Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

Solar Purchase Volume Prediction Using The K-Nearest Neighbor Algorithm Based On Backward Elimination

Aries Alfian Prasetyo[1*], Yudi Pramono[2], Laily Ulfiyah[3], Misbakhul Fattah[4]
[1,3,4] Madura State Polytechnic, Indonesia, [2] Bojonegoro Regency, Indonesia,
[1]aries@poltera.ac.id, [2]ypramono@gmail.com, [3]laily@poltera.ac.id, [4]misbakhulfatah@poltera.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 16-11-2024
Accepted: 28-11-2024
Published: 30-11-2024

Keywords:
Solar, Volume, K-Nearest Neaighboor, Backward Elimination, Prediction

ABSTRACT

The benefits obtained by gas stations from fuel purchases each period and sales according to volume. With the prediction method, turnover increases with efficient purchases. Using the k-NN algorithm in data prediction. With secondary data of liter volume in the period January 2020-November 2024. The k values used are k = 1, K = 4, k = 5 and k = 7. Before calculating with k = 1, 70% training data and 30% testing data are determined. Then the initial cluster of interval class training data. cluster on testing data with K = 4, k = 5 and k = 7. Analysis using K-NN 2,3,4 and 5 periods produces the 3rd smallest K, so that the 3rd will be measured with K-NN backward elimination. Produces backward elimination predictions, attribute weights xt - 3 and xt 1 which are the reference for prediction, because their weights are 1. K = 13 is the best K for carrying out the Analysis process with K-NN to predict fuel purchase volume, with a K = 4 value of 45556,788. So the best prediction of fuel oil volume data, diesel type, K is K = 13 with the K-NN Analysis method with the backward elimination process. The results above show that xt3 or week 3 and week xt1 to 1 in the last period of 2024 can be used as a reference for purchases in the following year, namely 2025.

INTRODUCTION

The profits obtained by a Public Fuel Filling Station come from purchasing fuel each period and selling according to the volume. So the right purchasing volume will determine your monthly turnover. However, profits or turnover are often irregular, there are several causes, including fuel prices which tend to change, order volumes that do not match consumer demand (R. Risdiyanta., 2022). These changes affect the amount of profit obtained, so a prediction method is needed to determine the turnover that will be received by gas stations. With this prediction method, it is hoped that the turnover obtained will increase with more efficient purchasing. (Fitria et al., 2023).

There are still many business people who do not think about efficiency in increasing business profits by making sales plans or predictions based on goods sales transactions that have already been carried out in order to minimize the company's investment expenditure. (Amelia S et al., 2023).

In this research, efforts were made to process data on the results of fuel purchases, especially diesel fuel, using one of the data mining methods, namely using the k-NN method The problem studied in this research is how to predict fuel purchase result data using the k-NN algorithm with attribute selection.

Several studies on the k-NN algorithm have been carried out by several researchers. Research has been carried out on grouping Indonesian language documents using the k-NN algorithm. Classification of texture images using k-NN based on feature extraction using the co-occurrence matrix method has also been carried out (Hassani, H., Beneki et al., 2020). The International Conference on Data Mining (ICDM) lists the k-NN algorithm as one of the best algorithms in the Top 10 algorithms in data mining. These algorithms include C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, k-NN, Naive Bayes, and CART (L. Xu, C. Jiang, et al., 2014).

This research was conducted to study the k-NN algorithm and then apply the k-NN algorithm in data prediction. The data used is secondary data in the form of data on the number of fuel purchases in liter volumes in the period January 2020 - November 2024.

The k values used are k=1, K=4, k=5 and k=7. Before the calculation with k=1 is carried out, training data and testing data are determined. In this study, 70% of the training data was determined and 30% of the testing data was determined. Then the initial cluster is determined from the training data based on interval classes (MA. Lusiandro, SM, et al., 2020). Meanwhile, clusters in testing data are determined based on testing with K=4, k=5 and k=7. The next process is calculating the percentage value of cluster members obtained from the data testing process for the values K=4, k=5 and k=7.

LITERATURE REVIEW

The method used is the new K-nearest functionality (KNN) for real-time online forecasting. The result is found that KNN is especially effective for forecasting heterogeneous auction populations.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

Shortly after the implementation of the fuel price increase in early September 2022, the Twitter timeline was flooded with the keywords "BBM naik" (fuel oil is rising); the emergence of many positive and negative opinions to be interesting to study. This study aims to determine the results of sentiment analysis, the results of accuracy precision and recall from data ratio 60:40 and 70:30 using the K-Nearest Neighbor algorithm. This study used the K-Nearest Neighbor (KNN) algorithm because KNN handled training data more accurately. The results of the classification using the KNN algorithm after testing got a percentage of 66.67% for accuracy, 73.02% for negative precision, 48.89% for positive precision, 80.00% for negative recall, and 39.29% for positiverecall as a result from a data ratio of 60:40. While at a data ratio of 70:30, the results obtained were 70.31% for accuracy, 76.92% for negative precision, 54.05% for positive precision, 80.46% for negative recall, and 48.78% for positive recall. The results showed that there were 65.7% negative opinions and 34.3% positive opinions, this is because it can harm people who use vehicles in their daily activities. (Chrisley H et al., 2022).

Village potential data in this research using data mining methods using K-Means and the new GAP indicator model to select which indicators could be used as priority development indicators compared to other indicators. For evaluation and validation, a Neural Network model is used to predict whether the recommendation results will have an impact on increasing village status. Obtained a class recall for Developing Villages of 98.53% and 98.16% for predictions for Advanced Villages with accuracy in this model of 98.72%. (Prasetyo, A, et al., 2022).

The Neural Network Algorithm is an information processing system designed to imitate the way the human brain works in solving a problem that can be used to predict fuel sales. By using rapidminer software, the use of the Neural Network Algorithm method can be applied to PT. Dempo Migas/SPBU 24.31550, the number of records is 4324, the results of processing using the Neural Network Algorithm are obtained for sales in 2022, it is predicted that Biosolar sales will increase, this is seen from the results of processing with a data comparison, 60:40, 70:30 and 80:20, biosolar fuel has increased with an accuracy value of 99.77%, 99.60% and 99.83%., (Veronika, J et al., 2022).

Poverty is a problem that must be faced by the Government, poverty can also affect criminal acts. Therefore, special attention is needed to reduce poverty rates, the Government has made efforts to reduce poverty rates, including by providing various kinds of assistance to the poor based on the data obtained. In addition, the Government also needs to pay attention to the poverty index in each province, this aims to find out information on the poverty index within a certain time. This study discusses the prediction of the poverty index using the K-Nearest Neighbor (KNN) method in predicting the poverty index in each Province using data obtained from the Central Statistics Agency (BPS). (Faisal, M, et al., 2022).

PT. Kalonika Bara Kusuma adds or reduces units to adjust to the amount of turnover obtained in the previous month. However, after being evaluated, it turns out that this method is not effective. The results of the calculation above using multiple linear regression, then what affects the use of coal SFC is the use of BBM 140.162, SFC 250400, so it can be seen that the results of BBM use are 249088.5216. in designing data mining based on adopting the Multiple Linear Regression method can be used in solving problems (Augie, et al., 2024). The closest value of the largest case will be used as a reference by a decision maker in making decisions.

In this study, the patient's health history dataset will be analyzed using the Random Forest Tree and KNN algorithms. Both algorithms will be compared to find the most accurate model to implement in order to determine patterns in cervical cancer patients and predict patient screening results whether they are positive for cervical cancer or negative. The results of this study were processed using Python programming code for 214 test data from a total of 854 data. The final accuracy was shown to be 88.7% for Random Forest and 90.6% for KNN. The dataset used has four target classifications which are multilabel classifications. KNN has proven to be more advanced in predicting multilabel classifications in detecting patterns of cervical cancer cases (Andrian, et al., 2020).

So in this research a prediction analysis will be carried out using the K-Nearest Neighbor (KNN) data mining prediction method to find out how much accuracy there is in predicting the turnover of Fuel Filling Stations.

Algoritma K-Nearest Neighboor

The k-nearest neighbor (KNN) method algorithm is very simple, it works based on the shortest distance from the query instance to the training sample to determine the KNN. Training samples are projected into a multidimensional space, where each dimension represents a feature of the data. This space is divided into sections based on training sample classification `(Gasparetto, A., Marcuzzo, et al.2022). A point in this space is characterized by a class if class (c) is the most frequently found classification in K (the nearest neighbor of the point). Near or far neighbors are usually calculated based on Euclidean Distance. Euclidean distance is most often used to calculate a distance.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

K-Nearest Neighbor technique by carrying out steps, starting from input: Training data, label training data, k, testing data

1. Determine the parameter K (number of closest neighbors).
2. Calculate the square of the Euclidean distance of each object to the given sample data.
3. Sort the objects into groups with the smallest distance.
4. Collect category Y (nearest neighbor classification).
5. With the most nearest neighbor categories, the value of the query instance that has been calculated can be predicted.

To calculate the distance between two points x and y, you can use the Euclidean distance as follows:

$$d(X_1, Y_2) = \sum_{I} \left| \frac{n_{1i}}{n1} - \frac{n_{2i}}{n_2} \right| \quad \ldots\ldots (1)$$

Which one X_1, l = 1, 2 is a category attribute, and n_{1j}, n_1 represents the corresponding frequency. The k-NN algorithm is one method that uses a supervised algorithm. While in unsupervised learning, the data does not have any pattern, and the purpose of the unsupervised learning algorithm is to find patterns in the data. The purpose of the k-NN algorithm is to classify new objects based on attributes and training samples. Where the results of the new test sample are classified based on the majority of categories in k-NN (James G, et al. 2013). The k-NN algorithm uses neighborhood classification as a predictive value of the new test sample. The distance used is Euclidean Distance. Euclidean distance is the most commonly used distance in numeric data..

$$D(x_i, y_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(y_j))^2} \qquad 2.1$$

Information:

- R $D(x_i, y_j)$: *Euclidean distance*
- (x_i) : *record* th -i
- (x_j) : *record* th - j
- (a_r) : data th-r
- (i, j) : 1,2,3…n

The k-NN algorithm is between testing data training an algorithm that determines the distance value data based on the smallest value of the nearest neighbor value.

Backward Elimination

Create a model by entering all the variables then removing them one by one by testing the parameters using the partial F test. The smallest partial F-test (FL) value is compared to the F0 table (M. Arhami et al., 2020).

- If FL < F0, then the X in question is removed from the model and continues with creating a new model without that variable
- If FL>F0, then the process is stopped and the last equation is used/selected.

In accordance with the problems above, the framework for thinking used is:

Problem
How to predict the volume of Diesel Fuel purchases in each period so that turnover increases.
Objective
Provide recommendations for purchasing decisions regarding general diesel fuel so that turnover increases
Metodh
K-Nearest Neighbor + *Backward Elimination*
Results
Recommendations for making general fuel purchasing decisions to increase turnover

Figure 1. Framework for thinking

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

METHOD

The object of the research is the Syirkah Amanah Mandiri Public Fuel Filling Station in the Balen District area, Bojonegoro Regency, where Bojonegoro was chosen because of the easily accessible location and direct request from management, making it easier to collect data.

The profits obtained by a Public Fuel Filling Station come from purchasing fuel each period and selling according to the volume. So the right purchasing volume will determine your monthly turnover. However, profits or turnover are often irregular, there are several causes, including fuel prices which tend to change, order volumes that do not match consumer demand. These changes affect the amount of profit obtained, so a prediction method is needed to determine the turnover that will be received by.

Using data mining methods, namely the K-Nearest Neighbor (K-NN) prediction algorithm, to find decisions from multicriteria. The benefit of the KNN algorithm is to classify new objects based on attributes and example data. The results of the new test samples are classified based on the majority of categories in KNN (Statsoft, at al 2019). In the classification process, this algorithm does not use any model to match and is only based on memory.

In data collection there are data sources, data sources collected directly by researchers are called primary sources, whereas if they go through second hand they are called secondary sources. The data obtained is secondary data because it was obtained from the gas station sales database owned by PT. Syirkah Amanah Mandiri is located in Balen District, Bojonegoro Regency. The data obtained in this research are qualitative and quantitative data. The data used is time series data on purchasing and sales reports at gas stations. From the time series data, the main indicators that will be used as a benchmark are purchase volume, type of purchase, month of purchase, profit obtained. Apart from that, collecting library documentation; in the form of collecting information through books, magazines, articles and information on sites on the Internet.

Table 1. Recap Table of Fuel Purchases 2020 – 2024

MONTH	DATE	QTY PURCHASE (Liter)	UNIT PRICE (Rp)	AMOUNT (Rp)
January 2022	1-15 January	180.000	4.350	783.000.000
	16-31 January	212.000	4.350	922.200.000
February	1-15 February	204.000	4.310	879.240.000
	16-29 February	212.000	4.310	913.720.000
March	1-15 March	274.000	4.350	1.191.900.000
	16-31 March	302.000	4.350	1.313.700.000
January 2023	1-15 January	195.000	4.295	837.525.000
	16-31 January	205.000	4.295	880.475.000
MONTH	**DATE**	**QTY PURCHASE (Liter)**	**UNIT PRICE (Rp)**	**AMOUNT (Rp)**
February	1-15 February	225.000	4.295	966.375.000
	16-28 February	231.000	4.295	992.145.000
March	1-15 March	290.000	4.295	1.245.550.000
	16-31 March	294.000	4.295	1.262.730.000
	16-31 July	254.000	5.295	1.344.930.000
December	1-15 December	120.000	5.263	631.560.000
	16-31 December	128.000	5.263	673.664.000
January 2024	1-15 January	210.000	5.263	1.105.230.000
	16-31 January	206.000	5.263	1.084.178.000

The research instruments used include interview guides, field notes, observation guides, recording equipment and writing tools.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

The research that will be carried out refers to the framework as in Figure 2.

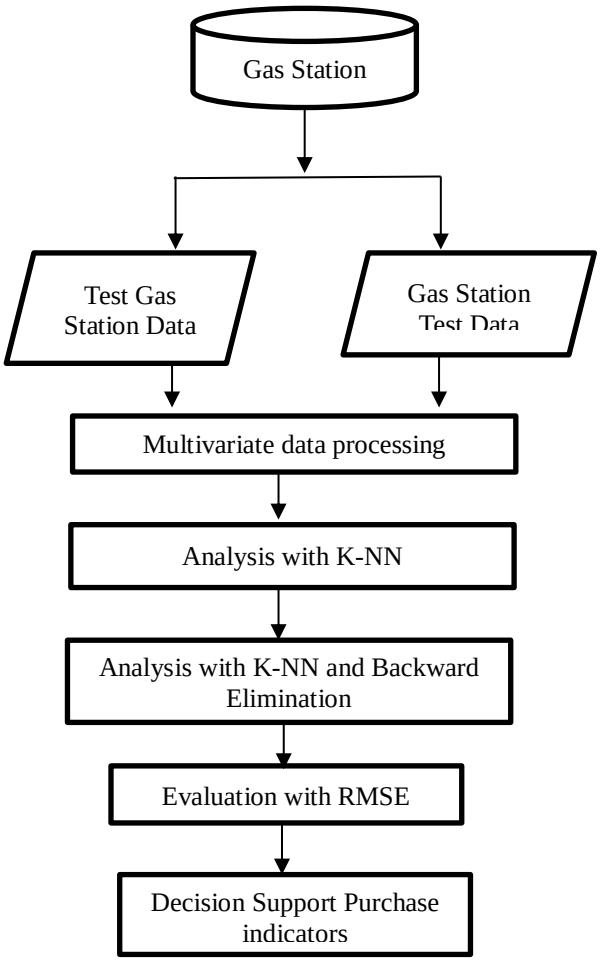

Figure 2. The research framework

This step aims to validate our proposed method. The testing model used in this research is a prediction model which is validated using root_mean_squared_error (RMSE). The results of calculating gas station redemption data using the normal K-NN method will then be compared with the data resulting from K-NN calculations with attribute selection, so which one produces the best K.

RESULT

The data used is time series data on purchasing and sales reports at gas stations owned by PT. Syirkah Amanah Mandiri is in Balen District, Bojonegoro Regency. From the time series data, the main indicators that will be used as a benchmark are purchase volume, type of purchase, month of purchase, profit obtained. In this research, the fuel data used is Diesel. This type of fuel is most often used. As shown in table 4.1 of the 2020 diesel ransom recapitulation data below:

Table 2. Solar Ransom Recapitulation Data for the 2020-2024 period

DATE	PURCHASE AMOUNT (Liter)	UNIT PRICE (Rp)
16-31 December	242.000	4.925
1-15 December	230.000	4.925
16-30 November	237.000	4.925
1-15 November	235.000	4.925
16-31 October	234.000	4.925
1-15 October	230.000	4.925
16-30 September	246.000	4.925
1-15 September	242.000	4.925
16-31 August	237.000	4.925
1-15 August	235.000	4.925
16-31 January	261.000	6.428

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

1-15 January	260.000	6.428
16-31 December	215.000	6.428
1-15 December	217.000	6.428
16-31 August	91.000	6.628
1-15 August	85.000	6.628
16-30 Juny	217.000	6.628
1-15 Juny	215.000	6.628
16-31 May	232.000	6.628
1-15 May	240.000	6.628
16-30 April	215.000	6.628
1-15 April	225.000	6.628
16-31 August	262.000	5.263
1-15 August	258.000	5.263

After getting normalized data, namely date or period data and purchase volume. In predicting the volume of diesel fuel purchases, it is based on the volume of the previous period, where the previous volumes are variables that influence the volume of the next period. The following is the process of univariate data becoming multi-variate:

Table 3. Results of changing univariate to multivariate with 3 predictions

xt-3	xt-2	xt-1	Xt
235000	237000	230000	242000
234000	235000	237000	230000
230000	234000	235000	237000
246000	230000	234000	235000
242000	246000	230000	234000
237000	242000	246000	230000
235000	237000	242000	246000
228000	235000	237000	242000
220000	228000	235000	237000
194000	220000	228000	235000
190000	194000	220000	228000
215000	190000	194000	220000
225000	215000	190000	194000
254000	225000	215000	190000
250000	254000	225000	215000
195000	189000	250000	254000
208000	195000	189000	250000
184000	208000	195000	189000
261000	184000	208000	195000
260000	261000	184000	208000
215000	260000	261000	184000
217000	215000	260000	261000
237000	217000	215000	260000
235000	237000	217000	215000
57000	235000	237000	217000
55000	57000	235000	237000

The next step is to carry out calculations using the K-Nearest Neighbor algorithm method on this data, so that the best method for predicting the volume of fuel purchases will be produced. The best method is calculated based on the calculation of the smallest root mean squared error (RMSE) of K used in the calculation. In the following analysis, the K that will be used is random but does not exceed the number of instances of the data that will be used. The K chosen determines the results of the prediction that will be obtained, and the K with the smallest RMSE is the best K for producing predictions.

In this research, the k values used are k=1, K=4, k=5, k=7 and k=13. Before the calculation with k=1 is carried out. Then the initial cluster is determined from the training data based on interval classes. Meanwhile, clusters in testing data are determined based on testing with K=4, k=5 and k=7. The next process is calculating the percentage value of cluster members obtained from the data testing process for the values K=4, k=5, k=7 and k=13. Data analysis in this research uses tools to determine data mining, namely Rapid Miner. In the rapid miner the data will be processed by selecting the data to be tested and testing it using the K-NN algorithm, with the following scheme.

The data used is the 2024 BBM Ransom Recapitulation Data, namely Diesel, as shown in table 4.3 above. Next,

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

to calculate the validation of the results obtained by calculating the K-NN algorithm, the root mean squared error (RMSE) calculation technique is used, where the smallest RMSE of K is used in the calculation. The following is a scheme for calculating RMSE.

From this scheme model, the results of the K-NN calculations are then validated with RMSE, producing the RMSE table as follows, from calculations for K = 1, K = 4, K = 5, K = 7 and K = 13.

Table 4. RMSE Value Prediction Results of Diesel Fuel Purchase Volume using testing data January 2020–November 2024.

kNN	RMSE
k=1	61730.493 +/- 10839.690
K=4	48649.070 +/- 13856.782
k=5	45567.429 +/- 13010.009
k=7	47057.164 +/- 12290.996
k=13	46734.280 +/- 13733.102

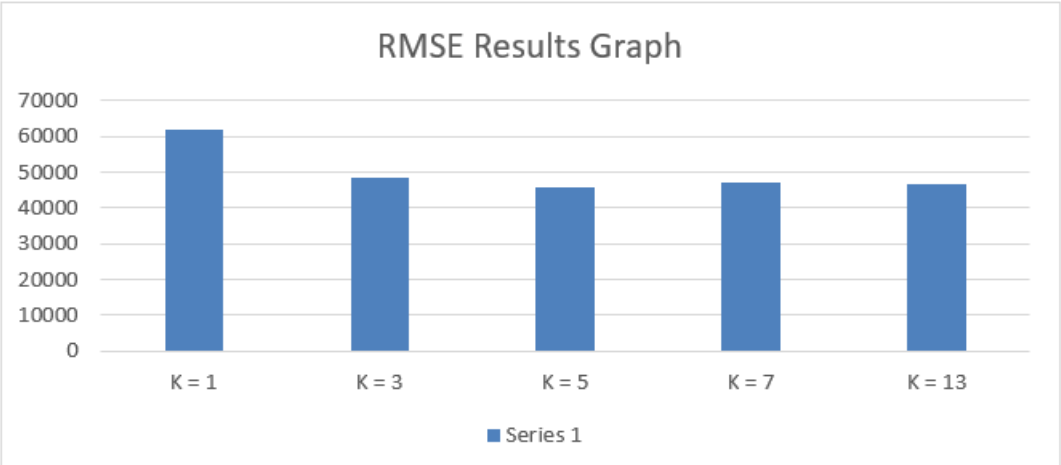

Figure 3 Graph of RMSE results for Solar data with K-NN

From table Table. 4. RMSE Value of Diesel Fuel Purchase Volume Prediction Results using testing data from January 2020 – November 2024 and Figure 4.2 Graph of RMSE results for Solar data with K-NN above, the results obtained are that for analysis of diesel purchase volume data using the K-NN algorithm, K is The best is K = 5, that is, with the smallest RMSE of 45567.429, so K = 5 will be used in the prediction process

After calculating the RMSE between multivariate data with 2 comparisons from the previous period and multivariate data with 3 comparisons from the previous period, the result is that K=5, the smallest K is 45567.429. The following is a comparison table of RMSE results from 2 and 3 comparisons.

Table 5. RMSE results from 2 and 3 periods

	RMSE			
kNN	Multivariat Periods 2	Multivariat Periods 3	Multivariat Periods 4	Multivariat Periods 5
k=1	59666.211	61730.493	55282.946	59683.276
K=4	50796.881	48649.070	48712.843	52612.915
k=5	48721.688	45567.429	46985.134	49408.358
k=7	46330.321	47057.164	46954.908	47724.549
k=13	45945,340	46734.280	46887.946	47246.726

After comparing the RMSE results between 2 periods, 3 periods, 4 periods and 5 periods, it was found that the smallest RMSE was in the 3rd period, namely K=5 which was worth 45567.429. For the next process, predictions will be carried out using backward elimination, with the aim of selecting which attributes in that period can be used in

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

predicting the volume of purchases of diesel fuel. To evaluate the results of the selected attributes, RMSE is again used to determine the smallest K from the K-NN analysis process and backward elimination. The following are the results of attribute selection from the results of prediction analysis with backward elimination.

Table 6. Prediction analysis results with backward elimination.

attribute	Weight
Xt-3	1
Xt-2	0
Xt-1	1

From the results of the prediction analysis with backward elimination, above, it was found that the attribute weights in period xt – 3 and in period xt 1 were chosen as references in the prediction process, because the weight was After obtaining the results from the analysis using backward selection on K-NN, the next step is to evaluate the results of the analysis, analysis using RMSE. The following are the results of analysis with RMSE.

Table 7. Evaluation Results of RMSE KNN Analysis with backward elimination

kNN	*RMSE*
k=1	61730.493 +/- 10839.690
K=4	48649.070 +/- 13856.782
k=5	45567.429 +/- 13010.009
k=7	47057.164 +/- 12290.996
K=13	45556.788 +/- 13972.028

From the evaluation results of the prediction analysis with backward elimination, above, it is found that the smallest K is K = 13, which has a value of 45556.788. So from the K-NN and backward elimination process and the evaluation results of these calculations using RMSE, we get the xt-3 and xt-1 attributes as the prediction period, and K=13 is the smallest K with a value of 45556.788.

Prediction Results of Diesel Fuel Purchase Volume Using K-NN and Backward Elimination After getting the analysis results for the right period to predict fuel volume purchases, which produces xt3 and xt1 values as a prediction reference, according to the table below.

Table 8. Prediction analysis results with backward elimination

attribute	Weight
Xt-3	1
Xt-1	1

The results above show that week 3 and week 1 in the last period of 2024 can be used as a reference for purchases in the following year, namely 2025. As shown by the following table.

Table 9. Results of analysis of predictions of diesel fuel volume purchases in 2025

MONTH	**DATE**	**AMOUNT OF PURCHASE (Liter)**	**DATE**	**PURCHASE QUANTITY (Liter)**
October	16-31 October 2024	234.000	16-31 October 2025	230.000
	1-15 October 2024	230.000	1-15 October 2025	246.000
September	16-30 September 2024	246.000	16-30 September 2025	242.000
	1-15 September 2024	242.000	1-15 September 2025	237.000
August	16-31 August 2024	237.000	16-31 August 2025	235.000
	1-15 August 2024	235.000	1-15 August 2025	228.000

DISCUSSION

Based on the analysis in this research, there are several suggestions that might be used as a basis for developing this research, namely:

1. Add more varied data from both physical and non-physical examinations.
2. Develop methods and combine other methods because in solving this problem, the KNN method is not the only method that can be used.

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4964

3. The system can also be developed by determining the K value and the optimal composition of the amount of training data and test data to get the smallest error.

CONCLUSION

From the analysis and evaluation process of fuel purchase volume prediction research using the fuel redemption dataset which is processed into multivariate data, the analysis process using the K-NN method using 2 periods and 3 periods produces the smallest K located in the 3rd period, so that For the 3rd period, predictions will be measured using K-NN based on backward elimination. With the aim of finding the best method for predicting the volume of fuel purchases, producing predictions with backward elimination, that the attribute weights in period xt – 3 and in period xt 1 are selected as references in the prediction process, because the weight is 1. K = 13 is K which best to carry out the analysis process with K-NN to predict the volume of fuel purchases, with a value of K = 4 of 45556,788. So in analyzing and predicting fuel oil purchase volume data, for diesel, the best K is K=13 using the K-NN analysis method with the backward elimination process.

The results above show that xt3 or week 3 and week xt1 1 in the last period of 2024 can be used as a reference for purchases in the following year, namely 2025.

REFERENCES

Augie S,Wawan Joko.(2024).Metode Regresi Linier Berganda Untuk Prediksi Pemakaian Bbm Pt.Kalonica Bara Kusuma.Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro dan Informatika.Vol.2, No.1

Amelia, N., & Aprianti, W. (2023). The Comparison of Decision Tree and K-Nearest Neighbor Performance for Determining Mustahik. Brilliance: Research of Artificial Intelligence, 3(2), 81-86. https://doi.org/10.47709/brilliance.v3i2.2953.

Andrian, Steele, Salim, E. S., Bindan, H., Pranoto, E., & Dharma, A. (2020). Analisa Metode Random Forest Tree dan K-Nearest Neighbor dalam Mendeteksi Kanker Serviks. Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI), 3(2), 97–101. https://doi.org/10.9767/jikomsi.v3i2.73

Chrisley Heltroyce1, Gerson Feoh2*, I Made Dwi Ardiada.sentiment analysis on the increase of fuel oil Prices using the k-nearest neighbor algorithm.Jurnal Kesehatan, Sains, dan Teknologi Vol. 3, No.1 April 2024

Faisal, M., Utami, W., & Parmica, S. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan. *Journal Sensi: Strategic of Education in Information System*, *9*(1), 11-23. https://doi.org/https://doi.org/10.33050/sensi.v9i1.2616

Fitria, D., pratama, S, C., Zs, Y, N., Pertiwi, E, D., (2023) The Effect Of Increasing Price And Scarcity Of Fuel (BBM) On Fisherman's Income In New Ketapang Village Semidang Alas Maras District Seluma District. Jurnal Ekonomi, Manajemen, Akuntansi dan Keuangan, Vol. 4 No. 3, Juli 2023 page: 575 – 584| 575

Gasparetto, A., Marcuzzo, M., Zangari, A. & Albarelli, A. (2022). A survey on text classificat ion algorithms: From text to predictions. Information, 13(2), 83.

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T. & Yeganegi, M. R. (2020). Text mining in big data analyt ics. Big Data and Cognit ive Computing, 4(1), 1.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning (p. 68)*. New York: Springer.

L. Xu, C. Jiang, J. Wang, J. Yuan and Y. Ren, "Information Security in Big Data: Privacy and Data Mining," IEEE Access: The Journal for Rapid Open Access Publishing, vol. 1, pp. 1149-1176, 9 October 2014.

MA. Lusiandro, SM Nasution, C. Setianingsih., 2020, Implementation of the Advanced Traffic Management System using k-Nearest Neighbor Algorithm, IEEE Access, Conference Paper. International Conference on Information Technology Systems and Innovation (ICITSI).

M. Arhami and M. Nasir., 2020, Data Mining Algoritma dan Implementasi. Penerbit Andi

Prasetyo, A. A., Suprapedi, S., Narulita, S., & Praharsena, B. (2022). Metode K-Means Berbasis Ordered Weighted Averaging (OWA) pada Data Potensi Desa untuk Penentuan Status Desa. *Jurnal Bingkai Ekonomi (JBE)*, *7*(2), 76-91. https://doi.org/10.54066/jbe.v7i2.224

R. Risdiyanta, “Membedah Stasiun Pengisian Bahan Bakar Umum Di Indonesia,” Swara Patra Maj. Ilm. PPSDM Migas, vol. 4, no. 3, Art. no. 3, Dec. 2022.

Statsoft Inc. (2019). Neirest Neighbor. Retrieved from http://www.statsoft.com/textbook/k-nearest- neighbors. https://www.statistikian.com/2024/11/analisis-multivariat.html

Veronika, J., & Andri, A. (2022). Penerapan Metode Algoritma Neural Network Untuk Memprediksi Penjualan Bahan Bakar Minyak. Journal of Information Technology Ampera, 3(2), 235–243. https://doi.org/10.51519/journalita.volume3.isssue2.year2022.page235-24