Prediction of School Uniform Sales Using the K-Nearest Neighbor Method

Nurwati[*1], Yudi Santoso[2], Nur Azizah[3]
[1,2]Information Systems Study Program, Faculty of Information Technology, Budi Luhur University
[3]Information Systems Study Program, Faculty of Science and Technology, Raharja University
Email: [*1]**nurwati@budiluhur.ac.id**, [2]yudi.santoso@budiluhur.ac.id, [3]nur.azizah@raharja.info

Abstract

Information regarding monitoring of operational expenditures of the Regional Public Service Agency at the Emergency Ambulance Service Unit of the DKI Jakarta Health Office. Because the finances issued or used must be accurately informed. In this case, monitoring the operational expenses of the Regional Public Service Agency at the Emergency Ambulance Service Unit of the DKI Jakarta Health Service still has many obstacles. Constraints faced are such as the technical executor. The activity of submitting an accountability letter still has a long stage, the Verifier is not there so that the process of checking the accountability letter stops, the input of reports is carried out one by one by the Spending Treasurer using Microsoft Excel. The process of monitoring the operational expenses of the Regional Public Service Agency at the Emergency Ambulance Service Unit of the DKI Jakarta Health Office takes a long time. And another obstacle is the occurrence of classifying the type of operational expenditure. In this data analysis method, using Balanced Scorecard (BSC) analysis, system design uses Visual Paradigm to create UML (Unified Modeling Language), PHP programming language, MySQL to create databases. The result of this research is that an information system for monitoring operational expenditures of the Regional Public Service Agency can create and submit a letter of accountability and verification letters are carried out computerized, the Expenditure Treasurer no longer enters the general cash books one by one, the Expenditure Treasurer no longer classifies the types of operational expenditure in recap operational expenditure.

Keywords *— Monitoring, Expenses, Operational*

1. INTRODUCTION

After more than 2 (two) years of being active and working from home due to the covid-19 pandemic, as time goes on the number of patients affected by COVID-19 has decreased so that people's activities have gradually improved as before the COVID-19 pandemic hit. The wheels of the economy began to slowly see transaction activities. Traditional markets and modern markets are starting to see sales activities, schools are starting to hold limited face-to-face learning (PTMT), offices are starting to open their business activities with a limited number of employees coming, and so on.

This Limited Face-to-face Learning Decision refers to "on May 11, 2022, the Minister of Education, Culture, Research, and Technology (Mendikbudristek), Minister of Religion (Menag), Minister of Health (Menkes), and Minister of Home Affairs (Mendagri) issued a Joint Decree (Four Ministerial Decree) Number 01/KB/2022, Number 408 of 2022, Number

HK.01.08/MENKES/1140/2022, Number 420-1026 of 2022 concerning Guidelines for Implementation of Learning in the Time of the COVID-19 Pandemic" [1] . With this 4 (four) ministerial decree, it brings fresh air to school uniform sellers , both online and offline sellers.

One of the online uniform sellers is the Mutiara Uniform shop with the website address www.nurwati.com which has been established since 2013. The profile of this Mutiara Uniform shop is " Mutiara Uniform online as an online school uniform shop has the goal of making a profit, has determined that business school uniform sales have a great opportunity in online sales. However, the current condition is not significant when viewed from the overall sales revenue of Mutiara Uniform" [2] . Mutiara Uniform Store provides school uniforms ranging from kindergarten, elementary, middle and high school as well as other school supplies. Even though the Mutiara Uniform Store already uses a desktop-based sales application, to see what products are most in demand, the data is still not optimal because entering product inventory data into the application is still negligent. So that sales predictions are needed to plan and schedule stock supplies of school uniform products in the future.

Analysis predicts sales in this study using data mining with the K-Nearest Neighbor method. Forecasting or "prediction or forecasting is the art and science of knowing or predicting what will happen in the future. Prediction becomes very important because of the preparation of a plan, one of which is based on a projection or forecast. The predictions made by the company aim to find out and estimate the number of future sales and the number of prediction errors, so that they can meet the needs of consumers and company management" [3] . Then using "data mining is the process of extracting data from very important information. Data mining is also a process of exploring patterns from data. Patterns are obtained from various types of databases, such as relational databases, data warehouses, transactional data, and object-oriented data. The use of data mining can help entrepreneurs make decisions quickly and accurately" [4] . The method used in predicting school uniform sales products is the K-Nearest Neighbor method, "k-NN is a method that uses a guided algorithm where the results of the new query instance are classified based on the majority of the categories in k-NN" [5] .

The next step is processing data with Microsoft Excel " Microsoft Excel is a software program that allows users to process and calculate numerical data (numbers). Data processing is carried out using formulas in spreadsheet sheets. Data in Microsoft Excel spreadsheets can be calculated and processed accurately using certain available formulas. In addition, there are several tools in Microsoft Excel that are able to present processed data results" [6] .

The problem formulated is how to predict the sales of school uniforms at Mutiara Uniform stores in the coming year so that they can prepare uniform products early.

As a source of reading in this study are as follows:

Table 1. Research Related to School Uniform Sales Predictions Using the K-Nearest Neighbor Method

Paper	Research purposes	Research result
1	Predict sales of clothing brands that are most in demand by consumers. This prediction aims to facilitate the company's management at PT. Matahari Department Store Tbk Binjai in planning the supply of clothing stock and can provide accurate information to companies about clothing brands that are most in demand by consumers [7] .	Produced a prediction system for the most desirable clothing brand using the K-Nearest Neigbor method, so that PT. Matahari Department Store Tbk Binjai benefited from planning and scheduling the supply of clothing stock for the future accurately. By calculating data mining using classification techniques with the K-Nearest Neighbor algorithm which is the majority, it can be predicted that the number of sales in the next period will increase with an average prediction of 14,900 per month and the most desirable clothing brand, Cardinal.
2	Meeting the wants and needs of customers and Pottery MSMEs also needs to pay attention to the production capacity they have, minimize waste products and be able to determine the amount of pottery production. This study will use a comparison of 3 data mining methods such as Decision Tree, Naive Bayes, K-NN to determine the amount of production to meet consumer demand. Applying the comparison of 3 methods in data mining in determining the amount of pottery production, it is hoped that MSMEs can overcome fluctuations in consumer demand with minimal production costs [8] .	Of the three Decision Tree, Naive Bayes and KNN methods, the method with the highest accuracy is the Naive Bayes method because the Naive Bayes method only ticks the laplacce correction parameters and results in sales predictions of 100.00%, for the Decision tre method changes the criterion value to information_gain, minimum leaf size becomes 1, the minimum leaf size for split becomes 1, the number of prepruning alternatives becomes 1, the sales prediction result is 75.00%, and for the KNN method changes the K value to 9, the sales prediction result is 90.00% .
3	Predict sales to see the potential of new customers and goods that customers like [9] .	To use training data which amounted to 99 data by testing testing data totaling 25 data, then the best k value to predict student study period was as follows: For the 100th buyer, the best k value to be used to predict the type of shoes to be purchased is value of k = 9 with a success rate of 49%. For the 100th buyer, the best value of k used to predict the purchase of this type of shoe is the value of k = 7 with a success rate of 74%. For the 100th buyer, the value of k = 1 is the best k value to be used to predict the student's study period with a success rate of 89%.

2. RESEARCH METHOD

The research method used in this study uses quantitative methods. This quantitative method "helps and directs in a research in the form of solving a problem in the form of research so that the achievement of the research objectives is more well-directed" [10].

This research begins with identifying problems by collecting sales data for 2020 and 2021, then the results of this initial stage are used to formulate problems. Furthermore, data

analysis is processed using the K-Nearest Neigbor (K-NN) method. The results of the K-NN were tested again using Microsoft Excel. The next stage is decision making. The stages of this research are as shown in Figure 1 below:

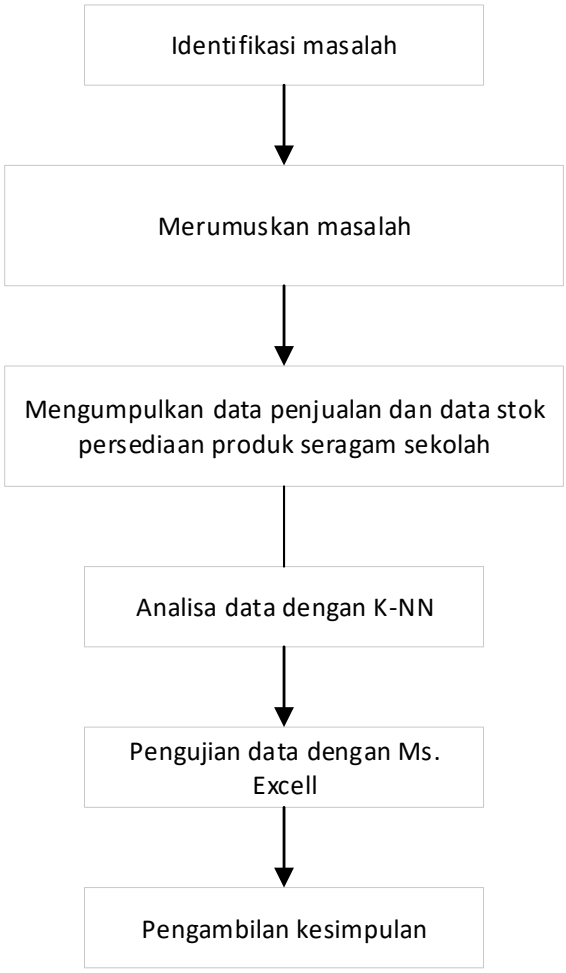

Figure 1. Research Stages

3. RESEARCH RESULTS AND DISCUSSION

3.1. Results

The product data used is sales product data for 2020 and 2021. With 19 (nineteen) types of products, namely short white shirts, rempel short red skirts, short red pants, short blue skirts, long blue pants, short gray skirts. ,span gray trousers, long white shirt, long red rempel skirt, long red pants, long black pants, short brown pants, long brown pants, scout short shirts, scouts long shirt, long gray rempel skirt, white skirt short rempel, long white skirt rempel and long black skirt rempel. The data was processed using the K-NN method with the Ms Excel application.

3.2. Discussion

a. Data Selection

The data used is sales data for 2020 and 2021 at the Mutiara Uniform store. Then the data is selected and processed to predict sales in 2022. The attributes used are product name, quantity and month, which can be seen in table 2 data selection:

Table 2. Data Selection

No	Name of goods	Quantity	Month
1	short white shirt	10	January
2	ruffled short red skirt	5	January
3	short blue skirt span	5	January
4	short white shirt	5	January
5	short white shirt	6	January
6	long red pants	5	January
7	short blue skirt span	12	January
8	short white shirt	20	January
9	short white shirt	10	January
10	scout short shirt	5	January
11	short gray skirt span	4	January
12	short white shirt	10	January
13	long red pants	5	January
14	short blue skirt span	4	January
15	short gray skirt span	4	January
...	.....	.....	.....
100	short white shirt	5	December

b. Preprocessing

At this stage "grouping the types of school uniform products based on the number of sales each month and year to facilitate the process of calculating predictions" [7] . Then after the data is grouped by type of product it will be able to see the number of sales. It can be seen in table 3 below;

Table 3. Preprocessing Data for Short White Clothes

Month	2020	2021
January	80	30
February	54	22
March	10	45
April	10	19
May	12	10
June	22	8
July	79	46
August	32	5
September	10	6
October	0	6
November	0	4
December	5	0

In table 3, total sales of 515 items are sold every month and year.

Table 4. Preprocessing Data for the Red Short Rempel Skirt

Month	2020	2021
January	10	8
February	8	2
March	0	5
April	0	9
May	12	10
June	17	8
July	5	20
August	4	0
September	1	6
October	5	6
November	0	0
December	5	4

In table 4, the results of the sale of short red rempel skirts are obtained every month and year with a total sale of 145 items.

Table 5. Preprocessing Data for Short Blue Skirt Span

Month	2020	2021
January	25	2
February	6	5
March	6	5
April	6	6
May	12	0
June	11	2
July	10	10
August	5	2
September	0	3
October	5	0
November	0	3
December	5	3

In table 5, the sales results of the short blue skirt span every month and year with a total sales of 132 items.

Table 6. Preprocessing Data for Long Red Pants

Month	2020	2021
January	20	8
February	4	4
March	4	6
April	6	0
May	4	0
June	7	2
July	9	6
August	5	2
September	0	3
October	0	3
November	4	0
December	0	0

In table 6, the results of the sale of long red trousers are obtained every month and year with a total sale of 97 items.

Table 7. Preprocessing Data for Scout Short Clothes

Month	2020	2021
January	6	4
February	2	3
March	0	0
April	0	0
May	4	0
June	2	0
July	5	1
August	5	0
September	2	0
October	0	0
November	4	0
December	0	0

In table 7, the results of the sale of short scout clothes are obtained every month and year with a total sale of 38 items

Table 8. Preprocessing Data of Short Gray Skirt Span

Month	2020	2021
January	0	2
February	2	3
March	0	0
April	0	0
May	4	0
June	2	1
July	5	1
August	2	2
September	0	0
October	0	0
November	4	0
December	0	0

In table 8, the results of the sale of a short gray skirt span every month and year with a total sale of 28 items

c. Transformation

This stage is called the transformation stage, namely "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 sales" [7] . This attribute will affect the prediction of the sale of the most desirable school uniform product and the data is called the target data ADDIN CSL_CITATION {"citationItems":[{"id":"ITEM-1","itemData":{"author":[{"dropping-particle":"","family":"Pratama","given":"Andrean","non-dropping-particle":"","parse-names":false,"suffix":""},{"dropping-particle":"","family":"Ginting","given":"Budi Serasi","non-dropping-particle":"","parse-names":false,"suffix":""}],"container-title":"Jurnal Abdi Ilmu","id":"ITEM-1","issue":"2","issued":{"date-parts":[["2021"]]},"page":"54-64","title":"Penerapan Data Mining Untuk Prediksi Merek Pakaian Yang Paling Diminati Dengan Metode K-Nearest Neighbor (Studi Kasus : Pt . Matahari Departement Store Binjai)","type":"article-

journal","volume":"14"},"uris":["http://www.mendeley.com/documents/?uuid=6f34cda0-fd98-468c-8ed7-56f2520494e7"]}],"mendeley":{"formattedCitation":"[7]","plainTextFormattedCitation":"[7]","previouslyFormattedCitation":"[7]"},"properties":{"noteIndex":0},"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}. The training data can be seen in table 9 below;

Table 9. Sales Training Data for Short White Clothes

No	month1	month2	month3	month4	month5	Target
1	80	54	10	10	12	22
2	54	10	10	12	22	79
3	10	10	12	22	79	32
4	10	12	22	79	32	10
5	12	22	79	32	10	0
6	22	79	32	10	0	0
7	79	32	10	0	0	5
8	10	0	0	5	30	22
9	0	0	5	30	22	45
10	0	5	30	22	45	19
11	5	30	22	45	19	10
12	30	22	45	19	10	8
13	22	45	19	10	8	46
14	45	19	10	8	46	5
15	19	10	8	46	5	6
16	10	8	46	5	6	6
17	8	46	5	6	6	4
18	46	5	6	6	4	0

Explanation of table 9, the training data used takes data on sales of uniform products in 2020 and 2021. “The training data is grouped into 2, namely input and target data. Where the input data is sales data from the 1st month to the 5th month, while the target data uses the 6th month data. The next input data uses data from the 2nd month to the 6th month and the 7th month as the target and so on, until the last data available” [11] .

3.3. Data analysis with K-NN

The next step is data analysis using the K-NN method with the following steps:

a. Determination of the value of k. The value of k used in this study is 3.
b. Calculate the distance between the training data and the test data in the transformation stage using the Euclidean Distance calculation. Existing test data (5,10,5,6,8,?) are as follows;
c. Stages of sorting the calculated data. The distance value that has been obtained is then sorted in ascending order, which is sorted from the closest distance value to the farthest distance value. The order is as follows: d8=25,19; d9=29.95; d17=36,18; d16=41.41; d18=41.51; d13=41.54; d15=42.19; d11=48,28; d10=48.73; d12=50,42; d2=51.55; d14=56,16; d3=73,28; d6=76.54; d4=78,88; d5=79.68; d1=87.28.
d. Determining the test data group based on the majority label from the k closest neighbors with a value of k = 3, the closest distance is d8 = 25.19; d9=29.95 and d17=36.18

e. By using the K-NN method, it can be predicted that sales of school uniform products will increase in the next period for short white shirts. For the sales prediction, only 2 (two) data are taken with the closest distance, namely d8 = 25.19 and d9 = 29.95 occurring between August and September. For the prediction of the increase in sales of short red rempel skirts, d12=3.87 and d9=8.18 occurred between December and September. For the prediction of the increase in sales of short blue skirts, the span is d2=7.93 and d4=8.12 between February and April. For the prediction of the increase in sales of red trousers is d3=4.47 ; d2=6.55 occurs between March and February. For the prediction of increased sales of scout shorts are d4=8.94; d7=9.27 occurs between April and June. For the prediction of increased sales of short gray skirts, the span is d4=10.34; d10=11.26 occurs between April and October.

f. The predictive value of selling short white shirts is 36.36%

4. CONCLUSION

Conclusion obtained is that the prediction of school uniform product sales using the K-Nearest Neighbor method can help Mutiara Uniform stores plan stock supplies for school uniform products in the future. By using the K-Nearest Neighbor category, the predictive value of selling short white shirts in 2022 is 36.36%.

5. SUGGESTED

As for suggestions that can be given for the development and improvement of this research, it is hoped that further development can be carried out by designing a school uniform product sales prediction application, by adding the number of product samples, varying k values and sales time spans for prediction.

6. REFERENCES

[1] Middle School Directorate Web Manager, "Latest Adjustment of Limited PTM According to the 4 Ministerial Decree," *https://ditsmp.kemdikbud.go.id/* , May 20, 2022.

[2] Y. Santoso and N. Nurwati, “Building a Business Model on Mutiara Uniform Online with a Business Model Canvas Approach,” *ICIT J.* , vol. 8, no. 1, pp. 79–86, 2022, doi:10.33050/icit.v8i1.2175.

[3] MM Hasmawati, Jumadil Nangi, "Application of Goods Sales Prediction Using the K-Nearest Neighbor (Knn) Method," vol. 3, no. 2, pp. 151–160, 2017.

[4] G. Aprilianur and EL Hadisaputro, “Application of Data Mining Using K-Means Clustering Method for Myam Hijab Penajam Store Sales Analysis,” *J. JUPITER* , vol. 14, no. 1, pp. 161–170, 2022.

[5] RR Waliyansyah and C. Fitriyah, "Comparison of the Accuracy of Teak Image Classification Using Naive Bayes and k-Nearest Neighbor (k-NN) methods," *J. Education and Researchers. information.* , vol. 5, no. 2, p. 157, 2019, doi:10.26418/jp.v5i2.32473.

[6] Z. Hardiansyah, “Knowing Microsoft Excel and its Benefits and Formulas,” *https://tekno.kompas.com/* , p. Kompas.com Tekno Apps & OS This article was published, Feb. 10, 2022.

[7] A. Pratama and BS Ginting, “The Application of Data Mining to Predict the Most Interested Clothing Brand Using the K-Nearest Neighbor Method (Case Study: Pt. Matahari Department Store Binjai),” *J. Abdi Sains* , vol. 14, no. 2, pp. 54–64, 2021.

[8] N. Nosiel, S. Sriyanto, and F. Maylani, "Comparison of Data Mining Techniques for Sales Prediction in Pottery MSMEs," *Pros. Semin. Nas. Darmajaya* , vol. 1, no. 0, pp. 72–86, 2021.

[9] B. Hardiyanto and F. Rozi, “Prediction of shoe sales using the k-nearest neighbor method,” vol. 04, pp. 13–18, 2020.

[10] BG Sudarsono and SP Lestari, “Clustering of Foundation Scholarship Recipients for Students Using the K-Means Method,” *J. Media Inform. Budidharma* , vol. 5, no. 1, p. 258, 2021, doi:10.30865/mib.v5i1.2670.

[11] M. Arifanto and E. Santoso, “Implementation of the K-Nearest Neighbor Method for predicting sales of skincare packaging at PT. Universal Jaya Perkasa,” *Technologic* , vol. 10, no. 8, pp. 1–9, 2015.

[12] F. Hermawan and H. Agung, “Implementation of the K-Nearest Neighbor Method in the Sales Data Application of PT. Multitek Mitra Sejati,” *J. Science and Technology.* , vol. volume 4 n, pp. 103–109, 2017, [Online]. Available: http://research.kalbis.ac.id/Research/Files/Article/Full/CDRHLKPOJ7STTQHS1YJF02JX.pdf.