Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

Application of the K-Nearest Neighbor Machine Learning Algorithm to Predict Sales of Best-Selling Products

Muhtajuddin Danny[1*], Asep Muhidin[2], Akhiratul Jamal[3]
[1,2,3]Informatics Engineering Study Program, Faculty of Engineering, Pelita Bangsa University, Indonesia
[1]utat@pelitabangsa.ac.id, [2]asep.muhidin@pelitabangsa.ac.id, [3]akhjaml@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 12-06-2024
Accepted: 13-06-2024
Published: 28-06-2024

Keywords:
k-nearest neighbor algorithm, machine learning, sales.

ABSTRACT

The development of increasingly intense competition in the business world, accompanied by advances in information technology, has brought retail companies into a situation of tighter and more open competition. PT LG Innotek Indonesia is the only company that produces tuners in Indonesia. Looking at consumer demand, PT LG Innotek must improve product quality, and add products that consumers like and frequently purchase. For this reason, PT LG Innotek Indonesia needs an analysis that can help the company identify products that tend to sell well. This analysis can be carried out through the application of machine learning algorithms, especially the K-Nearest Neighbor method. The aim of this research is to find out how the KNN algorithm performs in predicting products that are selling well and not selling well at PT LG Innotek Indonesia. Based on the analysis results, prediction results were obtained with an accuracy level of 94.74% and an error rate of 5.26%. With this high level of accuracy and low error rate, it can be concluded that the K-Nearest Neighbor method is effectively used to predict sales of PT LG Innotek Indonesia's best-selling products.

INTRODUCTION

The development of increasingly intense competition in the business world accompanied by advances in information technology, including in Indonesia, has brought retail companies into a situation of tighter and more open competition (A. A. Leasiwal et al., 2021). In facing competition to increase revenue, these companies need to design marketing strategies for the products they sell. Strategic decisions must be taken appropriately by considering current market conditions. To meet consumer preferences, related parties must focus on improving product quality, adding product types that are in demand by consumers, and selling products (Rosidi & Setiawan, 2024). One retail company, namely PT. LG Innotek Indonesia, which is a company operating in the manufacturing industry by producing electronic components. This company is the only company that produces tuners in Indonesia.

To increase sales effectiveness, predictions or forecasting are needed (Herlambang et al., 2023) Forecasting is the process of predicting future sales with the aim of determining the estimated sales volume and identifying potential markets that will be dominated in the future (Mhd Angga Sabda & Suhardi Suhardi, 2023). Apart from that, forecasting also plays an important role in determining stock availability plans (Ari Sanjaya & Tri Wahyana, 2022). By utilizing these predictions, sales output can be predicted better so as to reduce errors in planning as efficiently as possible (Dewi et al., 2022a). This is important to ensure the right direction in a business's marketing strategy, because the need for accurate information through sales data is very important. Efforts to obtain the latest unknown information from data sets are known as Data Mining. In Data Mining various techniques are used to reveal information and patterns in data, one of which is the K-Nearest Neighbor (K-NN) classification technique(Azis et al., 2024a). The K-NN algorithm is used to group objects by comparing training data with the closest distance to the object being tested (Rozzi Kesuma Dinata & Novia Hasdyna, 2020). The working principle of K-NN is to compare training data with test data to find training data that is closest to the data being tested (Nolly et al., 2023).

This company sells various products, some of which include RF Tuners, WiFi, and Power modules such as LED lights, LIPS, SMPS, Trans, Adapters, and so on. Looking at consumer demand, the LG Innotek company must improve product quality, and add products that consumers like and frequently purchase. In this context, the LG Innotek company needs analysis to predict products that will be of interest to consumers (Febriana Santi Wahyuni, 2024). The required analysis must be able to help companies identify products that tend to sell well by predicting sales data from previous years through the application of machine learning algorithms (Harahap et al., 2023). This machine learning approach is based on exploiting data to develop statistical models, which are then utilized by the system to make predictions about the future based on previous input data or to understand patterns in the data (Wardani et al., 2024). Machine learning is part of artificial intelligence, which involves programming so that computers are able to behave intelligently like humans and can improve their understanding through learning from experience automatically (Nur Fajri et al., 2022). One machine learning algorithm that can predict the most popular products is K-Nearest Neighbor (KNN). The K-

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

Nearest Neighbor (KNN) approach is a classification technique that uses the distance between certain data and other data. KNN is a type of supervised learning algorithm where the results of querying new instances are grouped based on the majority of categories in KNN. The class that appears most frequently among the nearest neighbors is the classification result. Implementing the K-Nearest Neighbor Algorithm can make things easier for PT. LG Innotek Indonesia in identifying best-selling products based on number of sales (Dewi et al., 2022b).

LITERATURE REVIEW

Implementation of Data Mining for Car Sales Using the Naive Bayes Method (Rifky et al., 2022) is research that uses the Naive Bayes Method, namely the classification and branching method of artificial intelligence. Artificial intelligence is the system's ability to interpret external data correctly, to learn from the data and to use this learning to achieve goals (Sari & Hayuningtyas, 2019).

Therefore, the various brands in this research will be formed into a class, namely Best Selling and Not Best Selling, so that consumers, producers and researchers can find out which car brands are the best sellers based on category and output (Rizki et al., 2020). Implementation of the Naive Bayes method can produce maximum accuracy with little training data.

Naive Bayes is a method suitable for binary and multiclass classification. This method, which is also known as the Naive Bayes Classifier, applies supervised object classification techniques in the future by assigning class labels to instances/records using conditional probability. Conditional probability is a measure of the chance of an event occurring based on other events that have (by assumption, presumption, statement, or proven) occurred (Prastiwi et al., 2022).

Based on research on predicting car sales, several conclusions can be drawn, namely the Naive Bayes method utilizes training data to produce the probability of each criterion for different classes, so that the probability values of these criteria can be optimized to predict car sales based on the classification process carried out by the Naive Bayes method itself. Based on car sales data used as training data, the Naive Bayes method succeeded in classifying 7 data out of 10 data tested. So the Naive Bayes method is successful in predicting the size of car sales with an accuracy percentage of 70%

Application of Data Mining to Predict Sales of Best-Selling Bread Products at PT. Nippon Indosari Corpindo Tbk uses the K-Nearest Neighbor method (Ike Yolanda & Hasanul Fahmi, 2021). PT. Nippon sells 20 types of bread products that are popular among Indonesian people. Seeing the large demand from consumers, this company needs to carry out sales predictions to determine which bakery products are selling best so that the production process can be managed more efficiently. The application of data mining to predict sales of best-selling bread products has become a system that helps PT. Nippon Indosari Corpindo Tbk in predicting sales of bakery products with high accuracy and efficiency. This is done by considering the criteria for product quantity and quantity sold. The data used as the basis for the research comes from sales information for the last three months held by PT. Nippon. In this research, the KNN algorithm is used to predict the types of bread that are best-selling by Indonesian people. The application of the KNN algorithm in a data mining framework becomes an application that can produce sales predictions for best-selling bakery products by adjusting criteria and using predetermined weights.

Furthermore, Sri Puspita Dewi, Nurwati, Elly Rahayu in their research "Application of Data Mining to Predict Sales of Best Selling Products Using the K-Nearest Neighbor Method"(Dewi et al., 2022b). The location of this research was UD Andar. This is because UD Andar sells various types of products, such as plastic bags, herbal medicine powder, food and drink ingredients, frozen food, and so on. To increase efficiency in planning stock inventory, UD Andar needs predictive analysis to identify products that have the highest sales levels. By considering these needs, a data mining approach by applying the K-Nearest Neighbor (KNN) algorithm can be used to determine the products with the highest sales based on sales data for the last year at UD Andar. Thus, the KNN algorithm can act as a decision support tool for companies in determining the best-selling products.

Research conducted by Rismala, Irfan Ali, Ade Rizki Rinaldi with the journal title "Application of the K-Nearest Neighbor Method to Predict Best-Selling Motorcycle Sales"(Azis et al., 2024b). This research was carried out because of the increasing demand for means of transportation, especially motorbikes, as well as many manufacturers introducing products with various brands and designs. To increase product sales, predictive analysis is needed to find out which motorbike products are selling best and which are not selling well. This research uses data mining techniques by applying the K-Nearest Neighbor algorithm to identify the best-selling and non-selling motorbikes. The research was carried out at PT. Sumber Rejeki Jabar during the period January to December 2022. The trial results show that the KNN algorithm can be effectively used in classifying motorbike sales data at PT. West Java's Source of Fortune. Based on testing, an accuracy rate of 96.15% was obtained, indicating that the dataset can be considered valid for use at the next stage. With this high level of accuracy, the K-NN model is a potential solution for forecasting motorbike sales based on previous sales data. Therefore, PT. Sumber Rejeki Jabar can utilize this model to support decision making and plan more effective marketing strategies.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

METHOD

Knowledge Discovery in Database (KDD) is an important process in data mining that aims to discover useful knowledge from big data. By following systematic steps from data selection to interpretation of results, KDD helps uncover insights that can support better decision making in various domains.

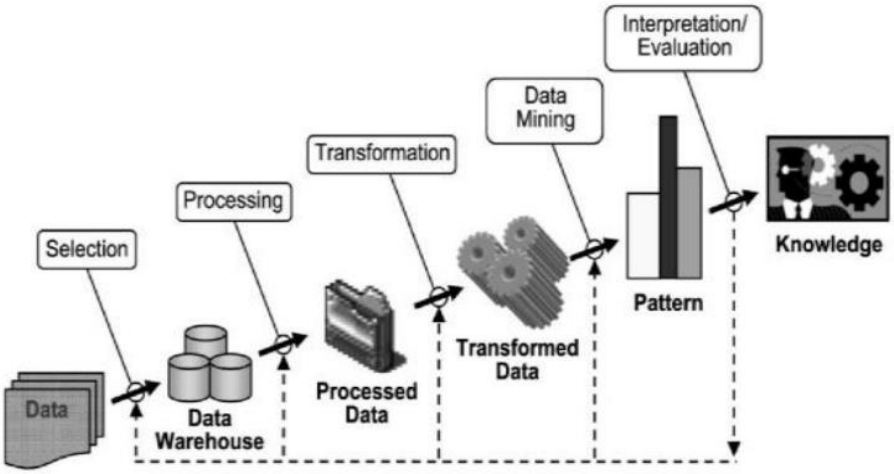


Figure 1. Knowledge Discovery Process in Database

Data Selection

The data used in this research is a dataset obtained from PT. LG Innotek Indonesia, which focuses on the number of sales of the company's products from January 2020 to December 2023. The next step in using KDD to dig up information is to select or select the required data from a large data set. The selected data will be used in the data mining process.

Table 1. Initial data before selection

month	OU	ORG	Global Customer	Customer	Model Code	Currency	2022-11 Price	2023-01 Qty	2023-01 Amt ($)	2023-01 Amt(KRW)
Jan-23	LGITKR	WF1	STARION	ctronics Inc., KoreaWF	ETWCFMBC02.IS20	USD	5,276	6,200	26,226	32,710,379
Jan-23	LGITKR	WF1	STARION	ctronics Inc., KoreaWF	ETWCERBC01.IS00	USD	5,095	22,550	92,117	114,892,616
Jan-23	LGITKR	WF1	STARION	ctronics Inc., KoreaWF	ETWCFRBC01.IS10	USD	5,099	4,100	16,761	20,904,908
Jan-23	LGITKR	WF1	STARION	ctronics Inc., KoreaWF	ETWCFLBC02.IS10	USD	5,436	400	1,743	2,174,256
Jan-23	LGITKR	WF1	STARION	ctronics Inc., KoreaWF	ETWCFMBC01.IS20	USD	6,151	6,600	32,551	40,599,484
Jan-23	LGITKR	WF1	»õ·Î´Ð½º	ctronics Inc., KoreaWF	TWCM-K505D(A)-F.IXM	USD	5,990	2,880	13,967	17,250,530
...	...	...	...	...	...	...	...	...	...	...
Dec-23	LGITKR	WF1	LGE	ynosa S.A.DE C.V.)WF	ETWCFMBC01.IS20	USD	6,228	12,251	58,615	76,304,890
Dec-23	LGITKR	WF1	LGE	ynosa S.A.DE C.V.)WF	ETWCHMBC01.IS20	USD	11,972	1,500	13,577	17,958,232
Dec-23	LGITKR	WF1	Medtronic	MedtronicWF	ETGBBTBP01.IS00	USD	10,438	4,608	36,864	48,096,461
Dec-23	LGITIN	WF2	DIRI PRATAMA	ARISAWF	TWFB-R301D-F.IXS	IDR	2,953	2,035	4,608	6,009,118
Dec-23	LGITIN	TU2	Sunjet	Components Corp.TU	TDQS-A701F.IXS	USD	1,011	10,333	8,008	10,442,376
Dec-23	LGITIN	TU2	Sunjet	Components Corp.TU	TDSY-H480F(L).IXS	USD	1,127	1,487	1,285	1,675,314
Dec-23	LGITIN	TU2	Sunjet	Components Corp.TU	TDSY-G430D.IXS	USD	1,033	2,902	2,298	2,997,042
Dec-23	LGITIN	TU2	Sunjet	Components Corp.TU	TDSY-G480D.IXS	USD	1,033	16,639	13,178	17,183,966
Dec-23	LGITIN	WF2	Sunjet	Components Corp.WF	ETWCARIC03.IS00	USD	4,088	1,960	6,145	8,012,436
Dec-23	LGITKR	WF1	Vivint	vivintWF	ETPFFRPP01.IS00	USD	35,693	43,200	1,186,272	1,541,916,346
Dec-23	LGITKR	WF1	Sunjet	Components Corp.WF	TWFB-R301D-F.IXS	USD	3,204	44,100	107,472	141,298,060

Initial data before processing in variable selection from PT LG Innotek Indonesia sales data from January 2023 to December 2023.

Table 2. Selection of Variables

Variable	Indicator	Indicator	Usage Details
X1	month	V	Used
X2	OU	X	-
X3	Global Costumer	X	-
X4	Customer	X	-
X5	Model Code	V	Used
X6	Currency	X	-

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

X7	2022-11 Price	X	-
X8	2023-01 Qty	V	Used
X9	2023-01 Amt ($)	X	-
X10	2023-01 Amt(KRW)	X	-

The table explains the variables that will and will not be used in this research. The "V" indicator indicates that the variable will be used, while the "X" indicator indicates that the variable will not be used or eliminated at the criteria determination stage. Removal of some variables is based on relatively similar model values and has no impact on the results of the assessment process.

Table 3. Data Selection

month	Model Code	2023-01 Qty
Jan-23	ETWCFMBC02.IS20	6200
Jan-23	ETWCERBC01.IS00	22550
Jan-23	ETWCFRBC01.IS10	4100
Jan-23	ETWCFLBC02.IS10	400
Jan-23	ETWCFMBC01.IS20	6600
Jan-23	WCM-K505D(A)-F.IX	2880
Jan-23	ETWCFMBC03.IS00	50
Jan-23	ETWCERBC01.IS00	32000
Jan-23	ETWCFMBC01.IS20	28800
Jan-23	ETWCFMBC02.IS20	12480
Jan-23	ETWCFRBC01.IS10	5300
…	…	…
Dec-23	TDSY-H480F(L).IXS	1487
Dec-23	TDSY-G430D.IXS	2902
Dec-23	TDSY-G480D.IXS	16639
Dec-23	ETWCARIC03.IS00	1960
Dec-23	ETPFFRPP01.IS00	43200
Dec-23	TWFB-R301D-F.IXS	44100

The calculated data is sorted based on distance from closest to furthest (ascending). Dataset that has been selected and will be used for the data mining process. The amount of data processed for testing was 191 data.

Data Preprocessing

To ensure the quality and accuracy of stored data, this step involves deleting incomplete or invalid data. This preprocessing stage involves cleaning missing value data, which refers to data that is inconsistent or empty.
Here is the process:

a. Determine the maximum and minimum values from the amount data contained in the table above.
   Maximum number of data = 5635621
   Minimum number of data = 5
b. Calculating (maximum amount – minimum amount) = 5635621 – 5 = 5635616
c. Calculate the Range with (max – min results) divided by the number of categories, Range = (max – min results) ÷ 2 = 5635616 ÷ 2 = 2817808.

Transformation

Based on the range value, the sales amount that is less than or equal to ≤2817808 is categorized as Not Selling, while the sales amount >2817808 is categorized as Selling. The dataset after transformation can be seen in Table 4 as follows:

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

Table 4. Sales Data After Transformation

No	Model Code	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	Septembe	Oktober	Novembe	Desembe	Keterangan
1	ERDACBDA01.IS00	10785	6052	9348	4393	4081	2945	4350	3080	2708	3500	2803	362	Tidak Laris
2	ERDACCDA00.IS00	803	1659	637	629	1605	1099	2149	3133	2432	3545	3224	790	Tidak Laris
3	ERDAGAKC01.IM00	261649	305468	266113	107654	130192	55886	4445	6036	5033	6007	781	0	Tidak Laris
4	ERDAGAKC01.IM10	0	0	0	0	52724	97348	64896	74952	0	40000	0	0	Tidak Laris
5	ERDAGAKC01.IS00	9571	161	55	0	0	0	0	0	384	1832	6637	4567	Tidak Laris
6	ERDAGAKC01.IS10	0	0	0	0	0	0	0	0	0	7264	15200	0	Tidak Laris
7	ERDAGAKC02.IS00	0	0	0	6000	0	229	0	0	0	0	0	0	Tidak Laris
8	ERDAGAKC03.IS00	0	18669	0	1607	0	0	0	0	0	0	0	0	Tidak Laris
9	ERDAGAKC04.IS00	0	0	0	0	0	0	0	172	0	0	0	3	Tidak Laris
10	ERDAGAKD01.IS00	1248	0	0	0	0	0	0	0	0	0	0	0	Tidak Laris
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
184	TWFB-R302D-F.IXM	0	0	0	0	0	0	0	0	0	0	0	100	Tidak Laris
185	TWFB-R321D-F.IXS	6260	3005	4097	2657	1000	980	1045	0	190	1250	960	0	Tidak Laris
186	TWFM-B006D(G)-F.IXS	240	0	0	0	0	0	0	0	0	0	0	0	Tidak Laris
187	TWFM-K008D(T)-F.IXM	0	0	0	0	0	0	0	1440	0	0	0	0	Tidak Laris
188	TWFM-K008D(T)-F.IXS	1593	6273	952	1047	0	1894	5457	2901	900	780	913	5040	Tidak Laris
189	TWFM-K304D-F.IXS	1340	1640	2768	5825	5945	2137	2611	6404	5096	2503	0	0	Tidak Laris
190	TWMU-Z001D-F.KXX	3	0	0	2	2	1	2	1	0	0	0	0	Tidak Laris
191	TWZU-V320D-F.IXS	880	1150	1645	1702	114	41	0	0	7	630	0	0	Tidak Laris

The data that is ready will then be classified into two data, namely training data and testing data. The division of training data and testing data, namely, the proportion of training data and testing data is 90%:10%, with each amount of data being 172 training data and 19 testing data.

Data Mining (K-Nearest Neighbor Algorithm)

The mining process used in this research is to predict sales of PT products. LG Innotek Indonesia is a K-Nearest Neighbor algorithm. Following are the steps of the K-Nearest Neighbor algorithm:

1. Set the parameter value k, which represents the number of neighbors to consider.
2. Calculate the squared distance using the Euclidean method between the object and the training data that has been provided.
3. Sort the objects into groups with the smallest distance.
4. Collect the Y category, which represents the nearest neighbor classification based on the predetermined k value.
5. Determine the classification results by looking for the majority label from the group.

Evaluation

Method testing is carried out to analyze calculation results and evaluate the performance of the function. After carrying out manual calculations, the data was tested using the RapidMiner tool to ensure the calculation results were appropriate in the best-selling products. Test results will be validated and evaluated. The evaluation in this study aims to assess the use of the K-Nearest Neighbor algorithm in predicting best-selling products. This evaluation includes measurements of accuracy, precision, recall, and other values that determine the performance of the algorithm.

Data resulting from the data mining process will be identified to reveal patterns that will be entered into a knowledge base and then analyzed. To evaluate the performance of the K-Nearest Neighbor algorithm, the confusion matrix method is used to calculate the accuracy and error rate. Furthermore, interpretation is carried out through visualization and presentation of information regarding the methods used to obtain knowledge or information that has been revealed by the user.

RESULT

Datasets

In this study, the number of datasets used was 191 data. PT. Sales product data. LG Innotek Indonesia. The amount of data obtained from this process was 191 data, consisting of the attribute "Sold in Selling" with 6 data, and "Not Selling" with 185 data.

K-Nearest Neighbor Calculation

a. Determine the value of k. Determining the k value is generally a minimum of 1 until the training data limit is reduced by one, based on the amount of training data in Table 7, the k value that can be used is from 1 - 179. In this study the k value used is 5.
b. Calculate the distance between training data and PT sales testing data. LG Innotek Indonesia. Because the testing data consists of 19 data, the calculations shown are only on 1 testing data. Distance calculations use the Euclidean Distance formula.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

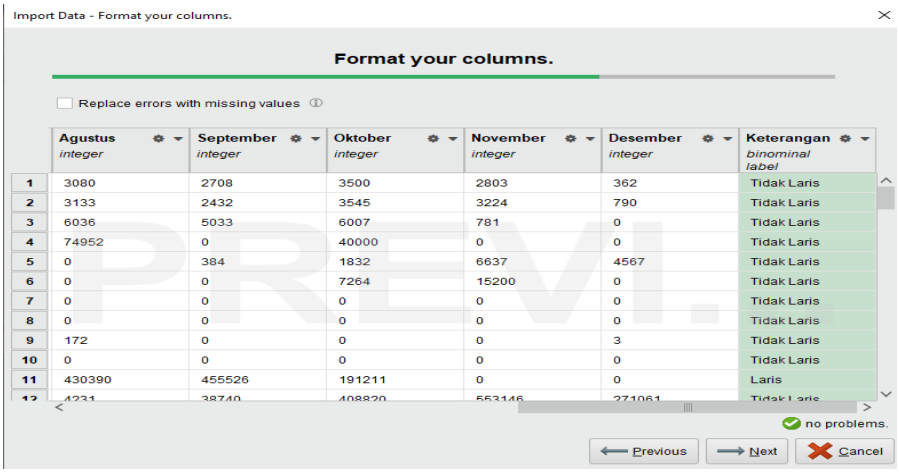
c. Sort the distances calculated from Euncliden Distance from smallest to largest.
d. Collect the Y category, which represents the nearest neighbor classification based on the predetermined k value.
e. Determine the classification results by looking for the majority label from the group. Based on Table 8, the prediction results from the calculation of Data Testing-1 with the model code ERDACCDA00.IS00, where the prediction results show that the largest number of neighbors that appear is in the Not Selling category.

Implementation of the K-Nearest Neighbor Method in RapidMiner

After carrying out manual calculations, next proof is using RapidMiner version 10.3. Testing was carried out using the K-Nearest Neighbor algorithm and the data used was 172 training data and 19 testing data. Following are the testing stages:

a. Data Import Process

At this stage, data on sales results of PT. LG Innotek Indonesia which has been classified, integration and selection imported into the RapidMiner 10.3 tool can be seen as follows:



	Agustus integer	September integer	Oktober integer	November integer	Desember integer	Keterangan binominal label
1	3080	2708	3500	2803	362	Tidak Laris
2	3133	2432	3545	3224	790	Tidak Laris
3	6036	5033	6007	781	0	Tidak Laris
4	74952	0	40000	0	0	Tidak Laris
5	0	384	1832	6637	4567	Tidak Laris
6	0	0	7264	15200	0	Tidak Laris
7	0	0	0	0	0	Tidak Laris
8	0	0	0	0	0	Tidak Laris
9	172	0	0	0	3	Tidak Laris
10	0	0	0	0	0	Tidak Laris
11	430390	455526	191211	0	0	Laris

Figure 2. Import Dataset in RapidMiner

b. Testing Process

1) Design

In the Design view, enter the data that was imported previously, the Split Data and Apply operators, and the Performance Model operator then connect the cables as shown below:

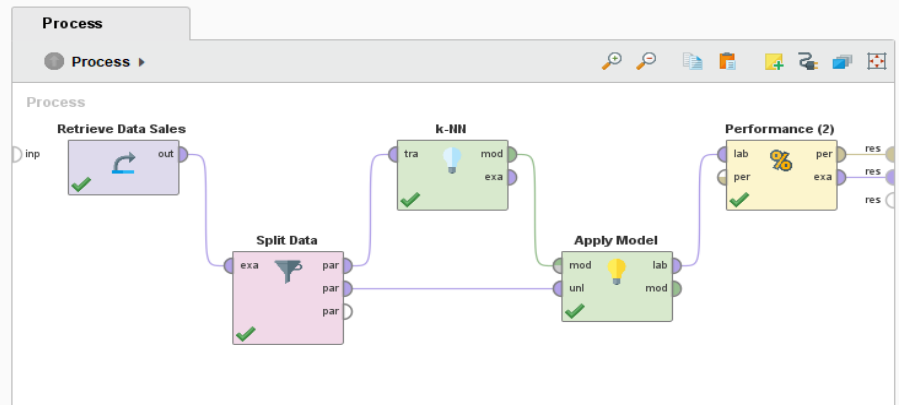


Figure 3. Initial Testing Process

2) Data Split Operator

Split Data Operator to divide a dataset into different subsets for use in the data analysis process. The proportions used in dividing training data and testing data are 0.9 and 0.1 respectively.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

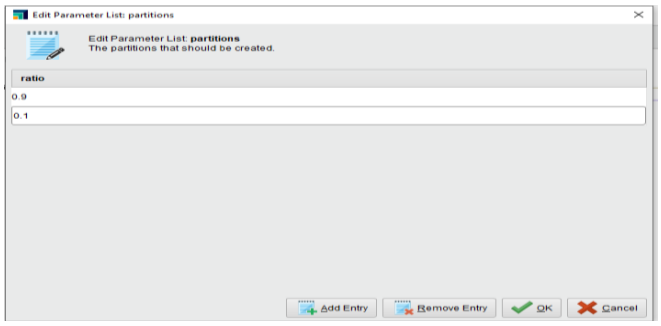


Figure 4. Split Data

3) K-NN Operator
The k-NN operator for building classification or regression models based on the K-Nearest neighbor algorithm. In this operator, the k value will be set to 15.

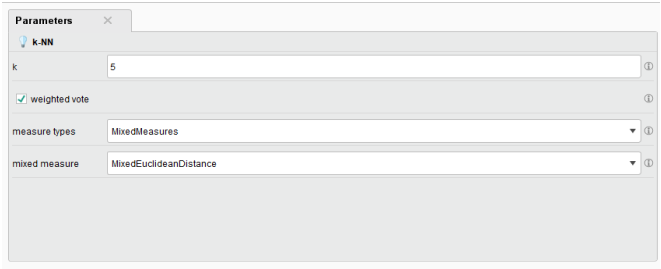


Figure 5. k-NN operator

4) Operator Apply Model
At this stage the apply model operator applies a trained model to an ExampleSet, namely a model is first trained in an ExampleSet, information related to the ExampleSet is learned by the model. Then the model can be applied to other ExampleSet to produce predictions from decision labels.
5) Operator Performance
The "Performance (Classification)" operator is used to evaluate the performance of the classification model and works in predicting classes from the dataset. In this operator, the Accuracy and Error values will be displayed.

DISCUSSION

The analysis was carried out descriptively in order to obtain an overview of the sales predictions for PT's best-selling products. LG Innotek Indonesia with various stages that are usually carried out in the decision to accept pencak silat athletes. Based on the research results that have been obtained, accompanied by existing data, the author will next carry out an analysis of the research results that have been described previously. To find out from the data analysis, you can find out by calculating the percentage of the analysis data.

Apply Model Results

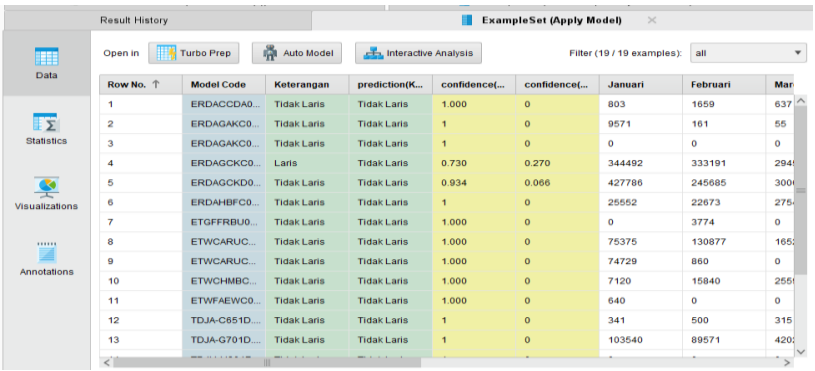


Figure 6. Results of Apply Model

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

It can be seen in Figure 6, testing1 data obtained predicted results. From the 18 testing data records that were read, it resulted in a predicted "Sold Out" decision of 0 data and a "Not Selling" decision of 19 data.

Performance Vectors

Figure 7. Performance Vector results

Figure 7 explains the confusion Matrix which provides details about the model performance by comparing the model predictions to the actual values. As for the classification error in this case, 5.26% means that of all the predictions made by the model, 5.26% of them are wrong

Performance Confusion Matrix

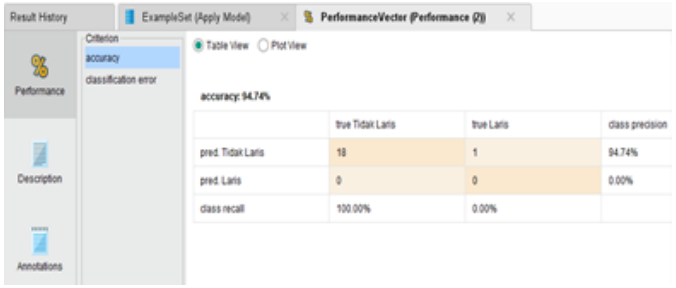


Figure 8. Accuracy Level Performance

Figure 8 shows that the model has a high accuracy of 94.74%, which shows that most of the model predictions are correct. The confusion matrix shows that there are no “Laris” cases in this dataset, indicating a significant class imbalance. The model tends to classify all cases as “Not Selling”.

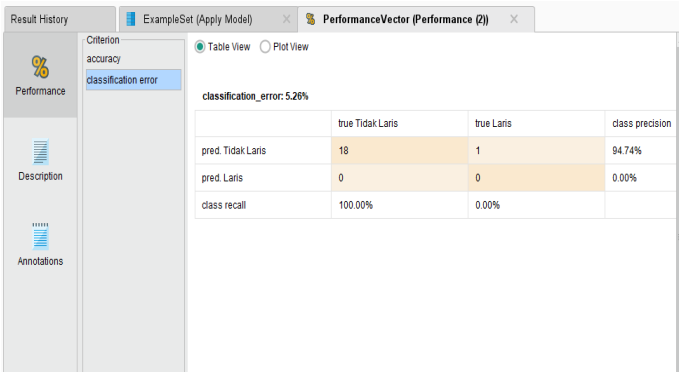


Figure 9. Error Rate Performance

In Figure 9 it can be seen that the model has a high accuracy of 94.74%, which shows that most of the predictions are correct. Low Classification Error. The classification error was 5.26%, meaning that only a small portion of predictions were wrong.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$
$$= \frac{18 + 0}{18 + 1 + 0 + 0} \times 100\%$$
$$= 94{,}74\%$$
$$Error = \frac{FP + FN}{TP + FP + TN + FN} \times 100\%$$
$$= \frac{1 + 0}{18 + 1 + 0 + 0} \times 100\%$$
$$= 5{,}26\%$$

Accuracy is a measure of how well a classification model predicts the correct classes from all predictions made. Based on Figure 9, an accuracy of 94.74% was obtained, which means that of all the predictions made by the model, 94.74% of them were correct according to the actual class. High accuracy indicates that the classification model has a good ability to differentiate between "Best Selling" and "Not Hot" products based on the given features. Error rate is the inverse measure of accuracy. This measures how often the model makes wrong predictions. Based on Figure 9, an error rate of 5.26% is obtained, meaning that of all the predictions made by the model, around 5.26% of them are wrong. A low error rate indicates that the model has a good ability to avoid making wrong predictions.

The application of the K-Nearest Neighbor (KNN) algorithm in predicting sales of best-selling products is a strategic step taken to increase the accuracy of sales predictions by using historical data and factors that influence sales. This research aims to evaluate the effectiveness of KNN in the context of sales prediction for best-selling products and compare it with other methods that may be more complex. This research follows the steps of Data Collection, Data Pre-processing, Data Sharing, Selection of K Values, Model Training, as well as prediction and Evaluation. After applying KNN to the product sales dataset, the following results were obtained: Cross validation shows that the optimal K value for this dataset is 5. This value provides a balance between bias and variance. Compared with linear regression, KNN shows better performance in terms of prediction accuracy, especially when the data has a non-linear pattern. Compared to more complex models such as Random Forest or Neural Networks, KNN is faster in the prediction process but may be less accurate if the data is very complex or the number of features is very large.

Simplicity and Ease of Implementation, KNN is easy to implement and does not require assumptions about data distribution. Ability to Handle Non-linear Data, KNN can handle data that has a non-linear relationship between features and targets. KNN Limitations, Large Storage Requirements. Since KNN stores the entire dataset to make predictions, storage requirements are large. Computation Speed: The computing process becomes slow when the dataset is very large because it is necessary to calculate the distance for each instance. Sensitive to Noise: The presence of outliers or irrelevant features can reduce the accuracy of the model.

Recommendations for Further Research, Use of Data Enhancement Techniques: Using techniques such as PCA (Principal Component Analysis) to reduce data dimensions can improve KNN performance. Exploration of Other Algorithms: Although KNN shows good performance, exploring other algorithms such as Gradient Boosting or Neural Networks can provide deeper insights. Integration with External Data, namely Combining external data such as market trends or socio-economic data can increase prediction accuracy. Thus, research using the application of the K-Nearest Neighbor algorithm in predicting sales of best-selling products shows quite satisfactory results. Even though it is simple, KNN is able to provide quite accurate predictions with the right selection of K values and good data pre-processing. However, it is important to consider the limitations of KNN and combine it with other methods for more comprehensive results.

CONCLUSION

By applying data mining techniques using the K-Nearest Neighbor method to predict product sales at PT LG Innotek Indonesia based on the number of sales over the last 1 year, a prediction result of 94.74% was obtained. With this high level of accuracy, it can be concluded that the K-Nearest Neighbor method is effectively used in classifying sales of PT's best-selling or non-selling products. LG Innotek Indonesia. With the existing classification results, this method is good enough to be used in identifying sales of best-selling or non-selling products in the future. of course, by implementing this classification model, you can obtain product sales information efficiently, identify the products most in demand by consumers, and organize raw materials so that remain sufficient without excess or shortage of stock.

REFERENCES

A. A. Leasiwal, A. M. Andrian, & I. V Masala. (2021). *Implementasi Algoritma C4. 5 Untuk Klasifikasi Harga Emas*. Universitas Katolik De La Salle.

Ari Sanjaya, & Tri Wahyana. (2022). Penerapan Metode K-Nearest Neighbour Untuk Sistem Prediksi Kelulusan Siswa MTs Nurul Muslimin Berbasis Website. *Journal Transformation of Mandalika*, *3*(2), 31–47.

Azis, A., Zy, A. T., & Sunge, A. S. (2024a). Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4063

Algoritma K-Nearest Neighbor. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *6*(1), 117–124. https://doi.org/10.47233/jteksis.v6i1.1078

Azis, A., Zy, A. T., & Sunge, A. S. (2024b). Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *6*(1), 117–124. https://doi.org/10.47233/jteksis.v6i1.1078

Dewi, S. P., Nurwati, N., & Rahayu, E. (2022a). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, *3*(4), 639–648. https://doi.org/10.47065/bits.v3i4.1408

Dewi, S. P., Nurwati, N., & Rahayu, E. (2022b). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, *3*(4), 639–648. https://doi.org/10.47065/bits.v3i4.1408

Febriana Santi Wahyuni. (2024). Penerapan Teknik Data Mining untuk Menentukan Rencana Strategi Penjualan. *JUPITER (Jurnal Pendidikan Teknik Elektro)*, *7*(1), 47–54.

Harahap, F., Fahrozi, W., Adawiyah, R., Siregar, E. T., & Harahap, A. Y. N. (2023). Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes. *JURNAL UNITEK*, *16*(1), 41–51. https://doi.org/10.52072/unitek.v16i1.541

Herlambang, H. P., Saputra, F., Prasetiyo, M. H., Puspitasari, D., & Nurlaela, D. (2023). Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor. *Jurnal INSAN - Journal of Information System Management Innovation*, *3*(1), 21–28. https://doi.org/10.31294/jinsan.v3i1.2097

Ike Yolanda, & Hasanul Fahmi. (2021). Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT.Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *3*(1), 9–15.

Mhd Angga Sabda, & Suhardi Suhardi. (2023). Implementasi Data Mining Dalam Memprediksi Penjualan Parfum Terlaris Menggunakan Metode K-Nearest Neighbor. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *5*(2), 415–422.

Nolly, R. A., Fitria, A., & Saputra S, K. (2023). Penerapan Algoritma K-Nearest Neighbors untuk Klasifikasi Fragmen Metagenom Berdasarkan Ekstraksi Fitur K-Mers. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, *17*(1), 52. https://doi.org/10.30872/jim.v17i1.5779

Nur Fajri, F., Tholib, A., & Yuliana, W. (2022). Application of Machine Learning Algorithm for Determining Elective Courses in Informatics Study Program. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(3). https://doi.org/10.28932/jutisi.v8i3.3990

Prastiwi, H., Jeny Pricilia, & Errissya Rasywir. (2022). Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, *2*(1), 141–148. https://doi.org/10.33998/jakakom.2022.2.1.34

Rifky, L., Nugraha, Z., Saputra, B., Pratama, D., Raswir, E., & Pratama, Y. (2022). Implementasi Data Mining Untuk Penjualan Mobil Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, *2*(2), 225–230. https://doi.org/10.33998/jakakom.2022.2.2.109

Rizki, F., Faisol, A., & Santi Wahyuni, F. (2020). PENERAPAN METODE NAIVE BAYES UNTUK MEMPREDIKSI PENJUALAN PADA UD. HIKMAH PASURUAN BERBASIS WEB. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *4*(1), 26–34. https://doi.org/10.36040/jati.v4i1.2379

Rosidi, R. P. M., & Setiawan, K. (2024). Implementasi Algoritma Naïve Bayes Terhadap Data Penjualan untuk Mengetahui Pola Pembelian Konsumen pada Kantin. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(1), 120–126. https://doi.org/10.35870/jimik.v5i1.407

Rozzi Kesuma Dinata, & Novia Hasdyna. (2020). *Machine Learning* (Dr., Fajriana, & M. Si. S.Si, Eds.; Vol. 7). Unimal Press.

Sari, R., & Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal on Software Engineering (IJSE)*, *5*(2), 51–60. https://doi.org/10.31294/ijse.v5i2.6957

Wardani, N. W., Nugraha, P. G. S. C., & Mahendra, G. S. (2024). Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel. *JST (Jurnal Sains Dan Teknologi)*, *12*(3). https://doi.org/10.23887/jstundiksha.v12i3.38605