International Journal of Economics, Management and Accounting
Volume. 1 No. 2 June 2024
e-ISSN: 3046-9376, Page 221-231
DOI: https://doi.org/10.61132/ijema.v1i2.68

Support Vector Machine: Classification of Trade Balance of Provincies in Indonesia Based on Gross Regional Domestic Product and Large Trade Price Index in 2023

Rati Assyifa Putri
Statistics, Universitas Halu Oleo

Bahriddin Abapihi
Statistics, Universitas Halu Oleo

Dian Christien Arisona
Statistics, Universitas Halu Oleo

Kampus Hijau Bumi Thidarma, Anduonohu, Kambu, Kendari Sultra, 93232
Corresponding author: ratiassyifaputri@gmail.com

Abstract: *The aim of this research is to classify Indonesia's trade balance data using the SVM (Support Vector Machine) method with two features, namely Gross Regional Domestic Product (X1) and Wholesale Price Index (X2). Classification is carried out by comparing two types of kernels, namely polynomial kernels and RBF (Radial Basis Function) kernels. Equality Hyperplaneobtained from the polynomial kernel is:* $-14{,}1785X_1 - 0{,}0202X_2 - 0{,}0151$. *The Hyperplane equation obtained from the RBF kernel is:* $-14{,}1785X_1 - 0{,}0202X_2 + 0{,}06$ *Experimental results show that classification with polynomial kernels provides better performance than RBF kernels. This can be seen in the evaluation results which show that the Polynomial kernel has an average model goodness of 75.93% and for the RBF kernel the average model goodness is 74.07%. Leave One Out cross validation (LOOCV) simulation for training data obtained an average accuracy of 76.67% for the polynomial kernel and 66.67% for the RBF kernel. This shows that in this classification context, kernel polynomials are more effective in separating data classes.*

Keywords: *SVM, Polynomial Kernel, RBF Kernel, Hyperlane, Leave One Out Cross Validation*

INTRODUCTION

The trade balance is important to research because it allows analysis of the flow of goods and services between countries and their impact on the global economy. This contributes to our understanding of the international economic principles underlying the exchange of goods and services between countries. The trade balance is the difference between the export value and the import value of a country in a certain time period. When the value of exports is greater than the value of imports it is called a trade surplus, whereas when the value of imports is greater than the value of exports it is called a trade deficit (Permana et al, 2023).

One of the economic problems faced by the Indonesian government was the occurrence of an international trade deficit from 2012 to 2014 and again in 2018. In 2019 Indonesia's trade balance again experienced a deficit. From 2020 to 2022, Indonesia's trade balance continues to increase and in 2023, Indonesia's trade balance experiences a decrease in value but there is no deficit (Ministry of Trade, 2024). Indonesia's trade balance deficit has an impact on the decline in the rupiah exchange rate because a lot of money circulates quickly, causing the price of

Received April 30, 2024; Accepted May 30, 2024; Published June 30 2024
* Rati Assyifa Putri, ratiassyifaputri@gmail.com

goods to rise (Silitonga et al., 2017). Cumulatively, Indonesia's trade balance has a deficit of US$ 8,698.6 million (Bank Indonesia,2018).

Research conducted by Liani (2021) shows that GRDP has a positive and significant effect on Indonesia's trade balance. The trade balance measures the difference between a country's exports and imports, while the Wholesale Price Index (IHPB) records changes in export and import prices. The two are interrelated because a healthy trade balance can influence currency exchange rates, inflation and economic growth, which are reflected in changes in the IHPB (Krugman et al, 2014).

Gross Regional Domestic Product (GRDP) makes a major contribution to industrial competitiveness at the domestic level. GRDP has a positive and significant influence on economic growth (Romhadhoni, 2019). Along with that, imports are also a critical factor that can affect trade balance, foreign exchange reserves and economic stability. Imports have a positive and significant effect on economic growth and Indonesia's trade balance (Asiyan, 2013).

The development of Indonesia's IHPB (Wholesale Price Index) during 2022 experienced an average increase of 4.74 percent compared to the previous year from 106.20 in 2021 to 111.23 in 2022. The highest change in the index occurred in the Mining and Quarrying Sector. with an increase of 9.17 percent, followed by the Industrial Sector which increased by 5.49 and the Agricultural Sector by 1.34 percent compared to the previous year (BPS, 2023).

Classification analysis is important because it helps understand data patterns, predict labels for new data, group information, optimize processes, detect anomalies, evaluate model performance, and support better decision making. One type of classification is Support vector Machine (SVM) (Hastie, 2009).

Support vector machine (SVM) is a technique for making predictions, both in the case of classification and regression. SVM has the basic principle of a linear classifier, namely classification cases that can be linearly separated, but SVM has been developed so that it can work on non-linear problems by including the kernel concept to deal with non-linear data to find a hyperplane solution that can maximize the distance (margin) between classes data (Santosa, 2007).

The advantage of SVM compared to other classifications, such as classification using hierarchical and non-hierarchical methods, is that it can handle non-linear data and is resistant to outliers, whereas classification using hierarchical and non-hierarchical methods depends on the distribution of the data, if there are outliers then they need to be handled. first before going into the analysis method and experiencing difficulties in handling non-linear data. SVM classification also does not depend on the size of the dataset (Hastie, 2009).

The trade balance, GRDP and IHPB are important things to study. Based on the explanation above, researchers are interested in classifying the trade balance based on GRDP and IHPB using the SVM classification.

THEORETICAL FRAMEWORK

SVM or Support vector Machine is a method in machine learning which aims to find the optimal boundary or distance that can separate two classes in a dataset, which is often referred to as a Hyperplane (Satriyo, 2003). Based on its characteristics, the SVM method is divided into two, namely Linear SVM and Non-Linear SVM. Linear SVM is data that is separated linearly, namely separating the two classes on the Hyperplane with a soft margin. Meanwhile, non-linear SVM applies the kernel trick function to high-dimensional space (Rachman, 2012).

LOOCV is a cross-validation method commonly used in machine learning and statistics to evaluate the performance of prediction models. This method divides the data into two parts: one part is used to train the model, while the other part is used to test the model's performance. Leave One Out cross validation (LOOCV) is an extensive method. In this method, each sample in the dataset is used as testing data in turn, while the other samples are used as training data. If the dataset has N samples, the model is trained and evaluated N times. This method is suitable for datasets with a limited number of samples, but requires higher computing time (Marchetti, 2021).

The trade balance is a record of the value of goods exported or imported by a country. The existence of a trade balance is intended to determine the development of world trade carried out by a country. Export activities will provide rights/benefits in the form of receiving payments or receivables, while import activities will require the country to pay other countries so that there is a high potential for increasing state debt. The trade balance is the difference between the export value and the import value of a country in a certain time period. When the export value is greater than the import value, it is called a trade surplus. When the value of imports is greater than the value of exports, it is called a trade deficit. The trade balance rate is the percentage increase in the difference between export value and import value from one period to the next (Fudllayati, 2021).

Gross Regional Domestic Product (GRDP) is a macroeconomic indicator that can provide an overview of the economic condition of a region (Prishadoyo, 2008).

The Wholesale Price Index (IHPB) is a number that describes the magnitude of price changes at the wholesale level of commodities traded in a region (country/province). The 2022 IHPB is calculated based on the 2018 base year whose commodity packages are obtained from the 2017 Provincial IHPB Weighing Diagram Preparation Survey (SPDT). The number of commodities covered in the 2022 IHPB is 687 commodities (BPS, 2023).

RESEARCH METHODOLOGY

The research used is quantitative research withThe data used in this research is secondary data, namely trade balance (Y), GRDP (X1) and IHPB (X2) obtained from the Ministry of Trade and BPS in 2023.

RESULTS AND DISCUSSION

Descriptive statistics

The characteristics of GRDP and IHPB can be known using descriptive statistics. The descriptive statistics used can be in the form of minimum, maximum, mean, variance and standard deviation values for each research variable. General information about the data used in this research can be seen in table 1 as follows:

Table 1 Descriptive analysis of the dependent variable (Y) and independent variable (X)

variable	Minimum	Maximum	Average	Standard Deviation
X1	35.05	60767.01	8207.56	19009.67
X2	107.2	116.2	112.8	2,164

Source: Processed by the Author

GRDP (X1)

Based on table1 it can be seen that in 2023 the largest GRDP, namely60767.01 million US$ and the lowest GDP was 35.05 million US$with an average GDP of8207.56Million US$ and a standard deviation of 19009.67.

IHPB (X2)

Based on table 1, it can be seen that in 2023 the largest IHPB is116.2% and the lowest IHPB is 107.2%with an average IHPB of112.8% and standard deviation of2,164.

Finding Hyperpline Equations

Dividing Data into Training Data and Testing Data

Data is divided into two, namely training data and testing data. 90% training data and 10% testing data, where the data is divided randomly, the data obtained is presented in table 2

e-ISSN: 3046-9376, Page 221-231

Table 2 Training Data

No.	Y	X1	X2
1	0	160.04	113.54
2	0	160.36	110.64
3	0	67.54	116.22
4	1	560.71	109.93
6	1	2224.12	110.34
7	0	35.05	116.14
8	0	188.43	112.41
9	0	1850.48	107.24
10	1	1222.35	112.73
11	1	2049.1	113.13
13	0	166.36	115.46
14	0	124.15	114.17
15	0	593.19	116.21
16	0	159.34	112.07
18	0	306.89	110.67
19	1	38.29	112.99
20	0	54507.76	114.53
21	0	112.3	112.73
22	1	84.24	112.56
24	0	104.89	111.03
25	0	604.38	113.74
26	0	40302.2	114.8
27	0	419.71	111.11
28	0	60767.01	115.05
29	0	116.93	112.47
30	1	112.39	112.45
31	0	211.16	114.23
32	0	402.61	115.28
33	0	2549.02	110.44
34	0	130.99	110.56

Source: Processed by the Author

Furthermore,To determine the goodness of the model obtained from training data, testing data will be used to see classification accuracy. Testing data is displayed in Table 3.

Table 3 Testing data

No.	Y	X1	X2
5	0	59029.22	113.62
12	0	301.59	113.96
17	1	194.53	112.76
23	0	49199.87	108.88

Source: Processed by the Author

*Hyperplane*on the Polynomial Kernel as follows:

$$f(x) = \sum_{i=1}^{m} a_i y_i K(x_i, x) + b$$

$$= \sum_{i=1}^{m} a_i y_i K(x_i, x) + 0{,}06$$

By using R-Studio, the following equation results are obtained:

$$f(x) = -14{,}1785X_1 - 0{,}0202X_2 + 0{,}06$$

Leave One Out cross validation simulation

LOOCV Polynomial Kernel

The principle of this method is to iterate n data and remove one piece of data sequentially at each iteration. Implementing the LOOCV method can be done with the help of R-Studio software

Table 4 Polynomial LOOCV Simulation

No	*Iteration*	*Accuracy*
1	1	100%
2	2	100%
3	3	100%
4	4	0 %
5	6	0 %
6	7	100%
7	8	100%
8	9	100%
9	10	0 %
10	11	0 %
11	13	100%
12	14	100%
13	15	100%
14	16	100%
15	18	100%
16	19	0 %
17	20	100%
18	21	100%
19	22	0 %
20	24	100%
21	25	100%
22	26	100%
23	27	100%
24	28	100%
25	29	100%
26	30	0 %
27	31	100%
28	32	100%
29	33	100%
30	34	100%
average		76.67 %

Source: Processed by the Author

LOOCV Kernel RBF

The principle of this method is to iterate n data and remove one piece of data sequentially at each iteration. Implementing the LOOCV method can be done with the R-Studio software.

Table 5 LOOCV RBF iteration

No.	*Iteration*	*Accuracy*
1	1	0 %
2	2	100%
3	3	100%
4	4	0 %
5	6	0 %
6	7	100%
7	8	100%
8	9	100%
9	10	0 %
10	11	0 %
11	13	100%
12	14	100%
13	15	100%
14	16	100%

15	18	0 %
16	19	0 %
17	20	100%
18	21	0 %
19	22	0 %
20	24	100%
21	25	100%
22	26	100%
23	27	100%
24	28	100%
25	29	100%
26	30	0 %
27	31	100%
28	32	100%
29	33	100%
30	34	100%
average		66.67 %

Source: Processed by the Author

Determining the Classification Type

SVM classification has two types of classification, namely linear SVM classification and non-linear SVM classification. If the classes in the observation data can be separated by a linear line then in the classification process linear classification can be used and if the classes in the observation data cannot be separated by a linear line then in the classification a non-linear classification is used. Data images based on class are presented in Figure 1

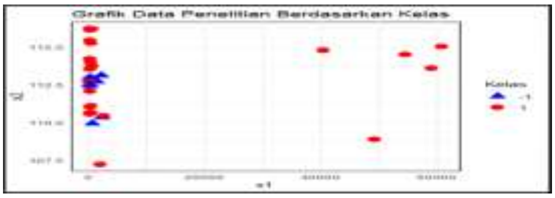

Figure 1 Research Data Graph by Class

Based on Figure 1, it can be seen that the red color represents the pattern with label 0, while the blue color represents the pattern in class 1. Based on Figure 1, it can be seen that the classes in the observation data cannot be separated by a linear line, so in the classification the non-classification will be used. linear by using the kernel method to create Hyperplane lines.

Predictions

Prediction on Polynomial Kernels

Predictions in SVM classification use training data to create a model and use testing data to see the goodness of the model. The data is divided into 90% training data and 10% testing data. Predictions on polynomial kernels can be done with the following hyperlane equation:

$$-14{,}1785X_1 - 0{,}0202X_2 - 0{,}0151$$

With the R-Studio software, prediction results are obtained as follows:

	Actual	
predicted	0	1
0	2	1
1	1	0

Predictions on the RBF Kernel

Predictions in SVM classification use training data to create a model and use testing data to see the goodness of the model. The data is divided into 90% training data and 10% testing data. Predictions on RBF kernel can be done with the following hyperlane equation:

$$-14{,}1785X_1 - 0{,}0202X_2 + 0{,}06$$

With the help of R-Studio software in Appendix 10, prediction results are obtained
as follows:

	Actual	
predicted	0	1
0	2	1
1	1	0

Evaluation of Results

Evaluation of Predictions Using Polynomial Kernels

Based on the prediction results from this classification, the next step is to determine the accuracy, and the SVM model is as follows:sensitifitasspecificity

$$\text{Akurasi } = \frac{TP + TN}{TP + TN + FP + FN} 100\%$$

$$= \frac{3}{4} 100\% = 75\%$$

The SVM model with a polynomial kernel succeeded in predicting Indonesia's trade balance class with an accuracy of 75%. This indicates that most of the testing data was successfully classified correctly. Of the 4 data samples tested, 3 of them were classified correctly according to the actual trade balance. The high level of accuracy shows that the SVM model is quite good at capturing patterns and relationships between GRDP and IHPB variables to predict the trade balance. The accuracy of the SVM method is presented in Table 6.

Table 6 Polynomial SVM Accuracy

Iteration	*Accuracy*
1	75 %
2	100%
3	50 %
4	75 %
5	75 %
6	100%
7	75 %
8	66.67 %
9	66.67 %
average	75.93 %

Source: Processed by the Author

Determining accuracy in SVM modeling was carried out 9 times to see the average accuracy. Table 6 above shows that the accuracy of the SVM classification of the model that has been created is 75.93%.

Prediction Evaluation Using the RBF Kernel

Based on the prediction results from this classification, the next step is to determine the accuracy of the SVM model as follows:

$$\text{Akurasi } = \frac{TP + TN}{TP + TN + FP + FN} 100\%$$

$$= \frac{2}{4} 100\% = 50\ \%$$

The SVM model with the RBF kernel succeeded in predicting Indonesia's trade balance class with an accuracy of 50%. This indicates that most of the testing data was successfully classified correctly. Of the 4 data samples tested, 2 of them were classified correctly according to the actual trade balance. The high level of accuracy shows that the SVM model is quite good at capturing patterns and relationships between GRDP and IHPB variables to predict the trade balance. The accuracy of the SVM method is presented in Table 7.

Table 7 *Accuracy*SVM RBF

Iteration	***Accuracy***
1	50 %
2	100%
3	75 %
4	50 %
5	100%
6	50 %
7	75 %
8	66.67 %
9	100%
average	74.07 %

Source: Processed by the Author

Determining accuracy in SVM modeling was carried out 9 times to see the average accuracy. Table 7 above shows that the accuracy of the SVM classification of the model that has been created is 74.07%.

CONCLUSION

In this research, classification of Indonesia's trade balance data was carried out using the SVM (Support Vector Machine) method with two features, namely Gross Regional Domestic Product (X1) and Wholesale Trade Price Index (X2). Classification is carried out by comparing two types of kernels, namely polynomial kernels and RBF (Radial Basis Function) kernels. Equality *Hyperplane* obtained from the polynomial kernel is:

$$-14{,}1785X_1 - 0{,}0202X_2 - 0{,}0151$$

The Hyperplane equation obtained from the RBF kernel is:

$$-14{,}1785X_1 - 0{,}0202X_2 + 0{,}06$$

Experimental results show that classification with polynomial kernels provides better performance than RBF kernels. This can be seen in the evaluation results which show that the

Polynomial kernel has an average model goodness of 75.93% and for the RBF kernel the average model goodness is 74.07%. The LOOCV simulation for training data obtained an average accuracy of 76.67% for the polynomial kernel and 66.67% for the RBF kernel. This shows that in this classification context, kernel polynomials are more effective in separating data classes.

REFERENCES

Asiyan, S. (2013). Pengaruh penanaman modal dalam negeri, penanaman modal asing, dan ekspor terhadap pertumbuhan ekonomi Jawa Timur. Jurnal Pendidikan Ekonomi (JUPE, 1(3).

Bank Indonesia. (2018). Nilai tukar mata uang asing terhadap rupiah Tahun 1998-2017. Diakses dari https://www.kemndag.go.id/id/ pada 21 Januari 2024, pukul 11.10 WITA.

BPS. (2023). Indeks harga perdagangan besar Indonesia. Diakses dari https://www.bps.go.id/id/ pada 19 Januari 2024, pukul 11.10 WITA.

Farina, F., & Husaini, A. (2017). Pengaruh dampak perkembangan tingkat empor dan impor terhadap nilai tukar negara ASEAN per-dollar Amerika Serikat (Studi pada International Trade Center periode tahun 2013-2015). Jurnal Administrasi Bisnis, 50(6).

Fudllayati, G. (2021). Analisis produk domestik bruto (PDB) dengan regresi nonparametrik kernel menggunakan estimator Priestley-Chao. Skripsi, Universitas Negeri Sunan Ampel, Surabaya.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. Springer.

Jacobus, A. (2013). Penerapan metode support vector machine pada sistem deteksi intrusi secara real-time. Berkala MIPA, 23(2), 4-6.

Kementerian Perdagangan. (2023). Nilai ekspor, impor dan neraca perdagangan tahun 2023. Diakses dari https://www.kemndag.go.id/id/ pada 21 Januari 2024, pukul 11.10 WITA.

Krugman, P. R., Obstfeld, M., & Melitz, M. J. (2014). International economics theory and policy. Pearson Education Limited.

Liani, J. (2021). Analisis peningkatan PDRB Provinsi Jawa Barat melalui kegiatan ekspor-impor. Jurnal Manajemen, Ekonomi, Keuangan dan Akuntansi, 2(1), 105-110.

Marchetti, F. (2021). The extension of Rippa’s algorithm beyond LOOCV. Applied Mathematics Letters, 120, 107262.

Permana, M. A., Sjahruddin, H., Satriani, R., Indrawati, I., & Caronge, M. A. A. (2023). Menelisik neraca perdagangan Indonesia dari faktor inflasi dan kurs. Jurnal Bina Bangsa Ekonomika, 16(2), 710-719.

Prishardoyo, B. (2008). Analisis tingkat pertumbuhan ekonomi dan potensi ekonomi terhadap produk domestik regional bruto (PDRB) Kabupaten Pati tahun 2000-2005. Jejak, 1(1).

Rachman. (2012). Perbandingan klasifikasi tingkat keganasan breast cancer dengan menggunakan regresi logistik ordinal dan support vector machine. s.l., s.n.

Romhadhoni, P., Faizah, D. Z., & Afifah, N. (2019). Pengaruh produk domestik regional bruto (PDRB) daerah terhadap pertumbuhan ekonomi dan tingkat pengangguran terbuka di Provinsi DKI Jakarta. Jurnal Matematika Integratif, 14(2), 113.

Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. Graha Ilmu.

Satriyo, A. (2003). Support vector machine. Teori dan aplikasinya dalam bioinformatika, 1, 1-11.

Silitonga, dkk. (2017). Pengaruh ekspor, impor, dan inflasi terhadap nilai tukar rupiah di Indonesia. Jurnal Ekonomi Pembangunan, 15(1), 53-59.

Sukirno, S. (2006). Pengantar teori makro ekonomi. PT Raja Grafindo Persada.