

IMPLEMENTATION OF SUPPORT VECTOR REGRESSION IN THE PREDICTION OF THE NUMBER OF TOURIST VISITS TO THE PROVINCE WEST NUSA TENGGARA (NTB)

Zaeniah¹; Zaenudin^{2*}; Bahtiar Imran³

Sistem Informasi¹, D3 Komputerisasi Akuntansi, S1 Teknik Sistem Komputer
Universitas Teknologi Mataram
Kota Mataram, Indonesia
<https://utmmataram.ac.id/>
zaen1989@gmail.com, zen3d.itb@gmail.com, bahtiarimranlombok@gmail.com

(*) Corresponding Author

Abstract — Indonesia has a variety of exciting tourist destinations to visit in each region. One area that is used as a favorite tourist destination is the Province of West Nusa Tenggara (NTB). Data, The number of tourists visiting the NTB province from 2014 to 2020 tends to change based on data obtained from the NTB Provincial Tourism Office website. The data on the number of visitors will continue to change, even if it is possible to increase. It can lead to the unpreparedness of the government and other tourism actors to provide the facilities and infrastructure needed by visitors when there is an increase in the number of tourist visits coming to NTB. Therefore, it is necessary to predict the number of tourist visits to NTB with accurate results. In this study, predictions of the number of tourist visits to the Province of NTB were made using the support vector regression method. This research resulted in an application to predict the number of tourist visits to NTB based on Event, Month, and Year. Provide predictive results close to the actual value under normal conditions. The data used in this study is data on the number of tourist visits in 2017-2021 and events held in 2017-2021.

Keywords: NTB Tourism, Support vector regression.

Abstrak—Indonesia mempunyai berbagai destinasi wisata yang menarik untuk di kunjungi pada masing-masing daerah. Salah satu daerah yang dijadikan sebagai tujuan wisata favorit adalah Provinsi Nusa Tenggara Barat (NTB). Data Jumlah wisatawan yang berkunjung ke provinsi NTB sejak tahun 2014 sampai tahun 2020 cenderung mengalami perubahan berdasarkan data yang diperoleh dari Website Dinas Pariwisata Provinsi NTB. Data jumlah pengunjung akan terus mengalami perubahan bahkan tidak menutup kemungkinan akan mengalami peningkatan. Hal ini dapat menyebabkan ketidaksiapan pemerintah dan pelaku wisata lainnya dalam menyediakan sarana dan prasarana yang dibutuhkan oleh pengunjung pada saat terjadinya peningkatan jumlah kunjungan

wisatawan yang datang ke NTB. Oleh karena itu, dibutuhkan adanya prediksi jumlah kunjungan wisatawan ke NTB dengan hasil yang akurat. Pada penelitian ini dilakukan prediksi jumlah kunjungan wisatawan ke Provinsi NTB menggunakan metode support vector regression. penelitian ini menghasilkan suatu aplikasi prediksi jumlah kunjungan wisatawan yang datang ke NTB berdasarkan Event, Bulan dan Tahun. Hasil prediksi yang mendekati nilai aktual dalam kondisi normal. Data yang digunakan pada penelitian ini adalah data jumlah kunjungan wisatawan tahun 2017–2021 serta event yang diselenggarakan pada tahun 2017–tahun 2021.

Kata Kunci: Wisata NTB, Support vector regression.

INTRODUCTION

Indonesia has various exciting tourist destinations in each region (Putriwijaya, Novi Nur & Wijayaningrum, 2019). Tourism has significantly contributed to the Indonesian economy (Kurniawati & Yantri, 2018). One of the reasons for the increasing regional income is the presence of tourists (Rahmawati, Yuniza, Rahma, & Zukrianto, 2020). One area that is used as a favorite tourist destination is the Province of West Nusa Tenggara (NTB). The tourism sector is the mainstay sector in the current development of the NTB Province (Rizal & Hartati, 2017). The tourism sector is also one of the largest sources of income for the people of NTB. The development of the tourism sector in the province of NTB is strongly influenced by the number of domestic and foreign tourist visits (Lestari, Albar, & Afwani, 2019). Data, The number of tourists visiting the province of NTB from 2014 to 2020 tends to change based on data obtained from the Website of the Tourism Office of the Province of NTB (<http://www.disbudpar.ntbprov.go.id/>).

In 2014 the number of tourist visits was 1,629,122 and increased in 2015, with the number of tourist visits amounting to 2,210,527. In 2016 the number of tourist visits increased by 3,094,437; in 2017, it also increased to 3,761,686. In 2018, the number of tourist visits to the province of NTB decreased to 2,812,379. In 2019 the number of tourist visits increased to 3,706,352. In 2020 the number of tourist visits decreased due to the covid 19 pandemic to 400,595 visitors coming to the province of NTB (Maulana, Setiawan, & Dewi, 2019). Development in the tourism sector must be adjusted to the number of tourist visits in previous years and make predictions about visits in the coming year (Rais et al., 2020). The data on the number of visitors will continue to change, even if it is possible to increase. An increase in the number of tourist visits must be anticipated with facilities that can support the progress of tourism (Rais et al., 2020). An increase in the number of uncertain visitors will impact the local government's level of readiness to deal with the surge in the number of tourists coming to NTB. Therefore, it is necessary to predict the number of tourist visits to NTB with accurate results. The prediction of the number of tourist visits has undoubtedly been carried out by the Tourism Office but only based on events that will take place in the future, so it does not produce prediction numbers that are close to the actual value. Prediction with accurate results is not enough just with events but must use the proper prediction method. In this study, predictions of the number of tourist visits to the Province of NTB will be made using the support vector regression method because, based on previous research, this method has a high level of accuracy.

Support Vector Regression (SVR) is a theory adapted from the theory of Support Vector Machine (SVM), namely machine learning theory, that is used to solve classification problems. (Maulana et al., 2019). One of the methods that can be used to make predictions that depend on time is Support Vector Regression (Sepri et al., 2020). SVR implements the SVM algorithm in solving regression cases (Furi, Si, & Saepudin, 2015). The SVM method produces values in the form of integers according to machine learning theory in the classification case, while the SVR method will produce outputs in the form of real numbers to solve the regression case. (Prakoso, 2019). The concept of the SVR algorithm can solve the overfitting problem so that it can produce better predictive values. Overfitting is the behavior of data during training that can produce almost perfect prediction accuracy (Maulana et al., 2019). The purpose of the SVR algorithm is to find the best hyperplane, called the dividing line (Maulana et al., 2019). The best dividing line can be found by measuring the margin with the dividing line

(Maulana et al., 2019). Margin is the distance from the dividing line (hyperplane) to the closest data. The data closest to the margin is called the support vector (Furi et al., 2015).

Prediction is a method that can be used to determine a value or a need in the next period (Prakoso, 2019). Generally, there are two types of predictions, qualitative and quantitative. Qualitative predictions are subjective predictions, and this is based on practical experience, decision-making intuition, and human emotions. Meanwhile, quantitative prediction is a prediction that is objective because it is based on actual data and processed using specific methods (Surtiningsih, Furqon, & Adinugroho, 2018).

A study by Vivi Nur Wijyaningrum and Novi Nur Putriwijaya in 2020 entitled "Support Vector Regression to Predict the Number of Foreign Tourist Visits on Bali" discusses specific predictions for the number of foreign tourist arrivals to Bali using the Support method. Vector Regression (Putriwijaya, Novi Nur & Wijyaningrum, 2019).

Komang Triantita Neti Lestari, Moh conducted the following research. Ali Albar and Royana Afwani in 2019, entitled "Application of the Backpropagation Method in Predicting Numbers. Tourist Visits to the Province of West Nusa Tenggara (NTB)" This study uses the Backpropagation method. It has not added an event as a variable to determine the prediction results. (Lestari et al., 2019). While the research will be conducted using the support vector regression method, which can produce better accuracy in the case of regression using data from 2014-2020 and adding event variables to determine prediction results with more accurate results. A literature review of no more than 1000 words state of the art in the researched field/technology developed—relevant library sources/primary references and prioritizing results research in scientific journals and up-to-date patents.

Teri Kincowati, Muhammad Tanzil Furqon, and Bayu Rahayudi's 2019 research entitled "Predicting the Number of Foreign Tourist Visits to Indonesia Using the Average-Based Fuzzy Time Series Models Method" this study discusses the prediction of the number of foreign tourist arrivals to Indonesia with the results which are pretty good using the Average-Based Fuzzy Time Series Models method (Kincowati, Furqon, & Rahayudi, 2019). Meanwhile, the research that will be conducted discusses the prediction of the number of tourists visiting NTB abroad and the archipelago using the support vector regression method, which can make predictions with very good or almost perfect results. Meanwhile, the research that will be conducted discusses the prediction of the number of tourists visiting NTB abroad and the archipelago using the

support vector regression method, which can make predictions with very good or almost perfect results. (Putri, Naja, Pratama, & Widodo, 2021). Meanwhile, in the research that will be conducted, it will discuss the prediction of the number of tourist visits coming to NTB both from abroad and the archipelago with the support vector regression method, which can predict with excellent results seen from the data of the last seven years and based on events that will be held on the next period in the province of NTB.

Aida Meimela's 2021 research entitled "Prediction of the Number of Foreign Tourist Visits to Indonesia" uses the seasonal ARIMA method to predict the number of foreign tourist visits to Indonesia, with prediction results having a good level of accuracy. (Meimela, 2021). While the research that will be conducted is predicting the number of tourist visits to NTB using the SVR method, which has perfect accuracy based on the results of previous studies.

From previous studies, many have made predictions on the number of tourist visits, with prediction results having a good level of accuracy by only calculating the prediction results from data on the number of tourist visits in previous years. In previous research, the stages carried out were only to evaluate the prediction results. Just. Meanwhile, in this study, it was carried out until the deployment stage so that it could be used to predict the number of tourist visits in the next period. In addition, the object of this research was carried out in the Province of NTB by calculating the number of tourist visits, both foreign and domestic, and predictions were made not only from visit data in the previous year but also based on events to be held in the next period to produce predictions with a high level of accuracy. Tall. The difference between the previous research and this research is that it uses different methods and different research objects. The tourism office can use the results of this prediction to predict the number of tourist visits in the next period to prepare resources and make decisions in the face of changes in the number of tourists visiting the province of NTB.

MATERIALS AND METHODS

The research method used in this study follows the steps in application development using the SKKNI standard no. 299 of 2020, this activity is carried out until the deployment stage.

1. Business Understanding: Determining Business Problems

Developing an intelligence system from data begins with determining the problem to be solved. The activities determine business and technical objectives and make project plans at this stage.

2. Data Understanding

Data understanding activities are activities to understand data more deeply. The three main activities that will be carried out at this stage are as follows: collecting data, reviewing data, and validating data. The data is taken from the website of the NTB tourism office at the following link: <https://data.ntbprov.go.id/dataset/jumlahkunjungan-wisatawan-ke-provinsi-nusa-tenggarabarat-ntb>

3. Data Preparation

Data preparation uses the analysis results to change the data so that the quality of the data increases and can improve the modeling process. The activities at this stage are sorting, cleaning, constructing, determining labels, and integrating data.

4. Modeling

The next stage is using Machine Learning (ML) algorithms to form a model from the data that has been repaired/prepared. The target of modeling is to find the best model that can be found using the processed data. The model used at this stage is the support vector regression model. The activity carried out at this stage is to build a model.

5. Model Evaluation

In this step, the measurement of the model's performance that has been obtained is carried out, and an analysis of whether the model is good enough from a technical and business point of view (domain) will be used. At this stage, the activities carried out are evaluating the modelling results and reviewing the modelling process.

6. Deployment

It is installing the model as part of the intelligence application (system) to be used according to the task and purpose of its manufacture. At this stage, what is done is to make a model deployment plan and deploy the model? The flowchart for the research stages is in Figure 1. Flowchart of Research Stages:

The results and discussion of this research are as follows:

1. Business Understanding

In this study, problems related to the number of tourist visits are uncertain and constantly changing, so local governments and other tourism actors need to predict the number of tourist visits to NTB and approach the number of original visitors. It can help local governments and other tourism actors prepare facilities and infrastructure visitors will use.

2. Data Understanding

This research produces an output in the form of an application that can predict the number of tourist visits to the NTB province using the support vector regression method. The dataset

needed is the number of tourist visits coming to NTB.

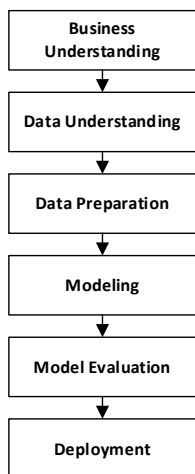


Figure 1 Flowchart of Research Stages

RESULTS AND DISCUSSION

The data downloaded from the tourism office website is for the last five years, from 2017 - 2021. The data used are.

- The number of tourist visits and events in 2017 from January to December
- The number of tourist visits and events in 2018 from January to December
- The number of tourist visits and events in 2019 from January to December
- Number of tourist visits and events in 2020 from January to December
- Number of tourist visits and events in 2021 from January to December

The attributes used in determining the number of tourist visits to NTB are year, month, and events held in a certain period.

Data correlation between attributes can be seen in Figure 2 below :

	Tahun	Bulan	Event	Jumlah_Wisatawan
Tahun	1.00	-0.04	-0.20	-0.62
Bulan	-0.04	1.00	-0.00	-0.09
Event	-0.20	-0.00	1.00	0.22
Jumlah_Wisatawan	-0.62	-0.09	0.22	1.00

Figure 2 Attribute correlation

According to Figure 2, the event is the correlation between the attributes that most influence the number of tourist visits coming to NTB. The data obtained is the problem to be solved to predict the number of tourist visits. It is necessary to have data on the previous year, month, and events.

3. Data Preparation

The data taken on the NTB tourism office website is very high-quality data at this stage, which is done only by changing the data format using the object format into a number format, namely the month and event attributes. The following are the first 10 data used as samples for changing object formats to number formats. The attributes used as the X variable are Year, Month, and Event. At the same time, the attribute that will be used as the Y variable is the Number of Tourists attribute.

Table 1 data preparation

	Tahun	Bulan	Event	Jumlah Wisatawan
0	2017	5	44	169954
1	2017	12	2	188484
2	2017	8	10	183310
3	2017	2	42	274332
4	2017	9	37	299220
5	2017	7	4	380119
6	2017	6	18	475420
7	2017	0	2	469215
8	2017	13	22	382789
9	2017	11	28	348751

4. Modeling

SVR modeling with linear kernel. This model will be built using the support vector regression method. Before building the model, the data will be separated for training and test data. In this study, 80% of the training data will be used, and 20% of the test data will be used. The following is a display of the X variable, which is used as test data can be seen in Figure 3 below:

```

X_test
array([[2017, 5, 44],
       [2017, 7, 4],
       [2020, 5, 25],
       [2020, 11, 13],
       [2018, 12, 6],
       [2021, 6, 20],
       [2019, 11, 27],
       [2021, 5, 25],
       [2018, 5, 26],
       [2021, 11, 19],
       [2020, 10, 17],
       [2021, 8, 8]], dtype=int64)
    
```

Figure 3 variables X Test Data

The following display of the y variable used as test data can be seen in Figure 4 below:

```

y_test
array([169954, 380119, 67864, 40490, 214185, 58677, 321217,
       68859, 194107, 95515, 45050, 66932], dtype=int64)
    
```

Figure 4 displays variable y

The following is a display of the X variable used by the training data can be seen in Figure 5 below:

```
X_train
array([[2019, 0, 21],
       [2017, 2, 42],
       [2021, 9, 34],
       [2018, 7, 34],
       [2017, 13, 22],
       [2017, 6, 18],
       [2020, 9, 34],
       [2017, 9, 37],
       [2020, 0, 1],
       [2018, 0, 1],
       [2019, 10, 33],
       [2021, 10, 17],
       [2019, 12, 5],
       [2021, 13, 23],
       [2018, 2, 43],
       [2019, 2, 15],
       [2017, 11, 28],
       [2019, 6, 11],
       [2019, 8, 43],
       [2018, 9, 30],
       [2019, 5, 26],
       [2021, 1, 3],
       [2017, 3, 39],
```

Figure 5 display of variable x

The following y variable used as training data can be seen in Figure 6 below:

```
y_train
array([[481159, 274332, 70779, 438699, 382789, 475420, 2610, 2
99220,
       29932, 19758, 347672, 62425, 134676, 80023, 456772, 2
56255,
       348751, 458910, 141559, 419317, 127269, 57752, 289733, 2
89672,
       89219, 9085, 60675, 413723, 27913, 188484, 58395, 1
83310,
       53539, 2790, 436892, 142538, 68061, 300359, 58812, 5
17032,
       50871, 83608, 469215, 18681, 209156, 297348, 70102,
41966],
      dtype=int64)
```

Figure 6 variable Y training data

The results of the prediction of the number of tourist visits to the province of NTB can be seen in the following picture can be seen in Figure 7 below

```
hasil=svr.predict([[2022,5,44],[2022,10,17]])
```

```
hasil
```

```
array([[142323.5, 137580.5])
```

Figure 7 Predicted Results of Tourist Visits

5. Evaluasi Model

Evaluation is critical to developing the model because evaluating it can provide knowledge to make improvements to the model in the future. The metric used to evaluate the model in this study is R-Squared. The R-squared metric is a metric that represents the extent to which the data fits the regression line obtained. The larger the R-squared, the better the line matching to the data will be. The best value obtained using this metric is 1.0 and can

be damaging. The results of the evaluation can be seen in the following figure can be seen in Figure 8 below.

```
from sklearn.metrics import r2_score
coef=r2_score(y_test,hasil)
coef
-0.00514387666852234
```

Figure 8 Evaluation Results

The results of testing the difference between the actual value and the predicted value can be seen in table 2 below.

Table 2 Evaluation results

No	Years	Month	Event	Actual	Prediction
0	2017	5	44	169954	142558.5
1	2017	7	4	380119	135580.5
2	2020	5	25	67864	139111.5
3	2020	11	13	40490	136969.5
4	2018	12	6	214185	135836.5
5	2021	6	20	58677	138185.5
6	2019	11	27	321217	139452.5
7	2021	5	25	68859	139064.5
8	2018	5	26	194107	139379.5
9	2021	11	19	95515	137966.5
10	2020	10	17	45050	137674.5
11	2021	8	8	66932	136079.5

From the table above, it can be analyzed that the prediction results will approach the actual value under normal conditions. 2019, 2020, and 2021 are still in pandemic condition, so the number of tourist visits has decreased.

6. Deployment

The deployment stage is the last stage of this research. The activities implement the test results into real applications with accurate data in the field. Applications made web-based with the following display can be seen in Figure 9 below:

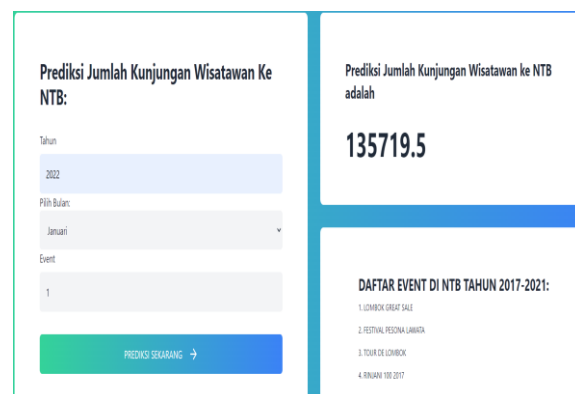


Figure 9 web-based tourist visit prediction application

From the results of the deployment in Figure 1.9, predictions of visits can be made in the years to come on an ongoing basis by events registered in previous years.

CONCLUSION

The number of tourist visits to West Nusa Tenggara (NTB) continues to change. It can lead to the unpreparedness of the government and other tourism actors to provide the facilities and infrastructure needed by visitors when there is an increase in the number of tourist visits coming to NTB. Therefore, this study resulted in an application to predict the number of tourist visits to NTB based on Events, Months, and Years. Prediction is made using the support vector method. Regression so that it can provide prediction results that are close to the actual value under normal conditions. The data used in this study is data on the number of tourist visits in 2017–2021 and events held in 2017–2021. Further research is expected to add new events held in NTB.

REFERENCE

- Furi, R. P., Si, M., & Saepudin, D. (2015). Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus : IHSG dan JII. *ISSN: 2355-9365 E-Proceeding of Engineering* ;, 2(2), 1–10.
- Kincowati, T., Furqon, M. T., & Rahayudi, B. (2019). Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Indonesia Menggunakan Metode Average-Based Fuzzy Time Series Models. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(6), 5250–5256.
- Kurniawati, E., & Yantri, O. (2018). Pemodelan Jumlah Kunjungan Wisatawan Mancanegara Di Batam Dengan Menggunakan Arima Dan Regresi Time Series Modelling Foreign Tourist Arrival in Batam Using Arima and Time Series Regression. *Dimensi*, 7(3), 593–606.
- Lestari, K. T. N., Albar, M. A., & Afwani, R. (2019). Penerapan Metode Backpropagation Dalam Memprediksi Jumlah Kunjungan Wisatawan Ke Provinsi Nusa Tenggara Barat (NTB). *Journal of Computer Science and Informatics Engineering (J-Cosine)*, 3(1), 39–48. <https://doi.org/10.29303/jcosine.v3i1.236>
- Maulana, N. D., Setiawan, B. D., & Dewi, C. (2019). Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2986–2995.
- Meimela, A. (2021). Prediksi Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia. *Media Wisata*, 19(1), 34–41. <https://doi.org/10.36276/mws.v19i1.64>
- Prakoso, B. H. (2019). Implementasi Support Vector Regression pada Prediksi Inflasi Indeks Harga Konsumen. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(1), 155–162. <https://doi.org/10.30812/matrik.v19i1.511>
- Putri, W. L., Naja, F., Pratama, T., & Widodo, E. (2021). Prediksi Kunjungan Wisatawan Mancanegara Ke Dki Jakarta Pada Masa Covid-19 Menggunakan Metode Des Holt. *Journal of Mathematics Education and Science*, 4(2), 81–90. <https://doi.org/10.32665/james.v4i2.243>
- Putriwijaya, Novi Nur & Wijyaningrum, V. N. (2019). Support Vector Regression Untuk Memprediksi Jumlah Kunjungan Wisatawan Mancanegara di Pulau Bali. *Machine Learning: Methods and Applications to Brain Disorders*, 11(January), 123–140.
- Rahmawati, Yuniza, Rahma, A. N., & Zukrianto. (2020). Prediksi Jumlah Wisatawan di Kota Pekanbaru pada Tahun 2019- 2023 Dengan Menggunakan Metode Fuzzy Time Series Chen. *Jurnal Pendidikan Matematika*, 2(1), 36–44.
- Rais, A. N., Rousyati, R., Thira, I. J., Kholifah, D. N., Purwati, N., Kristania, Y. M., & Kristania, Y. M. (2020). Evaluasi Metode Forecasting pada Data Kunjungan Wisatawan Mancanegara ke Indonesia. *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(2), 104–115. <https://doi.org/10.31294/evolusi.v8i2.8971>
- Rizal, A. A., & Hartati, S. (2017). Prediksi Kunjungan Wisatawan di Pulau Lombok dengan Menerapkan Recurrent Neural Network dengan Algoritma Training Extended Kalman Filter. *Jurnal Ilmiah ILMU KOMPUTER*, X(1), 7–18.
- Sepri, D., Fauzi, A., Wandira, R., Riza, O. S., Wahyuni, Y. F., & Hutagaol, H. (2020). Prediksi Harga Cabai Merah Menggunakan Support Vector Regression. *Computer Based Information System Journal*, 2, 1–5. Retrieved from <http://ejournal.upbatam.ac.id/index.php/cbis/article/view/1921>
- Surtiningsih, L., Furqon, M. T., & Adinugroho, S. (2018). Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2578–2586. Retrieved from <http://j-ptiik.ub.ac.id>