Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

Comparative Study of Iconnet Jabodetabek and Banten Using Linear Regression and Support Vector Regression

Muhaimin Hasanudin[1)*], Ifan Prihandi[2)], Sifania Nazua[3)]
[1)2)3)] Department of Informatics, Universitas Mercu Buana, Jakarta - Indonesia
[1)]muhaimin.hasanudin@mercubuana.ac.id, [2)]ifan.prihandi@mercubuana.ac.id, [3)]sifania.nazua.kerja@gmail.com,

ABSTRACT

PT PLN Indonesia Comnets Plus has little information regarding future customer growth, making it difficult to take steps to meet customer needs. This research aims to predict customer growth in the future based on quantitative data from the previous year, where the output provided produces data in the form of numbers that are analyzed using statistical methods. The hope is to provide information to maximize customer growth with the minimum area or bandwidth used by the company. This research uses linear regression and support vector regression (SVR) algorithms using a company secondary dataset of 252 data points with 5 attributes. Data was collected during the last one-year period, from January to December 2021. The results of the research show that predictions using both algorithms have increased, customer growth when viewed from the number of customers, bandwidth, and regional data has increased significantly. This can be seen from the value of the number of customers, which continues to increase, while the highest number of customers falls in December, the most requested bandwidth is 20 mbps, and the largest customer area is in the Depok area. The results of the research show that the SVR algorithm is superior in terms of mean absolute percentage error (MAPE): 0.02% MAPE, 0.10 MAE, and 0.99 RMSE, while for linear regression, the MAPE values were 36.28%, MAE 201, and RMSE 0.80.

Keywords: customer data, customer growth, linear regression, Predictions, support vector regression

1. INTRODUCTION

The emergence of telecommunications service providers is growing in view of the many needs of society. PT PLN Indonesia Comnets Plus (PLN ICON+) is one of the companies' providing services that meet the needs of information and communication technologies (Pratama & Noviana., 2023). The company generates significant sales every year. Where the value of each increasing customer makes it a huge success for the company itself (McKim & Hughes., 2001). Therefore, knowing future customer growth will help the company know which criteria will have more fans and which will not. The appropriate prediction methods for the available data are Linear Regression (Sandiwarno., 2018; Maulud, D., & Abdulazeez., 2020) and Support Vector regression (Anand & Chandra., 2020; Cindo & Ermatita., 2020; (Fitrianah dkk., 2022)). Linear regression is a statistical method for modelling the relationship between independent variables and dependent variables. Support Vector Regression (SVR) is a machine learning method used to perform regression that differs from conventional linear regression (Duan & Huang., 2019; Maulana dkk., 2019). SVR is a form of SVM implementation used for regression (Wen dkk., 2018; Dewi & Chen., 2019). The goal is to find a function that maps data into a feature space so that maximum prediction errors can be minimised.

Several related studies, among others, looked at predictions of smartphone sales (Sianturi & Sagala., 2020) using a linear regression method (He., 2023; Princess & Mashuri., 2020; Nastiti & Putrada., 2019). The research provides predictions of smartphone sales for the coming months, smartphone sales data for the previous year, and advertising costs of $2 million used as input. The resulting prediction accuracy is Mean Squared Error (MSE) 43,112% and Mean Absolute Percentage Error (MAPE) 20,001%. Other research is to determine the level of influence of weather on productivity in Indonesia using the double linear regression method and Support Vector Regression (SVR). Test results show that SVR and Linear Regression have the same level of accuracy. From the Linear Regression evaluation results, the rate of default of Root Mean Square Error (RMSE) was 26.4763 and Mean Absolute Error(MAE) was 21.7043, while the SVR showed the RMSE level was 25.8503 and the MAE level was 20.9812.

Based on this, one of the main issues to be explored is what the future growth of subscribers should be, so that it can help decision makers to get better information. Existing data such as the number of subscribers, bandwidth, and regional data have relationships between variables so that they can be accommodated in the model. Comparison of

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

results between the two algorithms used is very important and interesting for the author to conduct research. Comparison of Iconnet Jabodetabek and Banten Customer Growth Forecasting results in that the SVR Algorithm is superior in terms of mean absolute percentage error (MAPE): MAPE 0.02%, MAE 0.10, and RMSE 0.99, while for linear regression, the MAPE value is 36.28%, MAE 201, and RMSE 0.80.

LITERATURE REVIEW

Linear Regression

Linear regression is a statistical method used to evaluate the linear relationship between two or more variables. The main purpose of linear regression is to model the relationship between the independent variable (X) and the dependent variable (Y) so that it can be used to make predictions or estimates of Y values based on X values(Maulud, D., & Abdulazeez., 2020). The goal of linear regression is to determine the values of the parameters β so that the model fits the observed data as well as possible. This is done by minimising the sum of squared errors between the value predicted by the model and the true value of the dependent variable.

Support Vector Regression

Support Vector Regression (SVR) is a regression technique that uses the Support Vector Machines (SVM) approach to model the relationship between dependent and independent variables. Like SVM in classification, SVR is designed to handle cases where the relationship between variables is not linear or has complex patterns(Duan & Huang., 2019). Essentially, SVR conceptualises regression as an optimisation problem, where the main goal is to minimise the prediction error while maintaining maximum margins around the training data.

Dataset

This study uses the linear regression algorithm and Support Vector regression (SVR) methodology. Section I presents a description of the background. This section describes the data, data preparation, visualisation, and modelling of the LR and SVR methodologies. Meanwhile, Section IV presents an explanation of the evaluation and implementation. This study uses the Google Colaboratory for Environment. Also, several modules were utilised, including Python; Pandas for data exploration, filtering, and cleaning; Matplotlib for exploratory Data Analysis by creating visualisations such as histograms, scatter plots, box plots, and line plots; and NumPy as a fundamental package for scientific computing in Python.



	NOPA	NAMAPEL	ADDRESS	BANDWIDTH	ID	Month	Year	LATITUDE	LONGITUDE	PHONE_1_PELANGGAN	KELURAHAN	KECAMATAN	KABUPATEN
0	SPAUCT2101019770	Adhya Warman	SENTRALAND CLUSTER HAWAI BLOK D5 NO 40 SENTRAL...	20 Mbps	102120800042	January	2021	-6.350566	106.572646	087791934668	KABASIRAN	PARUNG PANJANG	KAB. BOGOR
1	SPAUCT2101019771	PURNOMO	PERUMAHAN SARANGAN PERMAI BLOK G3 NO 6 RT 06 R...	20 Mbps	102120800259	January	2021	-6.406175	106.775693	081392197813	SAWANGAN BARU	SAWANGAN	KOTA DEPOK
2	SPAUCT2101019773	Alamyah Alan Nugroho	PERUMAHAN TELAGA MURNI BLOK D 16 NO 35 RT 2 RW...	5 Mbps	102120801960	January	2021	-6.204496	107.111589	089646847524	TELAGAMURNI	CIKARANG BARAT	KAB. BEKASI
3	SPAUCT2101019774	Mukhlisin	JALAN PLN RAYA NO 2 JALAN PLN RAYA NO 2	20 Mbps	102120803261	January	2021	-6.345686	106.764819	08561852023	GANDUL	CINERE	KOTA DEPOK
4	SPAUCT2101019775	Lucky Jack Arthur Tongsiburg	PERUMAHAN PESONA CITRAYAM BLOK A9/7 PERUMAHAN P...	20 Mbps	102120806048	January	2021	-6.462426	106.792149	08211424660	SUSUKAN	BOJONG GEDE	KAB. BOGOR
...	...	...	...	...	...	...	...	...	...	...	...	...	...
10469	SPAUPG2112000137918	Purnomo Iskak	Jalan Sangrila Indah 2 No. 1/Jalan Sangrila In...	100 Mbps		NaN	December	2021	-6.237874	106.759033	0811134449	PETUKANGAN SELATAN	PESANGGRAHAN	KOTA ADM. JAKARTA SELATAN
10470	SPAUPG2112000129959	Nurohman Anshori	Perumahan Bumi Serang Damai D12 No.26 Drangong	100 Mbps		NaN	December	2021	-6.103555	106.134527	08128425194	DRANGONG	TAKTAKAN	KOTA SERANG
10471	SPAUPG2112000131604	VERA WATI	GRAND HARMONI 2 BLOK J3 NO 08 GRAND HARMONI 2	50 Mbps		NaN	December	2021	-6.160760	106.458668	081297349404	BUNAR	SUKAMULYA	KAB. TANGERANG
10472	SPAUPG2112000132961	Makhel Silalahipessy	Jl. Tembus No.41, Gandul, Kec. Cinere, Kota De...	100 Mbps		NaN	December	2021	-6.330451	106.792457	0811891814	CINERE	CINERE	KOTA DEPOK
10473	SPAUPG2112000134159	raden yana hardanaya soehardi	jl. benten I , no 3 rtrw 009/004 sumur batu ...	50 Mbps		NaN	December	2021	-6.161056	106.879686	08155005031	SUMUR BATU	KEMAYORAN	KOTA ADM. JAKARTA PUSAT

10474 rows × 13 columns

Fig 1. Data Raw Customer PLN Icon Plus

Types of Research

This study is part of quantitative research. Where the given output produces data in the form of numbers analysed using statistical methods. The first step in conducting this research is to collect data. The data collected is the customer data in PT PLN Icon Plus (Icon+) during the last one-year period in January 2021–December 2021. This raw dataset has 10474 records with 13 attributes.

Pre-Processing Data

This study is part of quantitative research. Where the given output produces data in the form of numbers analysed using statistical methods. The first step in conducting this research is to collect data. The data collected is the customer data in PT PLN Icon Plus (Icon+) during the last one-year period in January 2021–December 2021. This raw dataset has 10474 records with 13 attributes.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

Filtering Data

The process of selecting a subset of data that meets a specific criterion or removing unnecessary data. This process is done to filter the data so that it can focus on relevant information. At this stage, remove Nopa, Namapel, Address, Io, Latitude, Longitude, Phone_1_Client, Kelurahan, and Districts to focus on the subset of data to be used.

	BANDWIDTH	Month	Year	KABUPATEN
0	20 Mbps	January	2021	KAB. BOGOR
1	20 Mbps	January	2021	KOTA DEPOK
2	5 Mbps	January	2021	KAB. BEKASI
3	20 Mbps	January	2021	KOTA DEPOK
4	20 Mbps	January	2021	KAB. BOGOR
...	...	...	...	...
10469	100 Mbps	December	2021	KOTA ADM. JAKARTA SELATAN
10470	100 Mbps	December	2021	KOTA SERANG
10471	50 Mbps	December	2021	KAB. TANGERANG
10472	100 Mbps	December	2021	KOTA DEPOK
10473	50 Mbps	December	2021	KOTA ADM. JAKARTA PUSAT

10474 rows × 4 columns

Fig 2. Filtering Data

Cleaning Data

Data cleansing is the process of identifying, correcting, and removing inaccuracies, inconsistencies, or irrelevancies in a data set. At this stage there is no duplicate data, missing data, or data that does not match the specified format so there is no reduction in the amount of data and four attributes are used.

```
      BANDWIDTH  Month  Year                KABUPATEN
0       20 Mbps      1  2021                KAB. BOGOR
1       20 Mbps      1  2021                KOTA DEPOK
2        5 Mbps      1  2021               KAB. BEKASI
3       20 Mbps      1  2021                KOTA DEPOK
4       20 Mbps      1  2021                KAB. BOGOR
...         ...    ...   ...                       ...
10469  100 Mbps     12  2021  KOTA ADM. JAKARTA SELATAN
10470  100 Mbps     12  2021               KOTA SERANG
10471   50 Mbps     12  2021             KAB. TANGERANG
10472  100 Mbps     12  2021                KOTA DEPOK
10473   50 Mbps     12  2021   KOTA ADM. JAKARTA PUSAT

[10474 rows x 4 columns]
```

Fig 3. Turn the value in the month column into a number.

Labeling Data

Data labelling is the process of labelling or categorising data based on characteristics. In this phase, a new column is created called Code," and the area is labelled as the area code. Where to label with the following area designation:

1) Code Area 1: Jakarta Area (Central Jakarta Administrative City, West Jakarta Administrative City, East Jakarta Administrative City, South Jakarta Administrative City, North Jakarta Administrative City)
2) Code Area 2: Bogor Area (Bogor City, Bogor Regency)
3) Code Area 3: Depok Area (Depok City)
4) Code Area 4: Tangerang Area (Tangerang City, South Tangerang City, Tangerang Regency)
5) Code Area 5: Bekasi Area (Bekasi City, Bekasi Regency)
6) Code Area 6: Banten Area (Serang City, Serang Regency, and Cilegon City)

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

	BANDWIDTH	Month	Year	KABUPATEN	Kode
0	20	1	2021	KAB. BOGOR	2
1	20	1	2021	KOTA DEPOK	3
2	5	1	2021	KAB. BEKASI	5
3	20	1	2021	KOTA DEPOK	3
4	20	1	2021	KAB. BOGOR	2
...	...	...	...	...	...
10469	100	12	2021	KOTA ADM. JAKARTA SELATAN	1
10470	100	12	2021	KOTA SERANG	6
10471	50	12	2021	KAB. TANGERANG	4
10472	100	12	2021	KOTA DEPOK	3
10473	50	12	2021	KOTA ADM. JAKARTA PUSAT	1

10439 rows × 5 columns

Fig 5. Data after passing the labeling phase.

	Jumlah Pelanggan	BANDWIDTH	Month	Year	Kode
0	1	5	1	2021	5
1	3	10	1	2021	1
2	8	10	1	2021	2
3	7	10	1	2021	3
4	1	10	1	2021	4
...	...	...	...	...	...
247	4	100	12	2021	2
248	15	100	12	2021	3
249	9	100	12	2021	4
250	3	100	12	2021	5
251	2	100	12	2021	6

252 rows × 5 columns

Fig 6. Data through the preprocessing phase

After this stage, the next process calculates and displays the amount of empty data on the new column "Code" that has been created. The result shows there are 35 data points that do not fall into the category of the region, so the line is removed. The following is the data that has passed the stage of the data preprocessing process described in Figure 7. The results show as many as 5 columns with 10439 rows that are ready to enter the next stage, which is to count the number of Customers based on bandwidth, month, year, and code.

Modelling

Modelling in this phase is the process of dividing data into two sub-sets, namely data training and data test. Data trains are used to train predictive models, while test data is used to test the performance of trained models. At this stage, the data train will have a larger proportion than the test data (80% training and 20% test).

```
[22] from sklearn.model_selection import train_test_split
     x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)

     x_train.shape, x_test.shape, y_train.shape, y_test.shape

     ((201, 1), (51, 1), (201, 1), (51, 1))
```

Fig 7. Modeling Data Train & Data Test

Train-data modelling and test data are important to test the performance of the model and ensure that the trained model can produce accurate predictions on data that has never been seen before. This helps to avoid overfitting or underfitting as well as ensuring the applicability of the model to real-world situations.

Visualisation

At the data visualisation stage, creating graphs is the best way to explore and communicate the findings. Data visualisation is the process of making graphic representations of existing data to facilitate its understanding and analysis. Data visualisation can be done using various types of graphics, such as bar graphs, line charts, circular diagrams, maps, or flow charts.

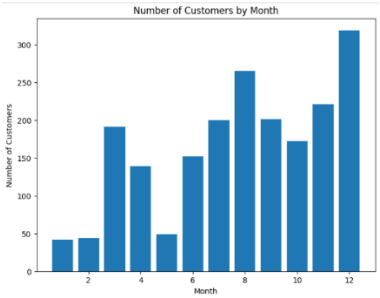


Fig 8. Visualitation Data Model Customers

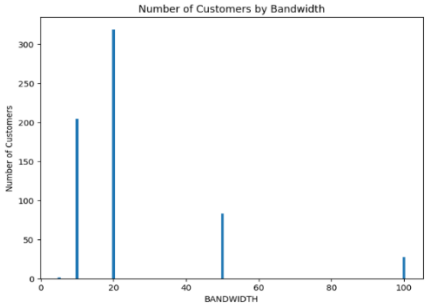


Fig 9. Visualitation Data Model For Bandwidth

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

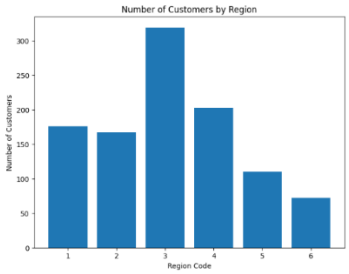


Fig 10. Visualitation Data Model For Region

In Figure 8. Visualitation Data Model Customers, it can be seen that customer growth in 2021 is on average the number of customers is increasing with the highest in December 2021 above 300 customers, while in Figure 9. customers who subscribe to 20 Mbps bandwidth have 300 customers. lastly in Figure 10. the depok area with area code 3 has 300 customers and that is the most.

Linear Regression for Count of Customers

It starts with the determination of the independent variable (axis X), which is the month, and determines the dependent variable, which is the Count of Customers. Then import a linear regression algorithm for modelling and display intercept values and data set feature coefficients.

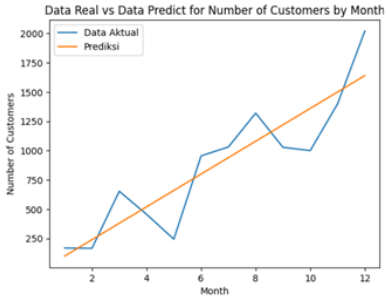


Fig 11. Prediction Count of Customers using Linear Regression

The above image shows a combination of actual customer data and predictive data. Next, represent data that has been in the process using diagrams to facilitate interpretation. You can see the image to the right, where on the 12th month (December), the Count of Customers shows both actual value and prediction.

Linear Regression for Bandwidth

The level is roughly the same as the customer data, where it determines the independent variable (axis X), which is bandwidth, and the dependent variable (axis Y), the Count of Customers. Then import a linear regression algorithm for modelling and display the intercept values and coefficients of the model used.

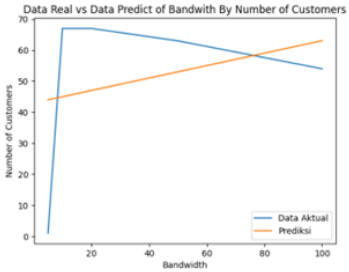


Fig 12. Prediction Bandwidth using Linear Regression

The next step is to represent the data that has been processed with the linear regression algorithm. Based on the image above, bandwidths 10 and 20 show the same results, but on the predictive line, set the bandwidth 100 that holds the highest Count of Customers.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

Linear Regression for the Region

At this stage, data is created to calculate customers based on the month based on their territory; the data is determined by independent variables (axis X) as the region and dependent variables as the Count of Customers. It then imports a linear regression algorithm for modelling and displays the intercept value and its coefficient value.



Fig 13. Prediction Region using Linear Regression

The above image shows actual data with predictive data that has values that are close to The data is processed using the linear regression algorithm. Where the prediction results show a decline, although the actual data (distributed by a dot pattern) displays a graph that is not much different from areas 1–6.

Support Vector Regression for the Count of Customers

Support vector regression (SVR) is a type of machine learning algorithm used for regression analysis. The goal of SVR is to find a function that approximates the relationship between the input variables and the continuous target variable, while minimizing prediction error. Unlike support vector machines (SVM) used for classification tasks, SVR seeks to find a hyperplane that best fits data points in a continuous space. This is achieved by mapping the input variables into a high-dimensional feature space and finding the hyperplane that maximizes the margin (distance) between the hyperplane and the nearest data points, while minimizing the prediction error. SVR can handle nonlinear relationships between input variables and target variables by using a kernel function to map data into a higher dimensional space. This makes it a powerful tool for regression tasks where there may be complex relationships between input and target variables. Support vector regression (SVR) uses the same principle as SVM, but for regression problems. Take a few minutes to understand the idea behind SVR.

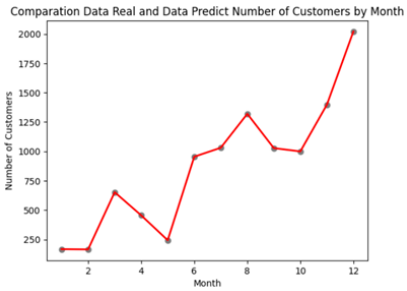


Fig 14 Prediction Count of Customers using SVR.

In the above implementation, 'X_train' and 'y_trains' are training data that contains the Value of Customer Number and Month, while 'X_test' is the test data with the value of the Count of Customers to be predicted. The variables `y_train` and `i_test` are the target values to be predicted, in this case month." As shown in the image above, the data shows month 12 (December) has the largest Count of Customers, which is the number of customers in its forecast for 2019.

Support Vector Regression for Bandwidth

The stages are like the customer number calculation: 'X_train' is the training data that contains the independent variable (Bandwidth), while 'y_trains' are the dependent variable. (Jumlah Pelanggan dalam kasus ini). The variable 'X_test' is the test data that you want to use to make predictions. The variable `y_pred` is the result of the prediction of the SVR model.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

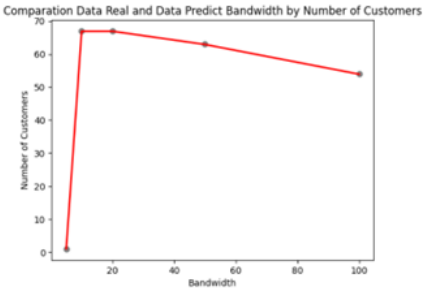


Fig 15 Prediction Bandwidth using SVR.

As shown in the image above, the data showed that bandwidths 10 and 20 had the same value, and bandwidth 5 showed the least number of Customers. This shows customer interest levels at 10 and 20 Mbps higher than others, while the bandwidth with the lowest interest is at 5 Mbps.

Support Vector Regression for Region

X_train is the training data that contains an independent variable (Areas), while 'y_trains' is a dependent variable (Jumlah Pelanggan dalam kasus ini). The variable 'X_test' is the test data that you want to use to make predictions. The variable `y_pred` is the result of the prediction of the SVR model.



Fig 16 Prediction Region using SVR.

From the image above, it shows pattern similarities between the predicted line and the actual line, where the area and Count of Customers have close values, as seen in Figure 16. This indicates that the use of the SVR model on growth predictions has values that are almost like actual values. The highest area is located on Code 3, or the Depok area.

RESULT

This research uses secondary data on the company PT PLN Icon+ from January to December 2021 for one year with 10474 data obtained and 12 columns/attributes with the link https://cssritel.iconpln.co.id/PA/RawData. The next process for the data in Pre-Processing includes cleaning the data so that there is no duplicate data, missing data, or data that does not comply with the specified format so that there is no reduction in the amount of data and the four attributes used as well as labeling the customer area. After the data is declared clean, then divide the data into two subsets, namely training data and test data. Train data is used to train a predictive model, while test data is used to test the performance of the model that has been trained. At this stage, the training data will have a larger proportion than the test data, namely 80% training and 20% test with a function to test model performance. The test results show that the Support Vector Regression (SVR) algorithm is superior because the error rate or mean absolute percentage error (MAPE) is smaller, namely 0.02%, while for linear regression the error rate or MAPE is greater, namely 36.28%, because SVR can handle relationships. nonlinearity between input and output variables as well as kernel functions to map data to higher dimensions, allowing the model to capture more complex patterns. In table 1 below, the test results from secondary data on the PT PLN Icon+ company for the 2021 period are given.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

Table I.
Calculation Of LR And SVR Models

Predict	Linear Regression			Support Vector Regression		
	RMSE	*MAPE %*	*MAE*	*RMSE*	*MAPE %*	*MAE*
Customers	0,80	36,2	201	0,99	0,02	0,10
Bandwidth	0,07	878	20	0,99	2	0,09
Area	0,46	19,36	6	0,99	0,25	0,09

DISCUSSIONS

As we can see in Table 1, the value of RMSE on the SVR algorithm is higher with a value of 0.99 Customer Number, 0.99 Bandwidth, and 0.99 Region compared with a linear regression RMSE value with a rating of 0.80 customer Number, 0.07 bandwidth, and 0.46 region. However, in the MAPE calculation, the Linear Regression value has a high accuracy score compared to SVR, where 36.28% Customer Number, 878% Bandwidth, and 19.36% Region are compared with the SVR algorithm map that is 0.02% customer Number, 2% bandwidth, and 0.25% Region. The Linear MAE accuracy value is higher with 201 Customer numbers, 20 bandwidths, and 6 Regions compared to the SVR MAE value of 0.10 Customer totals, 0.09 bandwidths, and 0.09 Regions. The calculation value is obtained from the calculation formula for the values of RMSE, MAPE, and MAE in Chapter 2. This makes accuracy calculations through RMSE, MAPE, and MAE score values using the SVR algorithm better in its implementation than using the Linear Regression algorithm.

Based on the results of research based on comparative testing of Iconnet Jabodetabek and Banten Customer Growth Forecasting, it results that the Support Vector Regression (SVR) algorithm is superior because it has a smaller error rate or mean absolute percentage error (MAPE) of 0.02%, while for linear regression the error rate or MAPE is greater, which is 36.28%. In addition, SVR can handle nonlinear relationships between input and output variables and kernel functions to map data to higher dimensions, allowing the model to capture more complex patterns.

CONCLUSION

Based on the results of data visualisation that customer growth in 2021 the number of customers increased by 300 customers with a bandwidth of 20 Mbps and is in the Depok area. while the results of testing the comparison of Iconnet Jabodetabek and Banten Customer Growth Forecasting show that the Support Vector Regression (SVR) algorithm is superior because it has a smaller error rate or mean absolute percentage error (MAPE) of 0.02%, while for linear regression the error rate or MAPE is greater, which is 36.28%. In addition, SVR can handle nonlinear relationships between input and output variables and kernel functions to map data to higher dimensions, allowing the model to capture more complex patterns. Suggestions for the future should be data multiplied or added again and comparison trials with other algorithms are carried out so that significant and good results can be seen.

REFERENCES

Anand, P., Rastogi, R., & Chandra, S. (2020). A class of new support vector regression models. Applied Soft Computing, 94, 106446.

Dewi, C., & Chen, R. C. (2019). Random forest and support vector machine on features selection for regression analysis. Int. J. Innov. Comput. Inf. Control, 15(6), 2027-2037.

Duan, Z., Liu, Y., & Huang, K. (2019, May). Mobile phone sales forecast based on support vector machine. In Journal of Physics: Conference Series (Vol. 1229, No. 1, p. 012061). IOP Publishing.

Fitrianah, D., Gunawan, W., & Sari, A. P. (2022). Studi Komparasi Algoritma Klasifikasi C5. 0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir. Techno. Com, 21(1), 1-11.

He, F. (2023, August). iPhone Sales Prediction Based on Multilinear Regression Model: Evidence from Statista. In 2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE) (pp. 341-346). IEEE.

Maulana, N A. The faithful b. D., and C. D. (2019). Implementation of the Support Vector Regression (SVR) Method in Bread Sales Prediction (Studi Kasus: Harum Bakery). Journal of Information Technology and Computer Science Development.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3362

Submitted : Dec 23, 2023
Accepted : Dec 30, 2023
Published : Jan 1, 2024

Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.

McKim, B., & Hughes, A. M. (2001). How to measure customer relationship management success. Journal of Database Marketing & Customer Strategy Management, 8, 224-231.

Nastiti, M. D., Abdurohman, M., & Putrada, A. G. (2019, July). Smart shopping prediction on smart shopping with linear regression method. In 2019 7th International Conference on Information and Communication Technology (ICoICT) (pp. 1-6). IEEE.

Pratama, A. D., & Noviana, W. (2023). Perancangan Sistem Informas Aplikasi E-Booking Berbasis Website Menggunakan Metode Extreme Programming Pada PT Indonesia Comnets Plus. LOGIC: Jurnal Ilmu Komputer dan Pendidikan, 1(6), 1532-1535.

R. A. The Princess W. S. Winnie and M. Mashuri, "Application of Ridge Regression and Support Vector Regresion (SVR) Method for Predicting Coal Index in PT XYZ," J. Science and Arts ITS, vol. 9, no. 1, pp. 64–71, 2020, doi: 10.12962/j23373520.v9i1.51021.

Sandiwarno, S. (2018). Developing an E-Forum to Universitas Mercu Buana Alumni's to Improve Effective Communication and Educative by Technology Multimedia Acceptance Model. Int. J. Comput. Sci. Mob. Comput, 7(8), 113-122.

Sianturi, D. P. S., & Sagala, J. R. (2020). Prediction of 2020 Mobile Sales Trends Using the Weighted Product Method. Journal of Intelligent Decision Support System (IDSS), 3(4), 25-36.

Wen, Z., Zhang, R., Ramamohanarao, K., & Yang, L. (2018). Scalable and fast SVM regression using modern hardware. World Wide Web, 21, 261-287.

Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In Machine learning (pp. 123-140). Academic Press.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.