terbesar di Indonesia)," Jurnal Reformasi, vol. 3, no. 2, pp. 96-107, 2013. "

[5] Suprijadi,F. Faizal, and R. R. Septiawan, "Computational Study on Melting Process Using Smoothed Particle Hydrodynamics," Journal of Modern Physics, vol. 05, no. 03, pp. 112–116, 2014. [Online]. Available: https://www.scirp.org/pdf/JMP2014022411463120.pdf.

[6] Febrianto, A. Setianingsih and A. Riyani, "Determination of Free Fatty Acid in Frying Oils of Various Foodstuffs," Indonesian Journal of Chemistry and Environment, vol. 2, no. 1, pp. 1-6, 2019."

[7] R. A. Anggraini, G. Widgado, A. S. Budi and M. Qomaruddi, "Penerapan Data Mining Clasification untuk Data Blogger Menggunakan Metode Naive Bayes," JUSTIN (Jurnal Sistem dan Teknologi Informasi), vol. 7, no. 1, pp. 47-51, 2019.

[8] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," Jurnal Edik Informatika, vol. 2, no. 2, pp. 213-219, 2016.

[9] A. W. Syaputri, E. Irwandi and Mustakim, "Naive Bayes Algorithm for Classification of Student Major's Specialization," Journal of Intelligent Computing and Health Informatics, vol. 1, no. 1, pp. 15-19, 2020.

[10] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," Jurnal Teknik Informatika Kaputama (JTIK), vol. 3, no. 1, pp. 22-28, 2019.

[11] Normah , "Naive Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews," Journal Publications Informatics Engineering Research, vol. 3, no. 2, pp. 1-19, 2019.

[12] M. Awaludin, V. Yasin and M. Wahyuningsih, "Optimization of Naive Bayes Algorithm Parameters For Student Graduation Prediction At Universitas Dirgantara Marsekal Suryadarma," Journal of Information System Informatics and Computing, vol. 6, no. 1, pp. 91-106, 2022.

[13] N. Abdillah, S. Defit and Sumijan, "Analisis Kinerja Metode Klasifikasi Data Mining Menggunakan Algoritma Naive Bayes Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa (Studi Kasus STIKES Syedza Saintika)," in Prosiding Seminar Nasional STKIS Syedza Saintika, Padang, 2021.

[14] A. S. Fitriani , "Penerapan Data Mining Menggunakan Metode Klasifikasi Naive Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur," JTAM (Jurnal Teori dan Aplikasi Matematika), vol. 3, no. 2, pp. 98-104, 2019.

[15] N. Iriadi and N. Nuraeni, "Kajian Penerapan Metode Klasifikasi data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta," Jurnal Teknik Komputer AMIK BSI, vol. II, no. 1, pp. 132-137, 2016.

[16] N. A. Widiastuti, S. Santosa and C. Supriyanto, "Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Particile Swarm Optimization untuk Deteksi Penyakit Jantung," Jurnal Pseudocode, vol. 1, no. 1, pp. 11-14, 2014.

[17] D. Berrar, "Bayes' Theorem and Naive Bayes Classifier," January 2018. [Online]. Available: https://www.researchgate.net/publication/324933572Bayes%27 TheoremandNaiveBayesClassifier?enrichId=rgreq-a8005eeb8ea85a3d1 724c2b7ccc23e07-XXXenrichSource=Y292ZXJQYWdlOzMyNDkzMzU3MjtBUzo4MDQ4Mjk1 MzQzNzU5MzZAMTU2ODg5NzcyMjA3Mg%3D%3Del=1x2esc=publ. [Accessed 07 November 2022]

[18] D. Jurafsky and J. H. Martin, Speech and Language Processing, New Jersey: Pearson Prentice Hall, 2021.

[19] T. M. Mitchell , Machine Learning, McGraw-Hill Science, 2017.

[20] RC. M. V. Polette , P. R. Ramos, C. B. Goncalves and A. L. D. Oliveira, "Determination of fress fatty acids in crude vegetable oil samples obtained by high-pressure processes," Food Chemistry: X, vol. 12, pp. 1-8, 2021.