BIOS: Jurnal Informatika dan Sains
Volume 2 Number 01, 2024, DOI 10.54209/bios.v2i01
E.SSN 2988-6910(Online)
https://seaninstitute.or.id/bersinar/index.php/bios/index

Application of Data Mining to Predict Best-Selling Stationery Sales at Berkah ATK Store Using the K-Nearest Neighbor (K-NN) Method

Maida Indrayani
Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia

Article Info

Corresponding Author:
Maida Indrayani
E-mail:
maida.indrayani92@gmail.com

ABSTRACT

ATK is an abbreviation for office stationery used by various groups. Examples of ATK are pencils, pens, erasers and so on. Toko Berkah ATK is one of the stores that sells office supplies.Sales of stationery at the Berkah ATK Store often experience a decrease and increase every month. It depends on the needs of the community who buy at the Berkah ATK Store. To avoid the accumulation of unsold stationery items, data mining is applied using the K-Nearest Neighbor (K-NN) method to predict the best-selling stationery sales at the Berkah ATK Store.

Keywords:
Data Mining,K-Nearest Neighbor (K - NN), Prediction, ATK

INTRODUCTION

Berkah ATK Store is located in Binjai City, North Sumatra, which sells all kinds of office supplies such as books, pens, markers, rulers, folders, stamps, envelopes, erasers, etc. Sometimes some of the stationery supplies sold are sold out and some are not sold out at all. This is because the stationery equipment purchased is based on the needs of the community. To avoid this, a prediction of the most sold stationery sales is made using data mining techniques using the K-Nearest Neighbor (K-NN) method.

The K-Nearest Neighbor (K-NN) method is a method in data mining that is often used to predict sales of goods. Based on research conducted by Rismala et al. entitled "Application of the K-Nearest Neighbor (K-NN) Method for Predicting Best-Selling Motorcycle Sales" shows thatThe K-NN algorithm can be used in the classification of motorcycle sales data at PT. X. Based on the test, the accuracy results were 96.15%, which means that the data set can be used in the next stage as valid data for use. From this background, the authorusing the K-Nearest Neighbor (K-NN) method to predict the most sold stationery at the Berkah ATK Store to make it easier for the store to provide the best-selling stationery stock.

RESEARCH METHODOLOGY

Stages of Research Methods

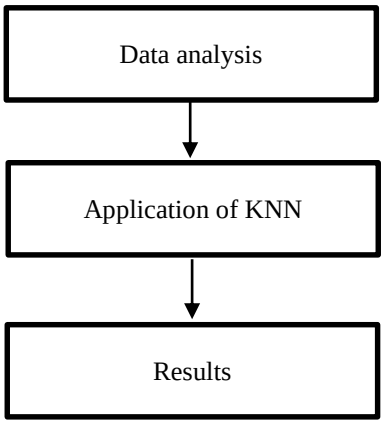

Figure 1. Research Framework

Steps for implementing the K-Nearest Neighbor (K-NN) method:

1. Determine k where k is the closest distance
2. Calculating distance using Euclidean Distance
   Formula : $\sqrt{\sum_{i=1}^{K}(X1-Y1)^2+(X2-Y2)^2}$
   Xi = Values in training data
   Yi = Value in test data
   k = Attribute dimension
3. Sort the results in ascending order to rank them from largest to lowest ATK sales data.

Data analysis

The training data used is stationery sales data for the last 6 months taken in January, February, March, April, May and June 2024.

Table 1. Stationery sales data for January, February, March, April, May and June 2024

No	ATK Name	Stock	Products Sold						Total Sales	Information
			Jan	Feb	Mar	Apr	May	June		
1	Fountain pen	100	10	24	8	10	15	10	77	Popular
2	Pencil	100	5	12	4	8	5	12	46	Popular
3	Highlighter	100	5	7	4	8	12	9	45	Popular
4	Eraser	100	4	6	7	12	8	4	41	Popular
5	Ballpoint Pen	100	10	5	15	8	12	5	55	Popular
6	Whiteboard marker	100	7	6	8	9	3	9	42	Popular
7	Colored pencils	100	5	8	9	5	10	3	40	Popular
8	Gel Pen	100	2	3	12	9	7	20	53	Popular
9	Roller Pen	100	5	6	5	15	5	4	40	Popular
10	Whiteboard Marker	100	9	9	6	8	3	8	43	Popular
11	Zippered Map	50	1	3	2	5	2	1	14	Not Selling Well
12	Notebook	10	3	6	4	1	5	0	19	Not Selling Well
13	Label	50	5	2	2	0	1	1	11	Not Selling Well
14	Ring Folder	50	9	3	2	1	4	4	23	Currently
15	Paper hole puncher	10	0	7	6	0	0	1	14	Not Selling Well
16	Punch Hole	10	4	8	6	0	2	2	22	Currently
17	Staples	50	8	4	5	3	2	1	23	Currently

No	ATK Name	Stock	Products Sold						Total Sales	Information
			Jan	Feb	Mar	Apr	May	June		
18	Glue	50	1	7	5	6	8	3	30	Currently
19	Scissors	20	4	2	8	3	3	2	22	Currently
20	Ruler	50	6	3	9	5	2	7	32	Currently
21	Stapler	50	2	0	3	5	6	1	17	Not Selling Well
22	Paperclip	50	3	1	3	0	0	6	13	Not Selling Well
23	Paper hole puncher	10	1	1	2	0	1	2	7	Not Selling Well
24	Pencil Case	50	0	5	3	1	1	1	11	Not Selling Well
25	Marker Eraser	50	1	5	3	3	3	2	17	Not Selling Well
26	Tissue box	20	2	7	3	2	0	0	14	Not Selling Well
27	Paperclip	10	3	0	3	1	0	0	7	Not Selling Well
28	Digital scales	5	0	0	1	0	0	1	2	Not Selling Well
29	Book Binder	50	4	5	3	7	3	8	30	Currently
30	Portfolio	50	2	3	4	5	7	5	26	Currently
31	Plastic Map	100	10	7	2	6	5	5	35	Currently
32	Cardboard Map	100	5	7	7	3	4	5	31	Currently
33	Agenda book	20	2	3	9	3	4	3	24	Currently
34	Clip Folder	50	4	3	2	2	6	2	19	Not Selling Well
35	Duty stamp	100	15	9	20	11	20	10	85	sLaris
36	Paperboard	50	5	3	1	5	2	2	18	Not Selling Well
37	Map Accordion	20	3	1	1	1	2	0	8	Not Selling Well
38	Acco Folder	20	3	1	1	1	2	3	11	Not Selling Well
39	Steel Staples	10	2	2	2	2	6	3	17	Not Selling Well
40	Plain Map	50	4	6	2	3	5	3	23	Not Selling Well

RESEARCH RESULT

Perform data normalization using the minimum and maximum sales values for the last six months. The following are the results of data normalization calculations using the minimum and maximum formulas:

Formula:

$$x' = x - \min(x) / \max(x) - \min(x)$$

Information:

x'= new scale value

x = data value before normalization

min = lowest data value

max = highest data value

Table 2. Results of Maximum and Minimum Normalization Calculations

No.	Min Max Stock of ATK	Results	Min Max ATK Sold	Results
1	100 - 5 / 100 - 5=	1	77 - 2 / 85 - 2 =	0.903
2	100 - 5 / 100 - 5=	1	46 - 2 / 85 - 2 =	0.530
3	100 - 5 / 100 - 5=	1	45 - 2 / 85 - 2 =	0.518
4	100 - 5 / 100 - 5=	1	41 - 2 / 85 - 2 =	0.469
5	100 - 5 / 100 - 5=	1	55 - 2 / 85 - 2 =	0.638
6	100 - 5 / 100 - 5=	1	42 - 2 / 85 - 2 =	0.481
7	100 - 5 / 100 - 5=	1	40 - 2 / 85 - 2 =	0.457
8	100 - 5 / 100 - 5=	1	53 - 2 / 85 - 2 =	0.614

9	100 - 5 / 100 - 5=	1	40 - 2 / 85 - 2 =	0.457
10	100 - 5 / 100 - 5=	1	43 - 2 / 8 5 - 2 =	0.493
11	50 - 5 / 100 - 5 =	0.473	14 - 2 / 85 - 2 =	0.144
12	10 - 5 / 100 - 5 =	0.052	19 - 2 /85 - 2 =	0.204
13	50 - 5 / 100 - 5 =	0.473	11 - 2 / 85 - 2 =	0.108
14	50 - 5 / 100 - 5=	0.473	23 - 2 / 85 - 2 =	0.253
15	10 - 5 / 100 - 5 =	0.052	14 - 2 / 85 - 2 =	0.144
16	10 - 5 / 100 - 5 =	0.052	22 - 2 / 85 - 2 =	0.240
17	50 - 5 / 100 - 5 =	0.473	23 - 2 / 85 - 2 =	0.253
18	50 - 5 / 100 - 5 =	0.473	30 - 2 / 85 - 2 =	0.337
19	20 - 5 / 100 - 5 =	0.157	22 - 2 / 85 - 2 =	0.240
20	50 - 5 / 100 - 5 =	0.473	32 - 2 / 85 - 2 =	0.361
21	50 - 5 / 100 - 5 =	0.473	17 - 2 / 85 - 2 =	0.180
22	50 - 5 / 100 – 5 =	0.473	13 - 2 / 85 - 2 =	0.132
23	10 - 5 / 100 - 5 =	0.052	7 - 2 / 85 - 2 =	0.060
24	50 - 5 / 100 - 5 =	0.473	11 - 2 / 85 - 2 =	0.108
25	50 - 5 / 100 - 5 =	0.473	17 - 2 / 85 - 2 =	0.180
26	20 - 5 / 100 - 5 =	0.157	14 - 2 / 85 - 2 =	0.144
27	10 - 5 / 100 - 5 =	0.052	7 - 2/ 85 - 2 =	0.060
28	5 - 5 / 100 - 5 =	0	2 - 2 / 85 - 2 =	0
29	50 - 5 / 100 - 5 =	0.473	30 - 2 / 85 - 2 =	0.337
30	50 - 5 / 100 – 5 =	0.473	26 - 2 / 85 - 2 =	0.289
31	100 - 5 / 100 - 5 =	1	35 - 2 / 85 - 2 =	0.397
32	100 - 5/ 100 – 5 =	1	31 - 2 / 85 - 2 =	0.349
33	20 - 5 /100 – 5 =	0.157	24 - 2 / 85 - 2 =	0.265
34	50 - 5 / 100 - 5 =	0.473	19 - 2 / 85 - 2 =	0.204
35	100 - 5 / 100 - 5 =	1	85 - 2 / 85 - 2 =	1
36	50 - 5 / 100 - 5 =	0.473	18 - 2 / 85 - 2 =	0.192
37	20 - 5 / 100 - 5 =	0.157	8 - 2 / 85 - 2 =	0.072
38	20 - 5 / 100 - 5 =	0.157	11 - 2 / 85 - 2 =	0.108
39	10 - 5 / 100 – 5 =	0.052	17 - 2 / 85 - 2 =	0.180
40	50 - 5 / 100 - 5 =	0.473	23 - 2 / 85 - 2 =	0.253

Calculate Euclidean Distance by specifying the parameter k, by calculating the distance between the testing data and the training data at the data transformation stage.

Table 2. Euclidean Distance Calculation Results

No.	Product name	Product Stock	ATK Sold	Euclidean Distance
1	Fountain pen	100	77	$\sqrt{1-(0)^2+(0{,}903\ -0)^2}$= 1,815
2	Pencil	100	46	$\sqrt{1-(0)^2+(0{,}530\ -0)^2}$= 1,280
3	Highlighter	100	45	$\sqrt{1-(0)^2+(0{,}518\ -0)^2}$= 1.268
4	Eraser	100	41	$\sqrt{1-(0)^2+(0{,}469\ -0)^2}$= 1,219
5	Ballpoint Pen	100	55	$\sqrt{1-(0)^2+(0{,}638\ -0)^2}$= 1,407
6	Whiteboard marker	100	42	$\sqrt{1-(0)^2+(0{,}481\ -0)^2}$= 1,231
7	Colored pencils	100	40	$\sqrt{1-(0)^2+(0{,}457\ -0)^2}$= 1.208
8	Gel Pen	100	53	$\sqrt{1-(0)^2+(0{,}614\ -0)^2}$= 1.376

9	Roller Pen	100	40	$\sqrt{1-(0)^2+(0{,}457\ -0)^2}$= 1.208
10	Whiteboard Marker	100	43	$\sqrt{1-(0)^2+(0{,}493\ -0)^2}$= 1.243
11	Zippered Map	50	14	$\sqrt{0{,}473-(0)^2+(0{,}144\ -0)^2}$= 0.493
12	Notebook	10	19	$\sqrt{0{,}052-(0)^2+(0{,}204\ -0)^2}$= 0.093
13	Label	50	11	$\sqrt{0{,}473-(0)^2+(0{,}108\ -0)^2}$= 0.484
14	Ring Folder	50	23	$\sqrt{0{,}473-(0)^2+(0{,}253\ -0)^2}$= 0.537
15	Paper hole puncher	10	14	$\sqrt{0,052-(0)^2+(0{,}337\ -0)^2}$= 0.165
16	Punch Hole	10	22	$\sqrt{0{,}052-(0)^2+(0{,}240\ -0)^2}$= 0.109
17	Staples	50	23	$\sqrt{0{,}473-(0)^2+(0{,}253\ -0)^2}$= 0.537
18	Glue	50	30	$\sqrt{0,473-(0)^2+(0{,}337\ -0)^2}$= 0.586
19	Scissors	20	22	$\sqrt{0,157-(0)^2+(0{,}240\ -0)^2}$= 0.214
20	Ruler	50	32	$\sqrt{0{,}473-(0)^2+(0{,}361-0)^2}$= 0.603
21	Stapler	50	17	$\sqrt{0{,}473-(0)^2+(0{,}180\ -0)^2}$= 0.505
22	Paperclip	50	13	$\sqrt{0{,}473-(0)^2+(0{,}132-0)^2}$= 0.490
23	Paper hole puncher	10	7	$\sqrt{0{,}052-(0)^2+(0{,}060\ -0)^2}$= 0.055
24	Pencil Case	50	11	$\sqrt{0{,}473-(0)^2+(0{,}108\ -0)^2}$= 0.484
25	Marker Eraser	50	17	$\sqrt{0{,}473-(0)^2+(0{,}180\ -0)^2}$= 0.505
26	Tissue box	20	14	$\sqrt{0{,}157-(0)^2+(0{,}144\ -0)^2}$= 0.177
27	Paperclip	10	7	$\sqrt{0{,}052-(0)^2+(0{,}060\ -0)^2}$= 0.055
28	Digital scales	5	2	$\sqrt{0-(0)^2+(0\ -0)^2}$= 0
29	Book Binder	50	30	$\sqrt{0{,}473-(0)^2+(0{,}337\ -0)^2}$= 0.586
30	Portfolio	50	26	$\sqrt{0{,}473-(0)^2+(0{,}289\ -0)^2}$= 0.556
31	Plastic Map	100	35	$\sqrt{1-(0)^2+(0{,}397\ -0)^2}$= 1.157
32	Cardboard Map	100	31	$\sqrt{1-(0)^2+(0{,}349\ -0)^2}$= 1.121
33	Agenda book	20	24	$\sqrt{0{,}157-(0)^2+((0{,}265\ -0)^2}$= 0.227
34	Clip Folder	50	19	$\sqrt{0{,}473-(0)^2+(0{,}204\ -0)^2}$= 0.514
35	Duty stamp	100	85	$\sqrt{1-(0)^2+(1\ -0)^2}$= 2
36	Paperboard	50	18	$\sqrt{0{,}473-(0)^2+(0{,}192\ -0)^2}$= 0.509
37	Map Accordion	20	8	$\sqrt{0{,}157-(0)^2+(0{,}072\ -0)^2}$= 0.162
38	Acco Folder	20	11	$\sqrt{0{,}157-(0)^2+(0{,}108-0)^2}$= 0.168

39	Steel Staples	10	17	$\sqrt{0{,}052-(0)^2+(0{,}180\ -0)^2}$= 0.084
40	Plain Map	50	23	$\sqrt{0{,}473-(0)^2+(0{,}253\ -0)^2}$= 0.537

Sort stationery sales data in ascending order

Table 3. Ascending Data

No..	Product name	Euclidean Distance	Ranking
1	Fountain pen	$\sqrt{1-(0)^2+(0{,}903\ -0)^2}$= 1,815	2
2	Pencil	$\sqrt{1-(0)^2+(0{,}530\ -0)^2}$= 1,280	5
3	Highlighter	$\sqrt{1-(0)^2+(0{,}518\ -0)^2}$= 1.268	6
4	Eraser	$\sqrt{1-(0)^2+(0{,}469\ -0)^2}$= 1,219	9
5	Ballpoint Pen	$\sqrt{1-(0)^2+(0{,}638\ -0)^2}$= 1,407	3
6	Whiteboard marker	$\sqrt{1-(0)^2+(0{,}481\ -0)^2}$= 1,231	8
7	Colored pencils	$\sqrt{1-(0)^2+(0{,}457\ -0)^2}$= 1.208	10
8	Gel Pen	$\sqrt{1-(0)^2+(0{,}614\ -0)^2}$= 1.376	4
9	Roller Pen	$\sqrt{1-(0)^2+(0{,}457\ -0)^2}$= 1.208	11
10	Whiteboard Marker	$\sqrt{1-(0)^2+(0{,}493\ -0)^2}$= 1.243	7
11	Zippered Map	$\sqrt{0{,}473-(0)^2+(0{,}144\ -0)^2}$= 0.493	30
12	Notebook	$\sqrt{0{,}052-(0)^2+(0{,}204\ -0)^2}$= 0.093	36
13	Label	$\sqrt{0{,}473-(0)^2+(0{,}108\ -0)^2}$= 0.484	26
14	Ring Folder	$\sqrt{0{,}473-(0)^2+(0{,}253\ -0)^2}$= 0.537	18
15	Paper hole puncher	$\sqrt{0,052-(0)^2+(0{,}337\ -0)^2}$= 0.165	33
16	Punch Hole	$\sqrt{0{,}052-(0)^2+(0{,}240\ -0)^2}$= 0.109	34
17	Staples	$\sqrt{0{,}473-(0)^2+(0{,}253\ -0)^2}$= 0.537	19
18	Glue	$\sqrt{0,473-(0)^2+(0{,}337\ -0)^2}$= 0.586	15
19	Scissors	$\sqrt{0,157-(0)^2+(0{,}240\ -0)^2}$= 0.214	29
20	Ruler	$\sqrt{0{,}473-(0)^2+(0{,}361-0)^2}$= 0.603	14
21	Stapler	$\sqrt{0{,}473-(0)^2+(0{,}180\ -0)^2}$= 0.505	23
22	Paperclip	$\sqrt{0{,}473-(0)^2+(0{,}132-0)^2}$= 0.490	25
23	Paper hole puncher	$\sqrt{0{,}052-(0)^2+(0{,}060\ -0)^2}$= 0.055	38
24	Pencil Case	$\sqrt{0{,}473-(0)^2+(0{,}108\ -0)^2}$= 0.484	27
25	Marker Eraser	$\sqrt{0{,}473-(0)^2+(0{,}180\ -0)^2}$= 0.505	24
26	Tissue box	$\sqrt{0{,}157-(0)^2+(0{,}144\ -0)^2}$= 0.177	32

27	Paperclip	$\sqrt{0{,}052-(0)^2+(0{,}060-0)^2}$= 0.055	39
28	Digital scales	$\sqrt{0-(0)^2+(0-0)^2}$= 0	40
29	Book Binder	$\sqrt{0{,}473-(0)^2+(0{,}337-0)^2}$= 0.586	16
30	Portfolio	$\sqrt{0{,}473-(0)^2+(0{,}289-0)^2}$= 0.556	17
31	Plastic Map	$\sqrt{1-(0)^2+(0{,}397-0)^2}$= 1.157	12
32	Cardboard Map	$\sqrt{1-(0)^2+(0{,}349-0)^2}$= 1.121	13
33	Agenda book	$\sqrt{0{,}157-(0)^2+((0{,}265-0)^2}$= 0.227	28
34	Clip Folder	$\sqrt{0{,}473-(0)^2+(0{,}204-0)^2}$= 0.514	21
35	Duty stamp	$\sqrt{1-(0)^2+(1-0)^2}$= 2	1
36	Paperboard	$\sqrt{0{,}473-(0)^2+(0{,}192-0)^2}$= 0.509	22
37	Map Accordion	$\sqrt{0{,}157-(0)^2+(0{,}072-0)^2}$= 0.162	35
38	Acco Folder	$\sqrt{0{,}157-(0)^2+(0{,}108-0)^2}$= 0.168	34
39	Steel Staples	$\sqrt{0{,}052-(0)^2+(0{,}180-0)^2}$= 0.084	37
40	Plain Map	$\sqrt{0{,}473-(0)^2+(0{,}253-0)^2}$= 0.537	20

From the results of the k data test where the value of k = 3 was taken, the best-selling ATK sales result with a value of 2 was stamps.

CONCLUSION

Based on the results of the calculation data using K-Nearest Neighbor, the most sold stationery product within 6 months at the Berkah ATK Store is stamps.

REFERENCES

[1] S. P. Dewi, Nurwati dan E. Rahayu, "Penerapan Data MiningUntuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Building of Informatics, Technology and Science (BITS) ,* vol. Volume.3 Nomor.4, p. 639–648, 2022.

[2] I. Yolanda dan H. Fahmi, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT.Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor," *JIKOMSI [Jurnal Ilmu Komputer dan Sistem Informasi],* vol. Volume.3 Nomor.3, pp. 9-15, 2021.

[3] A. U. Haspriyanti dan P. W. Prasetyaningrum, "Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-Nearst Neighbor," *JURNAL INFORMATION SYSTEM & ARTIFICIAL INTELLIGENCE,* vol. Volume 1, pp. 100-107, 2021.

[4] Rismala, I. Ali dan A. R. Rinaldi, "PENERAPAN METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI PENJUALAN SEPEDA MOTOR TERLARIS," *JATI (Jurnal Mahasiswa Teknik Informatika),* vol. Volume.7 No.1, pp. 585-590, 2023.

[5] C. Anisa dan Andri, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PREDIKSI PENJUALAN OBAT PADA APOTEK KIMIA FARMA ATMO PALEMBANG," *Bina Darma Conference on Computer Science,* pp. 199-208, 2020.

[6] A. Pratama, B. S. Ginting dan Nurhayati, "PENERAPAN DATA MINING UNTUK PREDIKSI MEREK PAKAIAN YANG PALING DIMINATI DENGAN METODE K-NEAREST NEIGHBOR (STUDI KASUS : PT. MATAHARI DEPARTEMENT STORE BINJAI)," *Jurnal Pancabudi,* vol. 14 No.2, pp. 54-64, 2021.

[7] Cholil, Saifur Rohman; Handayan, Titis; Prathivi, Rastri; Ardianita, Tria;, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian Journal on Computer and Information Technology) 6,* vol. 6 No.2, pp. 118-127, 2021.

[8] Putri, Vania Ariyani Prilia; Prasetijo, Agung Budi; Eridani, Dania;, "PERBANDINGAN KINERJA ALGORITME NAÏVE BAYES DAN KNEAREST NEIGHBOR (KNN) UNTUK PREDIKSI HARGA RUMAH," *JURNAL ILMIAH TEKNIK ELEKTRO,* vol. 24 (4), pp. 162-171, 2022.

[9] Mubarok, Muhammad Iqbal; , Purwantoro; , Carudin;, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DALAM KLASIFIKASI PENILAIAN JAWABAN UJISAN ESAI," *JATI(Jurnal Mahasiswa Teknik Informatika),* vol. 7 (5), pp. 3446-3452, 2023.

[10] Erika, Winda;, "Application of Data Mining to Predict Sales of the Best-Selling Dolls at Chudalla Store Using the K-Nearest Neighbor (K-NN) Method," dalam *International Conference Sciences Development and Technology*, Medan, 2023.