Forecasting Blood Demand Using the Support Vector Regression Method (Case Study: Blood Transfusion Unit-PMI Central Lombok)

Yati Apriati[1], Wafiah Murniati[2], Saikin[3], Sofiansyah Fadli[4], Hairul Fahmi[5]
[1,2,4,5]Teknik Informatika, STMIK Lombok, Praya, Indonesia
[3]Sistem Informasi, STMIK Lombok, Praya, Indonesia
Email: [1]apriatiyati750@gmail.com, [2]wafiah.mr@gmail.com, [3]eken.apache@gmail.com, [4]sofiansyah182@gmail.com,
[5]iroel.ami@gmail.com

Abstract *– Blood is an important component produced by the human body. Blood is also a very vital part of human survival. When blood levels in the human body are less than they should be, the way to overcome this is by donating blood or blood transfusion. The health facilities that organize blood donations, provide blood and distribute blood are called Blood Transfusion Units (UTD). UTD in carrying out its duties encountered several obstacles, such as blood only having a shelf life of 35 days from donation. If it has passed the expiration date, it cannot be used anymore for blood transfusions. Meanwhile, regarding the demand for blood, the need for blood is greater than those donating. Making it difficult for UTD if the demand occurs when the existing blood stock is not sufficient. And if the stock in UTD experiences an axcess, it can cause losses because the blood is wasted due to expiration. Apart form that. The problem is that in everyday life, many people's need for blood is reduced. Many of their families intervened directly to find available donors. They even search on social networks or social media such as WhatsApp, Facebook, Instagram and others. And this shows that many of them lack donors. To anticipate these problems. So it is necessary to carry out research on forecasting blood demand using the Support Vektor Regression method at UTD PMI Central Lombok. The aim of this research is to forecast or predict the demand for blood at UTD PMI Central Lombok in the coming period. To reduce the impact of lack or excess blood. SVR is the application of Support Vektor Machine (SVM) in the case of regression to find the best dividing line in the regression function. The advantage of the SVR model is that it can handle overfiting problems in the data. The tests used to measure the best model are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R2). The results of this research shows that the best model is Support Vektor Regression (SVR) with a polynomial kernel and based on the tuning results, the parameters used are C=10, degree=1, epsilon=1. The SVR model using a polynomial kernel produces a MAPE value of 18.7502% and RMSE value of 0.6919, which means the model has very good predictive ability. Prediction accuracy was achieved with an R2 value of 0.9936 or 99.36% and an MSE value of 0.4787, which means that the prediction of blood demand data at UTD PMI Central Lombok using SVR with a polynomial kernel function had very good prediction accuracy. With predicted result in january for blood type A it was 1654, B was 920, O was 2205 and AB was 1104.*

Keywords: blood, blood transfusion units (UTD), Support Vektor Regression (SVR), MAPE, RMSE, MSE, R2

I. INTRODUCTION

Blood is an important component produced by the human body. Blood is also the important part of human survival. In general, the function of blood is as a means of transportation in the body such as O_2, CO_2, nutrition, metabolism, hormones, heat and body immunity [1]. When blood level in the human body are less than they should be, the way to overcome this is by donating blood or blood transfusion [2]. Blood transfusion is the process of donating blood voluntarily by someone, which will later be used to help people in need [3][4].

In indonesian, the need for blood continues to increase every year, this happens because blood is needed for accidents and planned operations. Blood donation is a very important thing but is often forgotten. The indonesian Red Cross an one of the blood providers is currently experiencing limitations due to a lack of blood supply for patients, because according to WHO (World Health Organization), ideally blood availability is 2% of the population, meaning that indonesian currently needs 4.6

million bags of blood but PMI only received less than 0.5% of this need in 2005 [5][6]. In accordance with Government Regulation Number 7 of 2011 concerning Blood Transfusion Services, Minister of Health Regulation Number 83 of 2014 concerning Blood Transfusion Units and Minister of Health Regulation Number 14 of 2021 concerning Related Standards relating to commercial activities and products. In implementing risk-based commercial licensing. In the health sector, the health service facility that organizes blood donations, provides blood and distributes blood is the Blood Transfusion Unit (UTD) [[7][8]. UTD in organizing blood donations, providing and distributing blood, there are several obstacles such as the nature of blood which has a limited shelf life [9]. Or is easily damaged [10]. Quantity requested and amount of blood available at UTD. For example, blood has as a shelf life of 35 days from donation. If it has passed the expiration date, it cannot be used anymore for blood transfusion [11]. Meanwhile, regarding the demand for blood, the need for blood is greater than those donating, making it difficult for UTD if the demand occurs when the

JISA (Jurnal Informatika dan Sains) (e-ISSN: 2614-8404) is published by Program Studi Teknik Informatika, Universitas Trilogi
under Creative Commons Attribution-ShareAlike 4.0 International License.

147

existing blood stock is not sufficient. And if the stock in UTD experiences an excess, it can cause losses because the blood is wasted due to expiration [12]. Apart from that, the problem is that in everyday life, many of their families intervened directly to find willing donors. They even search on social networks or social media such as WhatsApp, Facebook, Instagram and others. And this shows that many of them have not had their blood needs met.

To anticipate losses if there is an excess or shortage of blood, forecasting is necessary. Forecasting is the activity of estimating, predicting future events to provide blood reserves to meet blood needs in the future [13]. The method used is Support Vektor Regression (SVR). The application of Support Vektor Machines (SVM) in regression cases with output in the form of real numbers or sequential data is called SVR [14]. SVR is an algorithm that can overcome the problem of overfitting where a good model is produced from training data and not test data. a data that at the training stage produces almost perfect prediction accuracy values is called Overfitting [15].

Research conducted by [16] with the title Support Vektor Regression for Forecasting Blood Demand: Case Study of UTD Branch-PMI Malang City to Determine Optimal Norms for SVR Parameters in predicting Blood Demand in PMI Malang City. Based on the results of blood data simulations for the 2010-2014 period, a minimum MAPE .value of 3,899% was obtained with parameter values

lambda=10, sigma=0.5, cLR= 0.01, C=0.1, epsilon=0.01, number of data features=4 and the number of iterations is 5000, from 12 test data used. The resulting MAPE value is <10% and can be said to be good for predicting the amount of blood needed. Meanwhile, research was also carried out regarding forecasting blood demand which was carried out by [17] by taking a case study in Malang City. In his research, it was concluded that based on the experimental result obtained, the accuracy value produced by the system in the Fuzzy Time Series Interval Optimization research using Particle Swarm Optimization in predicting blood needs was 92.49670% with an error rate (MAPE) of 7.50330% which was obtained from the error calculation results from actual data with forecasting results on 12 test data. in addition, research conducted by [18] used the SVR method to predict monthly rainfall. The SVR prediction results obtained show that the SVR prediction method is very effective in predicting rainfall with forecast accuracy obtained by testing the RMSE value against the SVR model parameter values. The best RMSE is 0.038800637 with gamma of 0.0005, C of 0.0001, and epsilon of 1. Based on the problems presented and research references, the author proposes research entitled "forecasting blood demand using the Support Vektor Regression method (Case Study: Blood Transfusion Unit – PMI Central Lombok)". Hopefully the results of this research can help blood transfusion units adjust blood stocks.

II. RESEARCH METHODOLOGY

The steps taken in this research is began with a literature riview, collecting data, pre-processing data, modeling, evaluating, implementing, and drawing conclusions. These processes shown in the diagram below in figure 1.
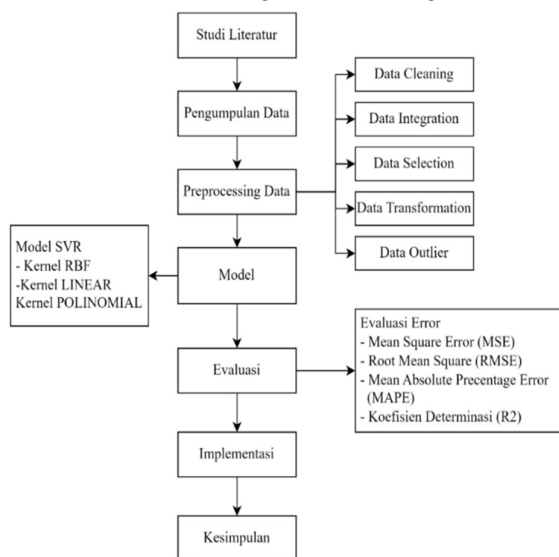

Fig 1. Research Steps

2.1 Literature Review

at this stage, a literature study is carried out or collecting information based on references from books, journals, and some research related to the study, a national and international research. Literature studies also discuss theories that are relevant to the research and

support the research.

2.2 Data Collection

This data collection was carried out by conducted a survey directly at the research site or field, namely UTD PMI Central Lombok. The data used in this research was incoming and outgoing blood data from March 2021 to December 2022 at UTD PMI Central Lombok.

	No	Desa	Nama	Usia	Gol	Tanggal_Masuk	Jumlah_Masuk	Total
0	1	KOPANG	NaN	NaN	A	2021-03-01	3	4
1	2	KOPANG	NaN	NaN	B	2021-03-01	3	4
2	3	KOPANG	NaN	NaN	O	2021-03-01	7	10
3	4	KOPANG	NaN	NaN	AB	2021-03-01	1	2
4	5	TIBU SISOK	NaN	NaN	A	2021-03-03	7	8
...	...	...	...	...	...	...	...	...
2700	2701	KOPANG	NaN	NaN	AB	2022-12-31	3	4
2701	2702	DD GEDUNG	NaN	NaN	A	2022-12-02	47	51
2702	2703	DD GEDUNG	NaN	NaN	B	2022-12-05	44	45
2703	2704	DD GEDUNG	NaN	NaN	O	2022-12-10	129	130
2704	2705	DD GEDUNG	NaN	NaN	AB	2022-12-08	6	7

2705 rows × 8 columns

Fig 2 Incoming Blood Data

2.3 Data Preprocessing

Data preprocessing is the stage for carrying out an initial process in data processing. Data Preprocessing aims to convert raw data into quality data so that the data is suitable for processing at the next stage. The stages in data preprocessing are [19]:

a. Data Integration

Data integration is the process of combining data from various databases to become one new database. The data needed in the data mining process does not only come from one database but can also come from several databases.

```
[33]: df=pd.read_csv('masuk.csv',sep=';')
      df1=pd.read_csv('keluar.csv',sep=';')

      data= pd.merge(df,df1, on='No')
      data.shape

[33]: (2705, 15)
```

Fig 3. *Integration Data Source Code*

b. Data Cleaning

Cleaning is the process of cleaning the data that will be used from unnecessary characters and even words. Data cleaning is carried out to clean unfalid data. unfalid data is data that in a dataframe has an empty value and the data has a Nan value. In the incoming and outgoing data, there are some blanks and some values, such as the name feature, age feature, aptave date feature and pocket number feature, have empty data. As in Figure 4 below, it shows missing data.
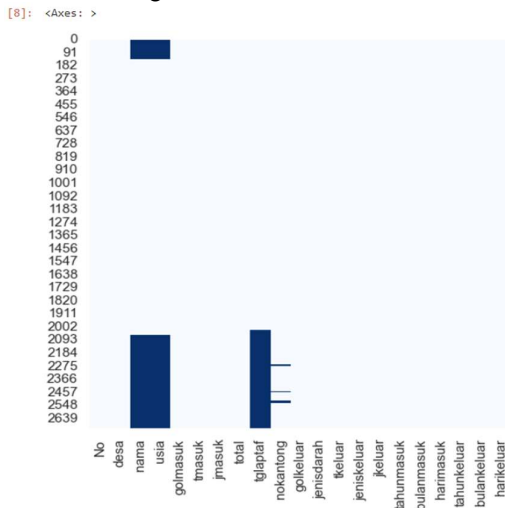

Fig 4. Data Missing

There are 2 steps used to clean a missing data, namely deleting missing features and unimportant features and inputting the average or mean value for features with few missing values. Here the features that are deleted such as village, aptaf date and pocket number because apart from having missing values, these features are also not very important. and the age feature is input based on the average value.

c. Data Selection

Data selection is data from a dataset which is be processed later. This data selection aims to select attributes that are considered to be attributes that influence blood data classification.

```
[31]: data= data[['usia','golmasuk','bulanmasuk','jmasuk','golkeluar','bulankeluar','jenisdarah','jkeluar','total']]
      data.head()
```

[31]:

	usia	golmasuk	bulanmasuk	jmasuk	golkeluar	bulankeluar	jenisdarah	jkeluar	total
0	36.376691	A	3	3	A	3	PRC	1	4
1	36.376691	B	3	3	B	3	PRC	1	4
2	36.376691	O	3	7	O	3	PRC	3	10
3	36.376691	AB	3	1	B	3	PRC	1	2
4	36.376691	A	3	7	O	3	PRC	1	8

Fig 5. Selection Data

d. Data Outlier

Outliers are data that deviate too far from other data in a data series (outliers). The existence of outliers will bias the analysis of a series of data.
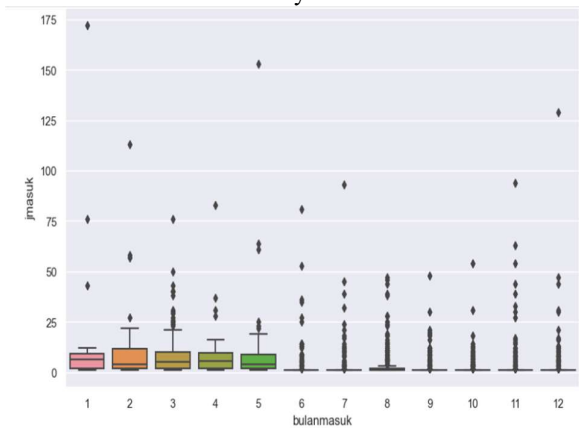

Fig 6. Outlier Data

e. Data Transformation

Transformation is the stage of making the data to be processed according to the model or algorithm that you want to use in the data processing stage. Data that past the data selection and data cleaning will be transformed, aiming to the distribution of the data so that the data becomes normal data and is easier to group. Data transformation here means changing categorical values into numerical values by utilizing the source code provided by Sklearn, such as the label encoder. In Figure 7, the columns that are transformed are the goal-in column, the goal-out column, the blood type column and the age column. The age column is binned first, after which it continues to the transformation stage.

```
from sklearn.preprocessing import LabelEncoder
from sklearn.compose import ColumnTransformer

le= LabelEncoder()
data['golmasuk']= le.fit_transform(data['golmasuk'])
data['golkeluar']= le.fit_transform(data['golkeluar'])
data['jenisdarah']= le.fit_transform(data['jenisdarah'])
data['usia']= le.fit_transform(data['usia'])
data.head()
```

Fig 7. Transformation Data Source Code

2.4 SVR Models

In this SVR model there are three kernels, namely the linear kernel, polynomial kernel and radial basis function (RBF) kernel. These three kernels will be tested using train data and good results will be selected for forecasting. a support vector regression diagram can be seen below:
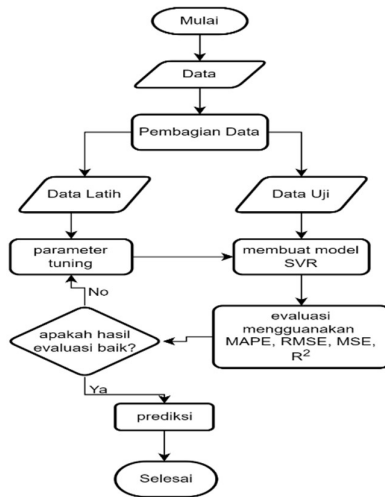

Fig 8. Support Vector Regression Steps

2.5 Evaluation

after getting good results from all three kernels. Then, an evaluation stage was carried out using MAPE, MSE, RMSE and d2. After the evaluation results for the SVR model and the three evaluation methods come out, the next stage is to make predictions. If the combination of kernel functions, parameters, and kernel evaluation used gives good results, then predictions can be made using the test data that was provided at the time of data sharing.

2.6 Implementation

This stage is carried out based on the results of tests that have been carried out previously. This implementation includes algorithm implementation and interface implementation using the Python programming language with the Jupyter notebook editor.

2.7 Conclusion

At this stage, conclusions are drawn after passing through the initial stages of literature study to system testing and analysis which the result in conclusions based on the results of the research carried out by the author. Drawing conclusions is used to answer the problem formulation that has been formulated before. And provide suggestions for further research.

III. RESULTS AND DISCUSSION

3.1 SVR Models

The data used in this research was 2075. The data was divided into 2, namely 80% training data or 2164 data, while 20% was test data amounting to 541 data. Then the training data is used to find the value of each kernel, including:

a. Linear Kernels

The linear kernel is to look for the x_train, y_train scores and x_test, y_test scores. The scores obtained

are for the x_train, y_train scores of 0.3703 and for the x_test, y_test scores of 0.4006. as in figure 9.

```
svr_linear=SVR(kernel='linear')
svr_linear.fit(x_train, y_train)

svr_linear.score(x_train, y_train),svr_linear.score(x_test,y_test)

(0.37036642225181704, 0.40062565856545373)
```

Fig 9. Source Code Linear Kernels

b. Rbf Kernels

Then, the formation of SVR uses Radial Basis Function (RBF). This is a way to look for x_train, y_train scores and x_test, y_test scores. The scores obtained are the scores for x_train, y_train of 0.2662 and the scores for x_test, y_test of 0.3210. as in figure 10.

```
svr_rbf=SVR(kernel='rbf')
svr_rbf.fit(x_train, y_train)
svr_rbf.score(x_train, y_train),svr_rbf.score(x_test,y_test)

(0.2662399274200602, 0.32104012447481844)
```

Fig 10. Source Code RBF Kernels

c. Polynomial Kernels

The next thing is to look for the x_train, y_train scores and x_test, y_test scores. The scores are the scores for x_train, y_train of 0.8131 and for x_test, y_test of 0.6856. As seen in Figure 11 below:

```
svr_poly=SVR(kernel='poly' )
svr_poly.fit(x_train, y_train)
svr_poly.score(x_train, y_train),svr_poly.score(x_test,y_test)

(0.8131136206308167, 0.6856683787743065)
```

Fig 11. Source Code Polynomial Kernels

To make it clearer, you can see in table:

Table 1 Comparison of SVR Models

SVR Models	X_train, y_train	X_test, y_test	Information
Kernel linear	0.3703	0.4006	Based on the score obtained, the linear kernel results are not good.
Kernel RBF	0.2662	0.3210	For the RBF kernel the results obtained are not good.
Kernel Poly	0.8131	0.6856	Meanwhile for poly kernels the results are much better than linear kernels and RBF kernels.

Based on table 1, it can be concluded that, based on the three kernels. The kernel used to make predictions is a polynomial kernel. After getting the best SVR model, proceed with improving the model using tuning parameters. Tuning parameters are used to increase the

score of x_train, y_train and x_test, y_test and find the best parameters. Based on the tuning results, it cn be seen that C=10, degree= 1, epsilon= 1 and cross validation(cv)=4 with a score of x_train, y_train of 0.9959 and for the score of x_test, y_test of 0.9946 with a best_score of 0.9944. So it can be concluded that the score has increased very well.

3.2 error evaluation

In this research, the calculation of prediction accuracy uses R2, while the prediction error calculation uses MSE, RMSE and MAPE.

Table 2 Error Evaluation

SVR Models	C	MSE	RMSE	MAPE	R2
Kernel polynomial	10	0.4064	0.6375	22.1498%	99.4620%

Based on table 2, it can ne explains that in this study the error calculation was that the MSE value was 0.4064, the RMSE value was 0.6375 and the MAPE value was 22.1498%. and the calculation of accuracy or precision of predictions using R2 is 99.46%. It means that the SVR model with a polynomial kernel is more accurate in predicting blood demand at UTD PMI Central Lombok. After calculating the error evaluation, the next step is to predict blood demand for the next period. This is the prediction of results to be obtaine for the next month:

```
model.predict(x_pred)

array([1654.51470459, 920.3394106 , 2205.14617507, 1103.8832341 ])
```

Fig 12. Prediction Results

Based on the pictures above about the prediction results for the next month. Or more details can be see below:

	usia	golmasuk	bulanmasuk	jmasuk	golkeluar	bulankeluar	jkeluar	jenisdarah	total
Januari	Remaja	A	Januari	9	A	Januari	9	PRC	1655.0
Januari	Dewasa	B	Januari	5	B	Januari	5	PRC	920.0
Januari	Remaja	O	Januari	12	O	Januari	12	PRC	2205.0
Januari	Dewasa	AB	Januari	6	AB	Januari	6	PRC	1104.0

Fig 12. DataFrame Prediction Results

Based on the pictures above, it can be explained that the prediction results are in January. For incoming blood group A of 9 and outgoing blood group A of 9, the prediction result obtained is 1655. Next, for incoming blood group B of 5 and outgoing blood group of 5, the prediction result obtained is 920. Furthermore, for the incoming blood group O of 12 and the outgoing blood group of 12, the prediction result is 2205. Meanwhile, for incoming blood group AB of 6 and outgoing blood group of AB of 6, the predicted result is 1104.

IV. CONCLUSION

Based on the discussion above, it can be concluded that the SVR model used is a poly kernel because the results shows was very good. With the parameters used, namely C= 10, degree = 1, epsilon=1, this means that the poly kernel is able to overcome the overfitting problem in predicting blood demand at UTD PMI Central Lombok. forecasting blood demand at UTD PMI Central Lombok using support vector regression with a polynomial kernel. then the MSE obtained is 0.4787, the RMSE is 0.6919 and the MAPE value is 18.7502%, which means that the forecasting accuracy is very good or the model has very good forecasting capabilities. The R2 value is 0.9936 or 99.36%, which means that the prediction of blood demand data at UTD PMI Central Lombok using SVR with a polynomial kernel function has very good accuracy or prediction accuracy results. With predicted results in January for blood type A of 1654, B of 920, O of 2205 and AB of 1104.

REFERENCES

[1] Y. Heryanita *et al.*, "4. The Value of Erythrocytes, Hemoglobin, and Hematocrit of Mice (Mus musculus) Exposed to Cigarette Smoke and Given Red Watermelon Extract (Citrullus vulgaris)," *J. Med. Vet.*, vol. 12, no. 1, pp. 24–31, 2018, doi: 10.21157/j.med.vet..v12i1.4106.

[2] G. C. Heriyanto, U. K. Petra, J. Siwalankerto, S. Jl, S. Surabaya, and J. Siwalankerto, "Prediksi Kebutuhan Darah Menggunakan Metode ARIMA Dengan Mempertimbangkan Faktor Deterioration," *J. Infra*, 2022, [Online]. Available: https://publication.petra.ac.id/index.php/teknik-informatika/article/view/12734%0Ahttps://publication.petra.ac.id/index.php/teknik-informatika/article/viewFile/12734/11036.

[3] H. H. Rohan, Y. Amalia, P. Ayu, and D. Reswari, "Journal of Community Engagement in Health Kegiatan Donor Darah Di Fakultas Ilmu Kesehatan Universitas Dr. Soetomo Surabaya Tahun 2018," vol. 4, no. 2, pp. 475–480, 2021, [Online]. Available: http://jceh.orghttps//doi.org/10.30994/jceh.v4i2.272.

[4] A. C. Cagliano, S. Grimaldi, C. Rafele, and C. Campanale, "An enhanced framework for blood supply chain risk management," *Sustain. Futur.*, vol. 4, no. July, p. 100091, 2022, doi: 10.1016/j.sftr.2022.100091.

[5] B. Apriliasari, N. S. Utami, L. Lidia, and A. Abdullah, "AKSI UNIASIAH (Blood Transfusion Unit) Based on Android With Waterfall," *J. Transform.*, vol. 17, no. 1, p. 57, 2019, doi: 10.26623/transformatika.v17i1.1410.

[6] Kemenkes RI, *PROFIL KESEHATAN INDONESIA 2018*. 2018.

[7] K. W. D. Nugraha, *Profil Kesehatan Indonesia 2021*. 2022.

[8] A. Datumaya, W. Sumari, A. K. Febrianto, and Y. Pramitarini, "Sistem Prediksi Permintaan Darah

Menggunakan Metode Regresi Linier (Studi Kasus Pada UTD PMI Kabupaten Bojonegoro)," *J. Inform. Polinema*, pp. 85–90, 2021, doi: 10.33795/jip.v7i2.495.

[9] A. Budipriyanto and F. Anggraini, "Manajemen Persediaan Darah Untuk Meminimalkan Shortage dan Wastage pada Bank Darah PMI Jakarta," *Pros. Senantias*, vol. 1, no. 1, pp. 1235–1244, 2020.

[10] O. S. S. Filho, W. Cezarino, and G. R. Salviano, *A Decision-making Tool for Demand Forecasting of Blood Components*, vol. 45, no. 6. IFAC, 2023.

[11] C. Rutherford, S. Y. Cheng, and K. Bailie, "Evidence of bullwhip in the blood supply chain," *23rd EurOMA Conf. Proc.*, no. June, pp. 1–10, 2016.

[12] M. Meneses, D. Santos, and A. Barbosa-póvoa, "Modelling the Blood Supply Chain," *Eur. J. Oper. Res.*, vol. 307, no. 2, pp. 499–518, 2023, doi: 10.1016/j.ejor.2022.06.005.

[13] N. Prastyanda, P. Pratama, T. Sukmono, B. I. Putra, and W. Sulistiyowati, "Forecasting the Amount of Blood Storage Using the Support Vector Machine (Svm) Method Peramalan Jumlah Permintaan Darah Menggunakan Metode Support Vector Machine (Svm)," vol. 3, no. 07, 2022.

[14] A. D. Dubey, "Gold price prediction using support vector regression and ANFIS models," *Int. Conf. Comput. Commun. Informatics, ICCCI*, no. January 2018, 2018, doi: 10.1109/ICCCI.2016.7479929.

[15] R. Melati, T. W. Purboyo, and M. Kallista, "Prediksi Penderita Tuberkulosis Menggunakan Algoritma Support Vector Regression (SVR) Prediction Of Tuberculosis Patients Using The Support Vector Regression (SVR)," vol. 10, no. 1, pp. 736–741, 2023.

[16] M. R. Rifqi, B. D. Setiawan, and F. A. Bacthiar, "Support Vector Regression Untuk Peramalan Permintaan Darah : Studi Kasus Unit Transfusi Darah Cabang – PMI Kota Malang," vol. 2, no. 10, pp. 3332–3342, 2018.

[17] A. D. A. Rifandi, B. D. Setiawan, and Tibyani, "Optimasi Interval Fuzzy Time Series Menggunakan Particle Swarm Optimization pada Peramalan Permintaan Darah : Studi Kasus Unit Transfusi Darah Cabang - PMI Kota Malang," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 7, pp. 2770–2779, 2018, [Online]. Available: http://j-ptiik.ub.ac.id.

[18] N. A. Siregar, "Peramalan Curah Hujan Di Kota Medan menggunakan Metode Support Vector Regression," vol. 1, no. 1, pp. 7–9, 2022.

[19] E. Virantika, K. Kusnawi, and J. Ipmawati, "Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1657, 2022, doi: 10.30865/mib.v6i3.4325.