Prediction Of Laptop Sales Using The K-Nearest Neighbor Method At The MVP Computer Mawar Store, In Takengon

Adi Kurniawan [1], Rayuwati [2], Ira Zulfa [3]
[1,2,3] Universitas Gajah Putih

Abstract. This research relates to predictions of laptop sales in computer shops in Central Aceh, with a focus on laptop brands Acer, Asus, HP and Lenovo. Over the last three years, sales of these laptops have reached 1,629 units, with a monthly average of between 108 and 150 units. Business owners today prefer brands with the highest percentage of sales, but this can lead to dead stock problems. Therefore, the author proposes using data mining techniques, especially the K-Nearest Neighbor (K-NN) method, to make recommendations for the number of products to be purchased by business owners based on past sales data. The K-NN method requires complete, structured and continuous sales data. It is important to choose an appropriate K value, and other factors such as weather, seasons, promotions, and special events also affect laptop sales. K-NN models may need to be combined with other data to improve prediction accuracy. It is hoped that this research will provide academic benefits in expanding knowledge about the use of the K-NN method in sales prediction, as well as practical benefits for business owners in planning their sales strategies. The research conclusions highlight the importance of good data collection, choosing the right K value, and considering external factors in the laptop sales prediction process.

Keywords: Sales, Prediction, K-Nearest Neighbor

INTRODUCTION

Based on the sales data obtained, the most laptop brands sold over the last three years in sequence were Acer with 456 units (27.9% of total sales), Asus brand with 450 units (27.6% of total sales), HP brand with 378 units (23.2% of total sales), and the Lenovo brand with 345 units (21.1% of total sales), with total sales of all brands being 1,629 units. The monthly sales trend over the last three years has been in the range of 108 units to 150 units (average monthly percentage is 6.6% - 9.3%).

Thus, business owners need planning recommendations according to the number of sales in the past for planning for the future. In making sales recommendations, you can use data mining techniques, namely by analyzing sales data using methods contained in data mining as a result, you can find news or sales patterns that are useful for making recommendations for the number of products that will be purchased by business owners.

In this research, the author uses a well-known and practical method, namely the k-nearest neighbor method and has a higher level of accuracy than other classification methods. In Anggriandi's (2021) research on predicting laptop sales using the k-nn method, it obtains RMSE values between 1.325 to with 54,055. Also, Yanti's (2022) research on laptop sales predictions produced high accuracy with an error of 2% - 3.6%. In research by Abdullah, et al (2022) regarding predictions of best-selling sales, the accuracy was 83.3%.

The advantage of this method is that it can analyze the closeness and similarity between one data and another data using a distance calculation formula based on the number of nearest neighbor distances which are considered k. The value of k generally amounts to more

* Adi Kurniawan

than one and is usually used in odd numbers. The weight of the distance value from k will be calculated, unlike the evidence value which is always fixed for each class in one sample. The value of the posterior will later be compared with the posterior values of other classes to determine what class a sample will be classified into, to carry out a prediction process on data on the number of future sales. So based on these two studies the author will use the k-NN method with laptop sales data in Central Aceh entitled "Laptop Sales Prediction Using the K-Nearest Neighbor Method at the Mawar computer Mvp Store, in Takengon".

LITERATURE REVIEW

Prediction

Forecasting is the process of systematically predicting what is most likely to happen in the future, based on past and present information, so that errors (the difference between what happens and what is predicted) can be minimized. Predictions do not need to provide an absolute answer to whatwhat will happen, but trying to find answers as close as possible using what will happen (Roza et al., 2020).

Data mining concept

Data mining, in simple terms, is an extraction step to obtain important information that is implicit and unknown. Apart from that, data mining is related to various fields including statistics, machine learning, pattern recognition, computing algorithms, database technology, and high performance computing.

K-Nearest Neighbor

K-NN is a method used in data classification. In general, K-NN functions to find the shortest distance between the data to be evaluated and the closest K-NN in the training data (Agusta, 2007).

Confusion Matrix

Evaluation of classification data mining types is carried out by testing for the object truth prediction process. The testing process utilizes a confusion matrix which places the predicted class at the top of the matrix then what is observed is placed on the left of the matrix. Each matrix cell contains a number that displays the actual number of cases of the observed class (Muslim, et al., 2019).

Orange Data mining

Orange is an open source based machine learning and data visualization application. Orange has a different user interface than what you are used to. In the middle is the Orange Canvas where the widget is placed. Each widget has a specific function (loading data, filtering,

fitting a specific model, displaying some visualization), and receives and/or provides data from/to other widgets. Widgets send signals to each other through data streams by organizing widgets into schemas and connecting them. Two widgets are only connected if their signal types match (Demsar, 2010).

METHODOLOGY

Research methodology

To solve the problems in this research, the author created a research methodology scheme using the data mining knowledge discovery in database (KDD) method. The following are the planned stages that will be implemented as in Figure 1 below:
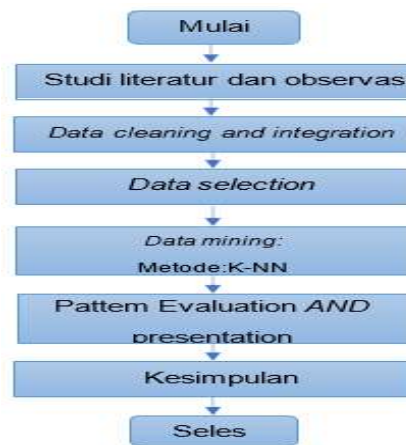
Figure 1 Flow of research methodology.

RESULTS AND DISCUSSION

Results

The results of implementing laptop sales predictions using the Orange data mining application are used.

Sales prediction results with k-nn

To get results from implementing laptop sales predictions using the orange data mining application with the k-nn method, go through the following stages:

1. Reading training data

To read training data in the Orange Data Mining application via a file block, as in the following image:
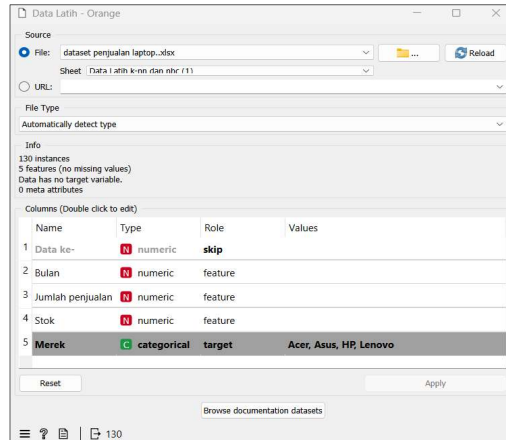
Figure 2 Training data import settings

Based on Figure 2 above, it can be explained that there are three feature variables (independent) and 1 target variable (dependent). Then after importing the training data, view the training data through the training data view block, so that the results of importing the training data can be seen in the following image:
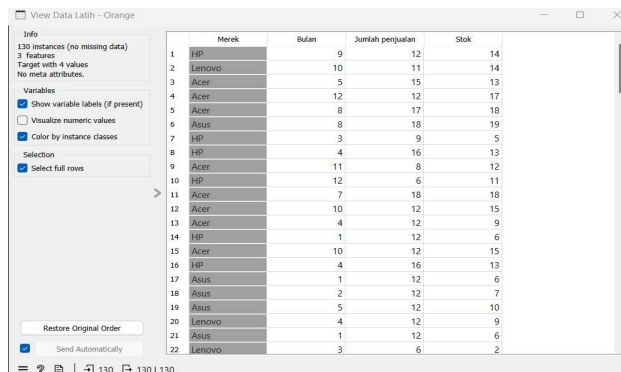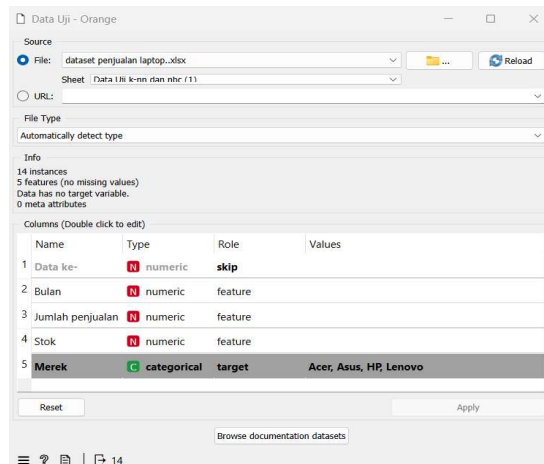

Figure 3 Results of training data view

Based on Figure 3 above, it can be explained that there are 130 training data with month variables, number of sales, and stock as feature variables, then brand variables as targets.

2. Test data reading

To read test data in the Orange Data Mining application via a file block, as in the following image:

Figure 4 Test data import settings

Based on Figure 4 above, it can be explained that there are three feature variables (independent) and 1 target variable (dependent). Then after importing the test data, view the test data through the test data view block, so that the results of the test data import can be seen in the following image:

Based on Figure 5 above, it can be explained that there are 14 test data with month variables, number of sales, and stock as feature variables, then brand variables as targets.

3. Variable selection

To select test data variables, use the select columns block, as in the following image:
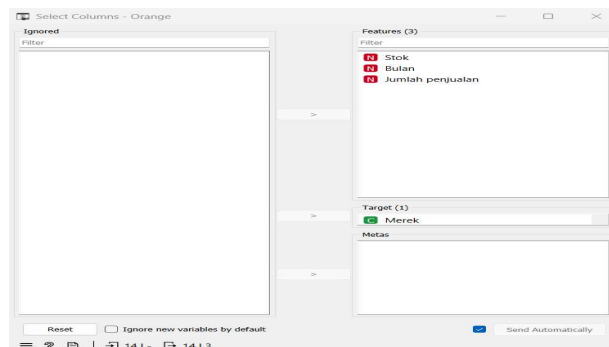
Figure 6 Selection of feature and target variables

In the picture above there is a selection of test data variables which contain stock, month, quantity and brand as targets, using the select columns feature, with the select columns feature in the orange application it is easier for us to select existing data.

4. k-nn process

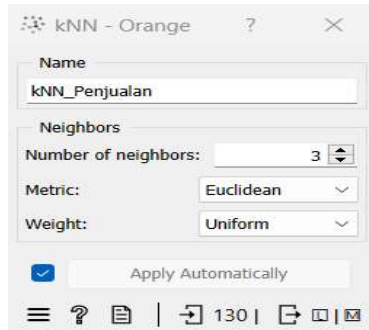To carry out the k-nn process, settings are required via the kNN block as in the following image:

Figure 7 K-NN process setup

In the picture above there is the process name, namely kNN_Sales, the number of k, namely 3, the type of distance measurement, namely Euclidean, and general weighting (uniform).

5. Prediction

So the prediction results from the k-nn modeling results which were tested with test data using the predictions block, can be seen in the image below

Figure 8 Prediction results of test data

Based on Figure 8 above, it can be seen that the values obtained by using the prediction feature in the Orange application will result in an accuracy of 0.969, precision of 0.893, and recall of 0.857.

6. Confusion matrices

So the results of the confusion matrix to see the amount of data categorized as true positive, false positive, true negative and false negative, can be seen in the following picture:
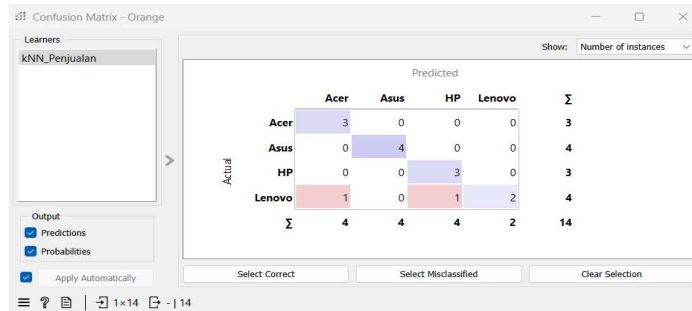
Figure 9 Confusion matrix results of sales test data

Based on Figure 9 above, it can be explained that the values for true Acer, ture Asus, true HP, true Lenovo are 12, false Lenovo are 2, with an error of 0.667 based on. So that the accurate test data is 12 out of 14 test data.

Discussion

In this section the author will discuss the performance level of the k-NN method, above in terms of accuracy, precision and recall based on laptop sales test data.

Accuracy:Accuracy measures the extent to which a classification model is correct in predicting the class of a data sample.

precession: Precision measures the extent to which positive predictions from a classification model are correct.

Recall: Recall measures the extent to which the classification model succeeds in identifying all instances that should fall into the positive class.

Performance of k-NN

Based on the data in Figure 9, namely the results of the confusion matrix of sales test data using the k-NN method, the performance percentage can be calculated as follows:

Table 1 Sales confusion matrix with k-NN

Actual	Prediction			
	Acer	Asus	MOBILE PHONE	Lenovo
Acer	3	0	0	0
Asus	0	4	0	0
MOBILE PHONE	0	0	3	0
Lenovo	1	0	1	2

Based on the Confusion Matrix calculation above, the performance of k-NN in sales prediction based on laptop brand is obtained, namely an accuracy value of 85.71%, a precision value of 87.5%, and a recall value of 87.5%.

Based on the calculations above, the accuracy, precision and recall values for sales

predictions using the k-NN method can be seen as follows:

Table 1 Accuracy, precision and recall of sales predictions

No.	Method	Accuracy	Precision	Recall
1.	k-NN	85.71%	87.5%	87.5%

From the table above it can be concluded that the k-NN method shows a high level of accuracy in predicting laptop brand sales, so as for the results of each table, the average accuracy is 85.71%, the average precision is 87, 5%, and an average recall of 87.5%, this shows that the k-NN method tends to provide accurate results.

The predictions for laptop sales based on brand in 2023 from January to October can be seen in the image below.

Figure 10 prediction results for 2023

Based on figure 10 above, it can be explained that the values for true Acer 9, true Asus 9, true HP 9, true Lenovo 8, total 35, while for false Acer 1, false Asus 1, false Hp 1, false Lenovo 2, total 5, based on figure 4.9. So the accurate 2023 prediction data is 35 out of 40 prediction data.

The details for monthly predictions based on laptop brand can be seen in the following image.

Picture 11 detailed monthly predictions.

Based on the image above, of the 5 incorrect or (false) predictions in the previous figure 11, it can be explained that the prediction error occurred in month 1 (January) for the HP brand, month 4 (April) for the Linovo brand, month 5 (May) for the brand linovo, month 6 (June) for the Acer brand, and month 10 (October) for the Asus brand.

CONCLUSION

Conclusion

Based on the results of the research and discussion above, in this chapter the author can draw several conclusions as follows: The conclusion is using the K-Nearest Neighbor (K-NN) method to predict laptop sales at the Mawar computer Mvp Store, in Takengon.

1. Using the K-NN method requires complete, structured and continuous laptop sales data. The more data available, the better predictions the K-NN model can provide.

2. Choosing the correct K value (number of nearest neighbors) is very important. A K value that is too small will make the model overfit, and a K value that is too large will make the model too general. Experimentation is needed to determine the optimal K value.

3. Other factors such as weather, seasons, promotions and special events will also influence laptop sales. K-NN models may need to be combined with other data to improve prediction

accuracy.

Suggestion

There are suggestions that the author wants to convey in predicting laptop sales using the K-Nearest Neighbor method at the Mawar Computer Mvp shop, in Takengon.

1. Toko Mawar Kompuer Mvp, needs to collect more data on laptop sales over time. You can train a K-NN model using historical data to improve accuracy.

2. Evaluate the K value Experiment with different K values to determine which value best fits the laptop sales data at Toko Mawar Kompuer Mvp.

3. Apart from laptop sales data, you should also pay attention to external factors that can influence sales, such as weather, seasons and promotions. Integrating this data into predictive models will improve accuracy.

4. Regular monitoring and evaluation, to ensure that the K-NN model is accurate over time, laptop sales forecasts should be evaluated periodically. If sales behavior changes, the model must be updated.

5. Use other analysis methods, such as regression, to understand more deeply the factors that influence laptop sales. Combining multiple analysis methods can provide better insights.

BIBLIOGRAPHY

Abdullah Robi Wariyanto, et al (2022). Application of Data Mining to Predict the Number of Best Selling Products Using the Naive Bayes Algorithm Case Study (Toko Prapti). GLOBAL INFORMATICS SCIENTIFIC JOURNAL, 13(1), 20-27

Anggriandi, Dendi (2021). The data mining application uses the k-nn method in predicting laptop sales (case study: PT UNIVERSAL SELLUER COMPOTINDO). STMIK Palangkaraya.

Agusta, Yudhi. (2007). "k-Means Application Problems and Related Methods". JournalSystems and Informatics Vol.3 No.1: 47-60

Muslim, Much Aziz, et al., (2019). Data Mining Algorithm C4.5 Accompanied by case examples and their application with computer programs. ILKOM UNNES, Semarang.

Roza, et al, (2020). Information System Tutorial for Predicting the Number of Customers Using the Web-Based Multiple Linear Regression Method Using the Creative CodeIgniter Framework, Indonesia.

Yanti, Selfi (2022). Prediction of Goods Sales at UD. Computer Fortune Using the k-Nearest Neighbor Regression Method. Faculty of Engineering, White Elephant University.