


# Application Of Data Mining For Prediction Of Sales Of Best-Selling Electronic Products Using K-Nearest Neighbor Method

**Yulia Rizki Amalia**

Information Systems Study Program, Faculty of Science and Technology, Universitas Islam Negeri Raden Fatah Palembang

ARTICLEINFO	ABSTRACT
<p><b>Article history:</b> Received Dec 22, 2019 Revised Jan 02, 2020 Accepted Jan 15, 2020</p> <hr/> <p><b>Keywords:</b> Forecasting; Data Mining; K-Nearest Neighbor (KNN).</p>	<p>PT. Bintang Multi Sarana Palembang is one of the largest electronics companies in South Sumatra. This company has a wide variety of electronic products on offer. Judging from the large number of consumer demands for electronic products based on sales data for the last 3 years, predictions are needed for sales of the best-selling electronic products, in order to make it easier for the company to plan stock supply. To determine the sales of the best-selling electronic products, data mining classification techniques and the K-Nearest Neighbor algorithm are used. The results of this study are predictions of the best selling electronics sales as many as 6 types of products from 22 types of products sold, namely CTV, Refrigerator, DVD, Speaker, Washing Machine and LCD. Based on the accuracy value of the best-selling product sales classification of 92.51%.</p> <p><i>This is an open access article under the <a href="#">CC BY-NC</a> license.</i></p> 

## Corresponding Author:

Yulia Rizki Amalia  
Information Systems Study Program, Faculty of Science and Technology,  
Universitas Islam Negeri Raden Fatah Palembang  
E-mail: [yuliarizki@gmail.com](mailto:yuliarizki@gmail.com)

## 1. INTRODUCTION

Prediction or sales forecasting (forecasting) is a calculation to predict future conditions through testing conditions in the past. Forecasting future sales means determining the estimated amount of sales volume, even determining the potential sales and market area controlled in the future.(Juniarti, 2018). One of the uses of prediction is to help company owners in decision making in determining the amount of goods that must be provided by the company. In addition, predictions can help the company in planning the supply of stock, because this prediction can provide the best output so it is hoped that the risk of errors caused by planning errors can be minimized. Prediction is usually used to find information from large amounts of data, so data mining is needed(Mahena, Rusli, & Winarso, 2015).

Data mining is a multidisciplinary field that combines techniques from machine learning, pattern recognition, statistics, databases and visualization to solve problems of retrieving information from large database storage.(WIJAYA, nd). Data mining can be used to dig up information from big data so that information is obtained that can be used to predict sales. In data mining there are many techniques in the process, to find hidden patterns or information including clustering, regression, association, and classification.(Amalia, 2018).

The methods contained in data mining for prediction include the Naive Bayes method, a method that divides problems into classes based on the characteristics of similarities and differences using statistics that can predict the probability of a class, Naive Bayes assumes that the

value of an input attribute in a given class does not depend on other attribute values (Amalia, 2018), besides that there is also the K-Nearest Neighbor (KNN) method. This method is a method used to classify objects based on learning data that is closest to the object. The advantage of the KNN method in predicting is that it is strong against noise data training and is effective when the training data is large.

PT. Bintang Multi Sarana Palembang is an electronics and furniture company engaged in trading that provides a wide range of electronic and furniture products. PT. Bintang Multi Sarana Palembang currently uses a sales information system to support the company's performance as well as for sales information services. The company has also implemented an online sales system to promote its sales. Products sold include: Television (TV), Mobile (HP), Air Conditioner (AC), washing machines and other electronic products. Judging from the large number of consumer demands for electronic products, it is necessary to predict the sales of electronics that are most requested by consumers. This prediction aims to facilitate the supply of goods at PT.

## **2. METHOD**

### **2.1 Research methods**

The research method in this study is a descriptive analytical approach. Analytical descriptive research is a method that functions to describe or provide an overview of the object under study through data or samples that have been collected as they are without analyzing and making conclusions that apply to the public. In other words, analytical descriptive research takes problems or focuses attention on problems as they are when the research is carried out, the research results are then processed and analyzed to draw conclusions (MAULIDIA, 2017).

The steps that need to be taken in the descriptive analytical approach are as follows:

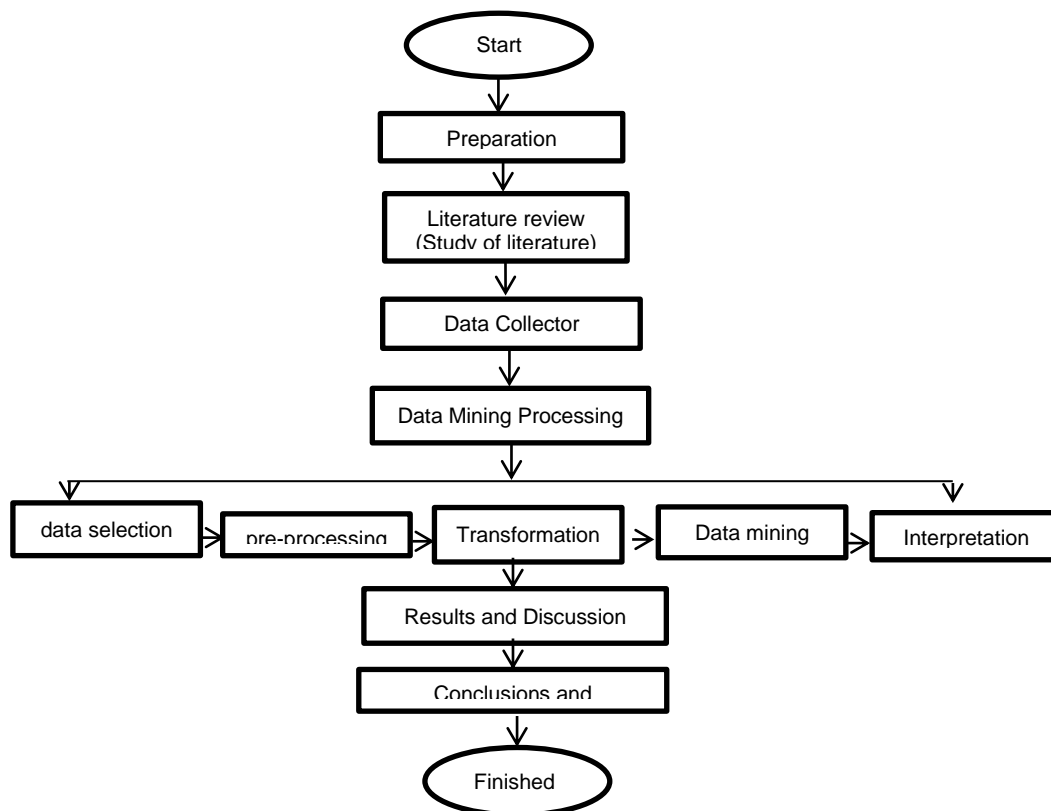
- a. Formulate research objectives
- b. Determine the unit of study and relate the things to be studied based on the characteristics possessed and the process that forms the basis of the research.
- c. Determine the design in selecting units and data collection techniques.
- d. Collecting data.
- e. Organizing information, collected data, and interpreting and generalizing.

### **2.2 Method of collecting data**

The data collection techniques that the researchers did in the study were as follows:

- a. Interview, According to (Indriani, 2015) said that the interview method is also called (interview) is a way of collecting data which is carried out through conversations between researchers (assigned people) and research subjects or respondents or data sources. This activity was carried out to find information about electronic sales at PT Bintang Multi Sarana Palembang. The data used for data mining calculations is electronic sales data from 2015 to 2017.
- b. Observation, according to Jogiyanto (2008:89) observation is a technique or approach to obtain primary data by observing directly the data object. The approach is taken by observing directly the events in the electronic sales system at PT Bintang Multi Sarana Palembang.
- c. Literature Study, is a data collection technique by conducting an in-depth study of books, literatures, notes and reports that have to do with the problem being solved. (Hidayat, 2015). Data collection is carried out directly from other sources such as books, theses, journals and other documents related to this research.

### **2.3 Research Stages**



**Figure 1** Research Stages

The research stages are outlined in this flow chart. Describes the research process that will be taken as well as describes the research as a whole.

- a. Preparation. This stage is the stage of the subject (population) of PT. Bintang Multi Sarana Palembang. The object taken is an electronic product. Limitation and develop a research plan.
- b. Literature review. In the literature review, a literature review and study of sales predictions and related matters is carried out.
- c. Data collection. Data was collected by interviewing the accounting, observation and document sections.
- d. Data mining processing. The collected data is processed according to the Knowledge Discovery in Database (KDD) stage.
- e. Results and Discussion. The discussion at this stage describes the results of the data mining process carried out using the K-Nearest Neighbor method.
- f. Conclusions and suggestions. Make conclusions from research results and provide suggestions for the company so that it can be even better.

### 3. RESULTS AND DISCUSSIONS

#### 3.1 Results

After analyzing the design with data mining stages to produce sales predictions for the best-selling electronic products at PT. Bintang Multi Sarana Palembang by using the K-Nearest Neighbor Algorithm, this analysis ends by carrying out the actual data mining process, the results achieved by the researchers are to find out the best-selling electronic sales predictions based on sales data of electronic products from sales data in 2015, 2016, 2016 & 2017 in PT. Bintang Multi Sarana Palembang. Researchers used the RapidMiner application to

simplify the data mining process, which resulted in the best-selling electronic sales prediction information. And the prediction results obtained from the sales of the best-selling electronic products as many as 6 types of products from 22 types of products sold including CTV, Refrigerator, DVD, Speaker, Washing Machine and LCD.

a. Data Selection

The data used in this study is data on sales of electronic products based on sales for the last 3 years from 2015, 2016, and 2017 originating from PT. Bintang Multi Sarana Palembang. Then the data is selected and will be used to be processed in predicting the best-selling electronic sales. The attributes used in determining the best-selling electronic sales predictions are the attributes of Item Name, Quantity and Month. Can be seen in Table 4.1.

**Table 1.**  
Data Selection Table

No	Name of goods	Quantity	Month
1	LED	1	JANUARY
2	LED	1	JANUARY
3	LED	1	JANUARY
4	REFRIDGERATOR	1	JANUARY
5	DVD	2	JANUARY
6	air conditioning	1	JANUARY
7	DVD	1	JANUARY
8	REFRIDGERATOR	1	JANUARY
9	SPEAKER	1	JANUARY
10	CTV	1	JANUARY
11	LED	1	JANUARY
12	LCD	1	JANUARY
13	REFRIDGERATOR	1	JANUARY
14	REFRIDGERATOR	1	JANUARY
15	WASHING MACHINE	1	JANUARY
16	WASHING MACHINE	1	JANUARY
17	CTV	1	JANUARY
18	REFRIDGERATOR	2	JANUARY
19	CTV	1	JANUARY
20	CTV	1	JANUARY
.....	.....	.....	.....
1174	LCD	1	DECEMBER

b. Preprocessing

The preprocessing stage is carried out for grouping the types of sales of electronic products based on the number of sales per month and year to simplify the process of calculating predictions. After the data is grouped, then all are added up so that it becomes sales data for all electronics. It can be seen in the following electronic product sales tables:

**Table 2**  
LCD Sales Data Sample

Month	2015	2016	2017
January	2	1	4
February	5	1	2
March	1	2	5
April	5	1	3
May	2	4	3
June	0	5	6
July	1	1	0
August	5	1	1
September	0	2	2
October	1	4	2
November	0	0	4
December	4	4	4

c. Transformation

At this transformation stage, the results of the preprocessing data grouping are then used for training data. The process of forming training data based on existing data, the data must be selected first to determine which attributes can affect the best-selling electronic sales called target data, where target data is data that contains attributes that will become relevant and support attributes in the data mining process. . The training data used are as shown in the following table:

**Table 3**  
LCD Sales Testing Data

BLN-1	BLN-2	BLN-3	BLN-4	BLN-5
2	4	0	4	4
4	0	4	4	2
0	4	4	2	5
4	4	2	5	3
4	2	5	3	3
2	5	3	3	6
5	3	3	6	0
3	3	6	0	1
3	6	0	1	2
6	0	1	2	2
0	1	2	2	4
1	2	2	4	4

The picture above shows that the data used for testing is sales data from January 2017 - December 2017 which will produce sales predictions in 2018.

d. K-Nearest neighbor

Based on the data mining stages for the K-Nearest Neighbor algorithm, as for the steps of the K-Nearest Neighbor:

- 1) Determination of the value of k. The determination of the value of k used does not have a standard rule, but in this study the value of k used is 3.
- 2) Calculate the distance between the training data and test data in the transformation stage using the Euclidean Distance calculation as follows:  
 $d1 = (2-1)^2 + (5-2)^2 + (1-4)^2 + (5-0)^2 + (2-4)^2 = 6.92820$   
 $d2 = (5-1)^2 + (1-2)^2 + (5-4)^2 + (2-0)^2 + (0-4)^2 = 6.16441$   
 $d3 = (1-1)^2 + (5-2)^2 + (2-4)^2 + (0-0)^2 + (1-4)^2 = 4.69041$
- 3) Sorting of calculated data. The distance that has been obtained is then sorted from the closest to the farthest (ascending).  
 After sorting obtained:  
 $d13 = 2.44948$ ,  $d6 = 3.46410$ ,  $d10 = 3.74165$ ,  $d16 = 3.87298$ ,  $d11 = 4.35889$ ,  $d3 = 4.69041$ ,  $d14 = 5.09901$ ,  $d8 = 5.38516$ ,  $d18 = 5.47722$ ,  $d17 = 5.65685$ ,  $d12 = 5.65685$ ,  $d4 = 5.83095$ ,  $d15 = 6$ ,  $d19 = 6.08276$ ,  $d2 = 6.16441$ ,  $d7 = 6.48074$ ,  $d9 = 6.55743$ ,  $d1 = 6.92820$ ,  $d5 = 7.41619$ .
- 4) Determine the test result data group based on the majority label from the k closest neighbors. Because the value of k = 3, the 3 smallest distances are taken, namely d13, d6, and d10.
- 5) By using the Nearest Neighbor category which is the majority, it can be predicted the number of sales in the next period.

e. K-NN Algorithm Results

This stage is the last stage of the rapidminer data mining process where after all operators are connected and then click the F11 button play icon, a results tab will appear, which contains a prediction of all data that meets the k = 3 value, the sales prediction results are obtained as shown below :

ExampleSet (12 examples, 1 special attribute, 5 regular attributes)

Row No.	prediction(T...	BLN-1	BLN-2	BLN-3	BLN-4	BLN-5
1	4	1	2	4	0	4
2	1	2	4	0	4	4
3	4	4	0	4	4	2
4	2.333	0	4	4	2	5
5	1.667	4	4	2	5	3
6	4	4	2	5	3	3
7	1.667	2	5	3	3	6
8	1.667	5	3	3	6	0
9	1.667	3	3	6	0	1
10	3.333	3	6	0	1	2
11	2.667	6	0	1	2	2
12	3.333	0	1	2	2	4

**Figure 1** LCD sales prediction results

The picture above is the result of LCD type sales predictions where the highest sales predictions are in January, March and June with a predicted value of 4.

#### 4. CONCLUSION

In this study, modeling was carried out using the k-nearest neighbor algorithm using data that was processed based on the Knowledge Discovery in Database (KDD) stages. Based on the results of data mining calculations using classification techniques and the k-nearest neighbor algorithm, the prediction results of the best-selling electronic product sales were 6 types of products from 22 types of products sold, namely CTV, Refrigerator, DVD, Speaker, Washing Machine and LCD based on the accuracy value of the sales classification of the best-selling products was 92.51%.

#### References

- Amalia, Y. R. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus: PT. Bintang Multi Sarana Palembang)*. UIN RADEN FATAH PALEMBANG.
- Hidayat, T. (2015). PENILAIAN YANG EFEKTIF BIDANG STUDI MATEMATIKA DITINJAU DALAM PERSPEKTIF KURIKULUM 2013. *Jurnal Penelitian Pendidikan*, 7(2), 1172–1180.
- Indriani, A. (2015). Penerapan Pembelajaran Tematik Kurikulum 2013 oleh Guru SD/MI di Desa Klepek Kecamatan Sukosewu Kabupaten Bojonegoro Semester Gasal Tahun Ajaran 2014/2015. *Jurnal Varidika*, 27(1), 43–49.
- Juniarti, A. T. (2018). *ANALISIS PERAMALAN PENJUALAN DALAM MENETAPKAN PERENCANAAN PRODUKSI TAHU SUSU PUTIH PADA PABRIK TAHU SUSU LEMBANG*. Perpustakaan Fakultas Ekonomi dan Bisnis Unpas Bandung.
- Mahena, Y., Rusli, M., & Winarso, E. (2015). Prediksi Harga Emas Dunia Sebagai Pendukung Keputusan Investasi Saham Emas Menggunakan Teknik Data Mining. *Kalbiscentia J. Sains Dan Teknol*, 2(1), 36–51.
- MAULIDIA, F. (2017). *ANALISA PRODUKTIVITAS DAN FAKTOR YANG MEMPENGARUHI KINERJA TUKANG PADA PEMASANGAN BATA DENGAN METODE PRODUCTIVITY RATING (Studi Kasus: Rumah Susun-SNVT Universitas Brawijaya)*. University of Muhammadiyah Malang.
- WIJAYA, I. H. (n.d.). *PREDIKSI MAHASISWA DROP OUT BERDASARKAN KLASIFIKASI ADMINISTRATIF*.