Data mining for forecasting community mobility denpasar city with long short-term memory method

I Wayan Agus Hery Setiawan [a,1], Evi Triandini [a,2,*], I Ketut Putu Suniantara [a,3], Djoko Kuswanto [b,4]

[a] Information System, Faculty Information and Computer, ITB STIKOM Bali, Denpasar, 80234, Indonesia
[b] Department of Medical Technology, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia
[1] setiawan@stikom-bali.ac.id; [2] evi@stikom-bali.ac.id; [3] suniantara@stikom-bali.ac.id; [4] crewol@prodes.its.ac.id
* corresponding author

ARTICLE INFO

Article history
Received August 8, 2024
Revised October 18, 2024
Accepted December 3, 2024

Keywords
Data mining
Forecasting
Long short-term memory
Community mobility

ABSTRACT

Denpasar City has a high potential for community mobility, this is supported by many public facilities. High and highly volatile human mobility causes the transmission of the COVID-19 virus to spread very quickly, so forecasting is needed to find out a picture of future community mobility using data mining techniques. Data mining is the process of solving problems by analyzing data that already exists in the database. Denpasar City community mobility data for the period September 1, 2021 – October 31, 2021 show that most of the high mobility is in the junior high school sector. The Long Short-Term Memory method was chosen as a method that can assist in forecasting community mobility. Long Short-Term Memory has the advantage of dealing with missing gradient problems and can be used on all types of data patterns, whether trend, cyclical, seasonal, or horizontal patterns. Hyperparameter tests were carried out including LSTM_units representing the number of Long Short-Term Memory units in each layer, Dropout, and Optimizer to obtain the optimal prediction method. this combination yields a total of 45 methods. The best hyperparameter obtained is at LSTM_units of 128, Dropout of 0.1, and Optimizer is Adam. The results obtained with this hyperparameter are the Root Mean Square Error (RMSE) value of 971,438687. This method results in forecasting the mobility of the people of Denpasar City from November 1, 2021 to November 7, 2021, reaching 9.550 total checkins which is close to the actual value of 10.219.

1. Introduction

The Sars-Cov-2 virus was first reported in the Wuhan city, China to be the cause of the Covid-19 pandemic to date. The COVID-19 pandemic has led to predictions that it will last a long time so that it presents a new order of life or a new normal [1]. All areas of life are affected by this pandemic so humans must be able to adapt. Human mobility is a very interesting aspect to discuss because it has a role in improving important aspects of life. The Covid-19 pandemic has made problem to it had a major impact on the speed of transmission of the Covid-19 virus [2].

In Fig. 1, it is recorded that positive confirmed cases worldwide as of November 17, 2021 have increased and are predicted to increase in the third wave [3]. This increase in confirmed cases is due to one of the factors of human mobility which has increased again in line with the implementation of the vaccine program. Denpasar City has the potential for high human mobility because it is the provincial capital and has a variety of cultural activities. Data from the Indonesian Ministry of Health as of October 14, 2021 with a positive case rate of 2.41/100.000 population/week, shows that the COVID-19 virus is experiencing a downward trend. However, based on the data, the tracing rate in Denpasar City is quite limited with a tracing ratio of 3.88/week. Several obstacles have caused the tracing ratio to be included in a limited category, one of which is because the public is still not open to providing close contact information and the emergence of negative stigma in the community regarding COVID-19 [4].

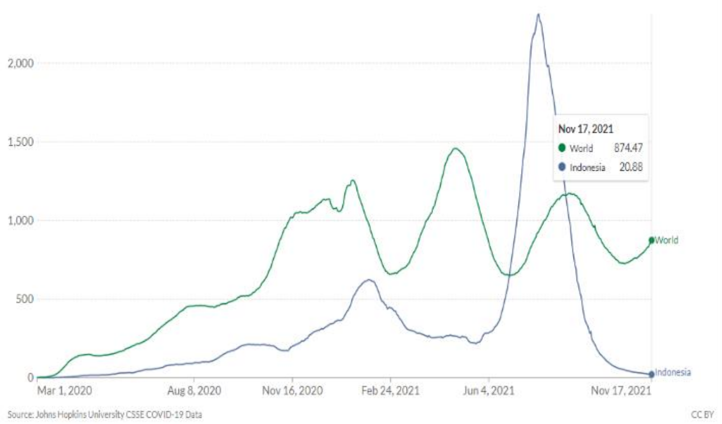

Fig. 1. Covid-19 cases

The Ideathon Bali Kembali Research conducted by ITB STIKOM Bali in collaboration with PT. Bamboo Media offers an application called SpeedID that was able to assist the tracing and tracking of public mobility. This system or application helps the community by scanning the qrcode that is available at public facilities. The data recorded to the database are form of place, time, name, telephone number and email information. The use of applications for a long period of time certainly results in big data [5]. Big data that has been collected basically has not been able to directly produce any information or knowledge. Science is developing rapidly, so is data. Currently data can become Big Data because it has a very large volume of data and has the potential to become even bigger over time. Data science can refer to the interdisciplinary science between statistics and technology by producing the ultimate goal of extracting or obtaining new information and insights from the data. Exploratory Data Analysis has a very important role to analyze, expose hidden information, identify anomalies in datasets [5]–[8].

Forecasting human mobility is an interesting discussion because mobility fluctuates rapidly [9], causing the transmission of the COVID-19 virus to be very fast. But on the other hand, with mobility, humans are able to maintain and improve their survival. One of the artificial neural network models that has the ability to predict or predict patterns of data types is Long-Short Term Memory (LSTM). LSTM is a development of the Recurrent Neural Network (RNN) model [10], [11]. LSTM has the advantage of RNN because it is able to overcome the vanishing gradient problem [12]. Research [13] focuses on the application of the LSTM model to forecast sales. The results of this study noted that the LSTM model proved to be better than the ARIMA, KNN, SVM and RNN models with the model evaluation using the Root Mean Square Error (RMSE) of 2595.96. Research [14] compared the accuracy of the LSTM and Gated Recurrent Unit (GRU) models to predict truck traffic flows. This study notes that both LSTM and GRU in general have very good performance in predicting truck traffic. But the GRU model in particular has higher accuracy than the LSTM model in the peak period. On the other hand, the LSTM model is able to predict better with a Mean Absolute Percentage Error (MAPE) value of 4.1% which is smaller than the GRU model. Based on the explanation above, the differences in research are data that fluctuates and adapts to the mobility conditions of the people of Denpasar City and the stages of data mining using the CRISP-DM method. Researchers conducted this study to analyze and predict the mobility of the people of Denpasar City using the Long Short-Term Memory (LSTM) method.

2. Method

2.1. Material Research

2.1.1. Long Short-Term Memory (LSTM)

LSTM was first introduced to the public in 1997 by Hochreiter and Schmidhuber. LSTM is a development of the Recurrent Neural Network (RNN) architecture because RNN cannot process long-term sequential data, causing missing gradients when using the backpopagation algorithm [15]–[17] as can see in Fig. 2.

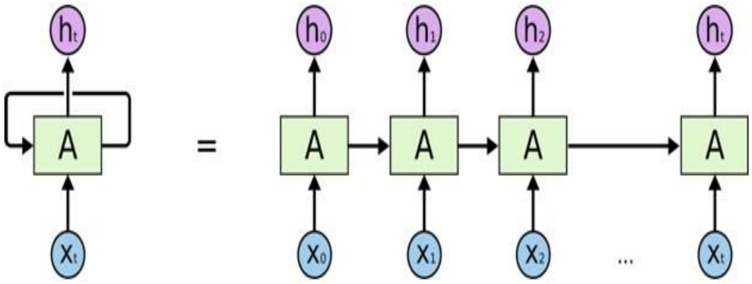

Fig. 2. RNN Architecture

LSTM has two functions, sigmoid and tanh activation functions. Not only that, LSTM has memory cells and gates which are divided into three namely forget gates, input gates, and output gates. Fig. 3 is an illustration of the LSTM method or architecture.

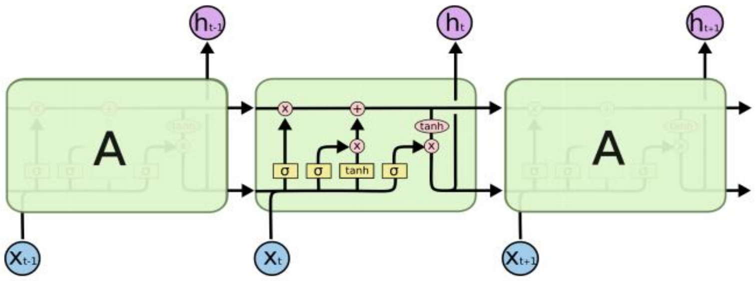

Fig. 3. LSTM architecture

Fig. 4 illustrates that one LSTM unit has 4 gates, namely forget gates, cell state, input gates, and output gates. The division of the 4 gates represents several gates tasks, namely Forget Gates is able to choose information that will or will not be used in the next process. Input gates represent new information used in the cell state. Memory block or called cell state has a function to store information in the t-1 period process and will be used in the t-period process [18]. Output gates will process the value generated from the entire process of 1 LSTM unit. The gates in the LSTM are processed with the sigmoid activation function so as to produce a value with a range of [0,1]. A value close to 0 represents that the value is not used while a value close to 1 will be used in the next process. The LSTM equation is as follows [19].

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{3}$$

Where f_t, i_t, o_t are Forget Gates, Input Gates, and Output Gates. σ is the sigmoid activation function, W is the weight of each gate, h_{t-1} previous output in period $t-1$, x_t is the input value in the t period and b represents the value of the bias at each gate. The equation for the new candidate cell state, cell state, and final output value is as follows [19]:

$$\bar{C}_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t \tag{5}$$

$$h_t = o_t * tanh\,(C_t) \tag{6}$$

Where C_{t-1} dan C_t represents the cell state value in the $t-1$ period, the cell state in the $t-th$ period or a memory block [19].

2.1.2. Normalization and Denormalization

Normalization is the process of changing the actual data into a range [0,1]. The purpose of normalization is to reduce errors and avoid data anomalies. One of the normalization techniques commonly used is min-max scaling. The min-max scaling technique aims to overcome the large difference in the actual data values. This technique will change the actual data value to a value with range [0,1] without changing the information available in the dataset. The equation or formula for this technique is as follow [20].

$$X' = \frac{(x - X_{min})}{(X_{max} - X_{min})} \tag{7}$$

Data denormalization is the process of changing or returning normalized data into actual data. The purpose of data denormalization is to facilitate understanding of the data so that when comparing it with actual data it can be done easily. Denormalization of data with range [0,1] can be expressed in the following equation or formula [21].

$$X_t = x(X_{max} - X_{min}) + X_{min} \tag{8}$$

2.1.3. Root Mean Square Error

Is a technique to measure error in a method when predicting or forecasting quantitative data. RMSE is able to overcome the unwanted use of absolute values in calculations. RMSE can be briefly explained as the difference between the total actual data minus the predicted data then squared and divided by the number of forecasting times. After that the value obtained will be rooted. The RMSE equation or formula is as follows [22].

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \bar{Y}_t)^2}{n}} \tag{9}$$

Some of the advantages of RMSE include: Sensitive to data ranges, so RMSE results are more accurate in assessing prediction errors. It is better to use it to compare errors in several methods used.

2.2. Method Research

Based on the Cross Industry Process for Data Mining (CRISP-DM) approach created by Daimler Chrysler in 1996, it has 6 stages of the process, namely (i) Business Understanding, (ii) Data Understanding, (iii) Data Preparation, (iv) Modeling, (v) Evaluation, and (vi) Deployment. The CRISP-DM method has been used in several studies [23]–[26]. The methodology implemented in this study can be seen in Fig. 4 [24].

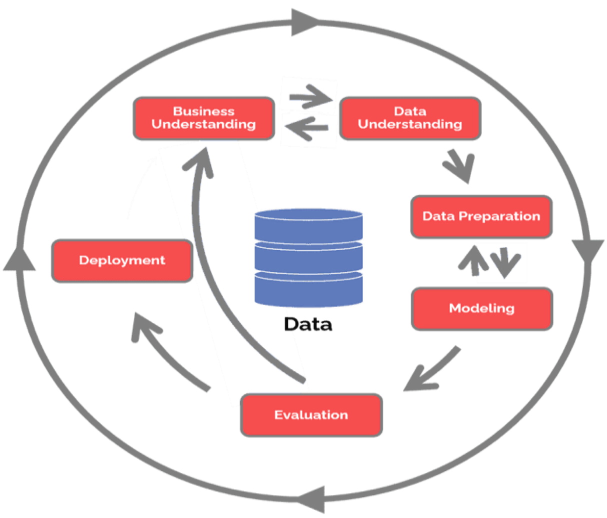

Fig. 4. CRISP-DM

2.2.1. Business Understanding

The purpose of this study is to analyze and predict the mobility of the people of Denpasar City for the next 7 days (1 November – 7 November 2021). The limitation in this study is that there are no sources outside the data source as attributes that can affect the data mining process such as rush hour.

2.2.2. Data Understanding

Data was obtained from the collaboration between ITB Stikom Bali and PT. Bamboo Media in Ideathon Bali Kembali Research with the title "Sistem Tracing dan Tracking Mobilitas Publik Berbasis QRCode dan Big Data, Untuk Mendukung Pembukaan Pariwisata Bali". The data collected is checkin area data for the period September 1, 2021 to October 31, 2021. The data that has been collected will be preprocessed and visualized so that the information you have can be conveyed.

2.2.3. Data Preparation

At this stage, the first thing to do is normalize the data to avoid data anomalies. Next, divide the data into 2, train data and test data with a proportion of 90%: 10%. The final step is to reconstruct the data so as to produce sub-sequential data using sliding windows (time steps).

2.2.4. Modelling

This stage is divided into 4 processes, namely: first, initiating early stopping to avoid overfitting. Next, initiate max epochs and batch sizes. After this it initiates a hyperparamter consisting of LSTM_units which represents the number of LSTM units, Dropouts, and optimizers to be combined into several different methods. Finally, train the training data with a combination of hyperparameters

2.2.5. Evaluation

Consists of several stages, namely, testing the results of the training method on the data testing. Denormalize data to return the actual range of data. Evaluation of the prediction results of testing data with Root Mean Square Error (RMSE). Choose the best method with the smallest RMSE value (close to 0) to be used in the next stage.

2.2.6. Deployment

At this stage the process carried out is to forecast the mobility of the people of Denpasar City for the next 7 days outside the dataset owned.

3. Results and Discussion

Forecasting the mobility of the people of Denpasar City with the LSTM method is one of the uses of information technology that is able to predict the amount of mobility of the people of Denpasar City in the specified period and can provide information to assist the government in determining policies that involve community or organizational mobility and event activities in regulating the number of in and out which takes into account the scale of community mobility.

3.1. Business Understanding

The purpose of this study is to analyze and predict the mobility of the people of Denpasar City for the next 7 days. The limitation in this study is that there are no sources outside the data source as attributes that can affect the data mining process such as rush hour. The programming language used is Python using the LSTM method which is available in the KERAT framework.

3.2. Data Understanding

At this stage the researcher collects research data through the web admin and then it is recorded and stored in a file with the extension Comma Separated Value (CSV). The period of recorded data is from September 1 to October 31, 2021. Several libraries used in this study are Pandas, NumPy, Scit-Learn, and Tensorflow. The data that has been acquired has a total of 147 rows and 63 columns. Data is preprocessed by changing data types, melt data, delete empty data. After preprocessing, the highest mobility case is obtained as can see in Fig. 5.

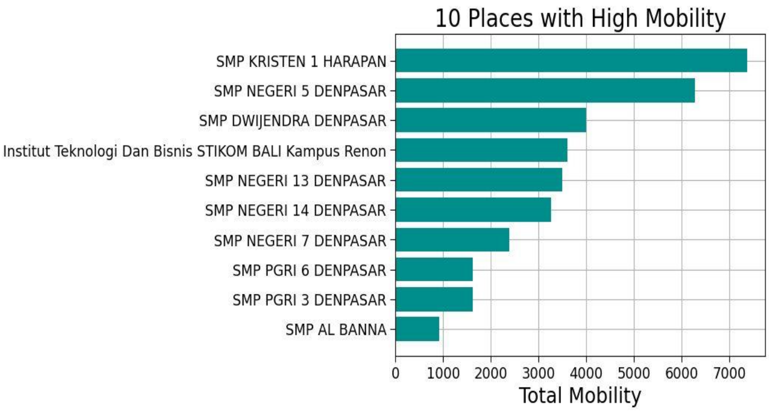

Fig. 5. Top 10 community mobility

The information obtained is that the level of mobility or scan mostly occurs in junior high schools. This is because Denpasar City in October 2021 has held limited face-to-face meetings (PTM). Information that can be sought again is how the mobility pattern of Denpasar City during that period was as can see in Fig. 6.

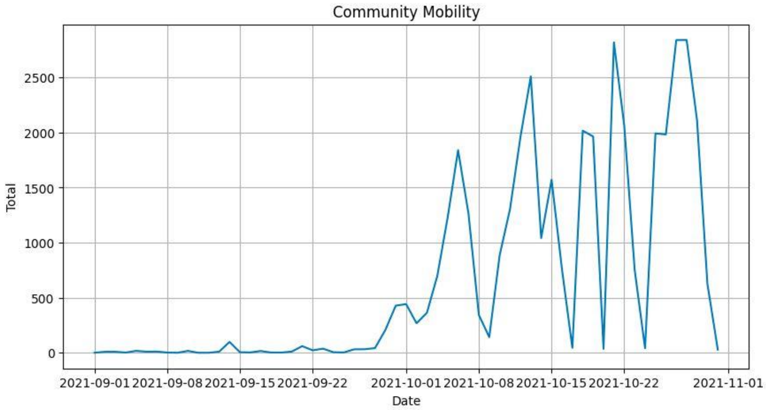

Fig. 6. Denpasar City Community Mobility Pattern

The results of the visualization above illustrate that the mobility of the people of Denpasar City in September tends to be lower than in October. This is in accordance with the conditions in September, that the city of Denpasar had not held a limited face-to-face meeting (PTM) so that the month of October had a fairly high mobility due to the implementation of the PTM [27]. In October, community mobility tends to increase and experiences a fairly high pattern of ups and downs.

3.3. Data Preparation

3.3.1. Data Normalization

The data on the mobility of the people of Denpasar City that has been collected has a large range as can see in Fig. 7.

count	61.000000
mean	653.721311
std	895.169521
min	0.000000
25%	10.000000
50%	60.000000
75%	1231.000000
max	2842.000000

Fig. 7. Descriptive Statistics of Community Mobility in Denpasar

The average daily data reaches 653 checkins with a fairly large standard deviation of 895. The standard deviation value is influenced by the data range from a minimum value of 0 to a maximum value of 2842. To avoid data anomalies in the LSTM method training process, the data is normalized with Equation 7 and produces the following values as can see in Table. 1.

Table.1 Data Normalization Results

Date	Amount	Scaled
2021-09-01	0	0
2021-09-02	9	0,003166783955
2021-09-03	9	0,003166783955
2021-09-04	2	0,0007037297678
2021-09-05	17	0,005981703026

3.3.2. Split Data

At this stage, the dataset is divided into 2 for training and testing needs. The normalized dataset is divided into 2 with a data ratio of 90:10. Where 90% is for training data and 10% is for testing data. Proportion of data train and data test as can see in Table 2.

Table.2 Proportion of Data Train and Data Test

Unit	Data Train	Data Test
Percent	90%	10%
Data Amount	54	7

3.3.3. Data Recontruction

Community mobility data collected is sequential data. The training process using the LSTM method requires more than 1 sequential data, for this research reconstruct sequential data into sub-sequential data. This sub-sequential data if illustrated as can see in Fig. 8.

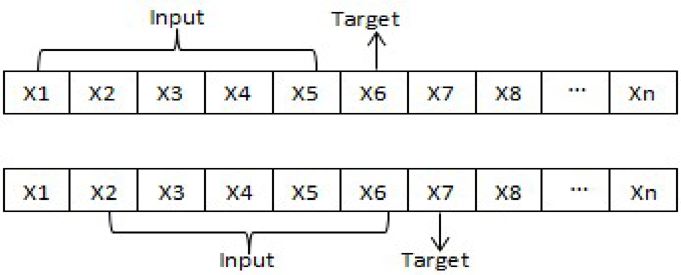

Fig. 8. Data reconstruction ilustration

The illustration above produces input data from X1 to X5 with a target value of X6. This process is also known as sliding windows, which is sliding sequential data to produce sub-sequential data [28]. Community mobility data if the sliding windows process is carried out will produce sub-sequential data as can see in Table. 3.

Table.3 Data Recontruction Results

X_1	X_2	X_3	X_4	X_5	$Target$
0	0.0031667	0.0031667	0.0007037	0.0059817	0.0035186
0.0031667	0.0031667	0.0007037	0.0059817	0.0035186	0.0038705
...	...	...	...	...	...
0.7100633	0.6914144	0.0123152	0.9926108	0.7241379	0.2.642505
0.6914144	0.0123152	0.9926108	0.7241379	0.2.642505	0.0144264

The results of the data reconstruction resulted in the amount of testing data of 49 sub-sequential data and training data of 7 sub-sequential data. The LSTM method has 3 input dimensions, namely the first dimension is the amount of data (batch size), the 2nd dimension is the sliding window or time steps and the 3rd dimension is the output dimension. In the regression problem for mobility forecasting, the output only has 1 dimension, namely how many predictions for mobility in the period t+1.

3.4. Modelling

3.4.1. Initiate Early Stopping

EarlyStop aims to avoid overfitting and underfitting the trained method. Overfitting is a phenomenon that the training error of a method is very small but its generalization ability to other data is weak, while underfitting is a condition where the data training method created does not represent the entire data that will be used later. This results in poor performance in data training. One way to avoid Overfitting is to do Early Stopping. The Early Stopping process is carried out if the minimum loss value does not decrease for 20 epochs.

3.4.2. Initiate Hyperparamater

In this study, the max_epochs value is set at 500. In the regression problem, there are several loss functions, namely L1 Loss, L2 Loss, Smooth L1 Loss. The loss function used in this research is L2 loss or commonly called Mean Square Error (MSE). The mean squared error is calculated as the average of the squared differences between the predicted and actual values. The result is always positive regardless of the sign of the predicted and actual values and the perfect score is 0.0. Square means that the larger error produces more errors than the smaller error, which means that the method will be corrected for making the larger error [29]. The number of batch_size is 1 and RMSE metrics to assess the performance of the trained method. The next stage is the initiation of hyperparameter. In this study there are 3 hyperparameters with the following test scenarios: LSTM_units: 32, 64, 128, 512; Dropout: 0.1, 0.2, 0.3; Optimizer: SGD, RMSProp, Adam. At the initiation of hyperparameters there are no definite rules, but it is done by testing the data and the methods to be used. The combination of the above hyperparameters produces a total of 45 methods to be trained. The resulting total combination is as can see in Table. 4

Table.4 Data Recontruction Results

LSTM_units	Dropout	Optimizer	Method Name
32	0.1	SGD	Model1
32	0.1	RMSProp	Model2
32	0.1	Adam	Model3
32	0.2	SGD	Model4
32	0.2	RMSProp	Model5
32	0.2	Adam	Model6
32	0.3	SGD	Model7
32	0.3	RMSProp	Model8
32	0.3	Adam	Model9
64	0.1	SGD	Model10
64	0.1	RMSProp	Model11
64	0.1	Adam	Model12
64	0.2	SGD	Model13
64	0.2	RMSProp	Model14
64	0.2	Adam	Model15
64	0.3	SGD	Model16
64	0.3	RMSProp	Model17

LSTM_units	Dropout	Optimizer	Method Name
64	0.3	Adam	Model18
128	0.1	SGD	Model19
128	0.1	RMSProp	Model20
128	0.1	Adam	Model21
128	0.2	SGD	Model22
128	0.2	RMSProp	Model23
128	0.2	Adam	Model24
128	0.3	SGD	Model25
128	0.3	RMSProp	Model26
128	0.3	Adam	Model27
256	0.1	SGD	Model28
256	0.1	RMSProp	Model29
256	0.1	Adam	Model30
256	0.2	SGD	Model31
256	0.2	RMSProp	Model32
256	0.2	Adam	Model33
256	0.3	SGD	Model34
256	0.3	RMSProp	Model35
256	0.3	Adam	Model36
512	0.1	SGD	Model37
512	0.1	RMSProp	Model38
512	0.1	Adam	Model39
512	0.2	SGD	Model40
512	0.2	RMSProp	Model41
512	0.2	Adam	Model42
512	0.3	SGD	Model43
512	0.3	RMSProp	Model44
512	0.3	Adam	Model45

3.4.3. LSTM Method Training

At this stage, the method resulting from the combination of hyperparameters is trained with the LSTM architecture which is built to have 1 input layer consisting of 5 input units, 3 LSTM hidden layers consisting of a number of LSTM_units and 1 output layer. The LSTM parameters that have been built can be calculated based on n LSTM_units of m input dimensions, $(m+n+1)*4*n$, which means that each matrix weight x_t sized $m*n$ and weight matrix weight h_t sized $n*n$ and size bias $1*n$ [30].

3.4.4. Evaluation

At this stage, all methods that have been trained will be tested with data testing before being used in the deployment process. This test aims to test the ability of the method on data that has never been used during the training process. The first thing to do in the above process is to make predictions using the i-th method with data testing. Keep in mind that the X_test data still has a range of [0,1], so it needs to be denormalized to get the actual range of values. The RMSE value of each test data generated will be compared. The RMSE value is said to be the best if it has a value close to 0 so that the lowest RMSE value is generated by model21 of 971.438687, followed by model17 of 990.7274516 and model33 of 1014.299474. If you visualize the value of the testing data generated by the 3 methods as can see in Fig. 9.

Vol. 8, No. 2, December 2024, pp. 213-225

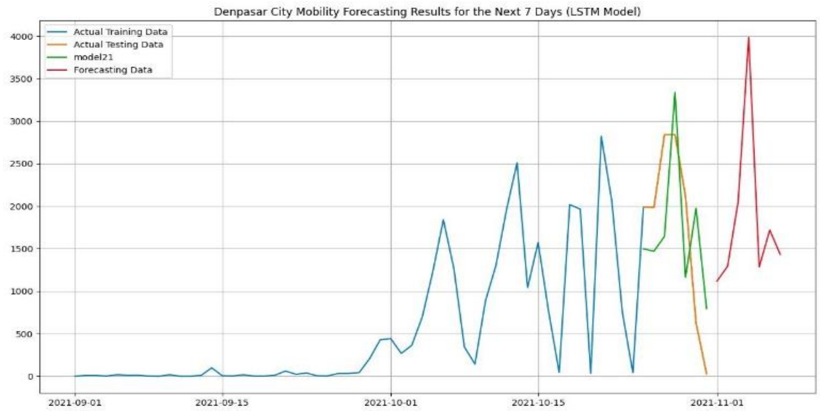

Fig. 9. Visualization Data Test

The visualization above illustrates that model21 has a pattern that is close to the actual testing data. When compared with model 17 and model 33 which have the 2nd and 3rd smallest RMSE values, the resulting pattern is still much different from the actual testing data. So from the RMSE value and the resulting pattern, model21 with 128 LSTM_units, 0.1 dropout, and the Adam optimizer can be used as the best method for the deployment stage. Model21 has a number of LSTM_units of 128. First hidden layer has 1 input dimension so that the first total parameter is (1+128+1)*4*128 ie 66,560 parameters. In the second hidden layer the number of input dimensions is 128 which is obtained from the output dimensions of the first hidden layer, so the total parameters of the second hidden layer are (128+128+1)*4*128 i.e. 131,584 total parameters. The third hidden layer has the same total parameters as the second hidden layer. In the output layers, there is 1 Dense layers unit with a total parameter, namely the total input dimension plus 1 for bias, so that the total parameters for the output layer are 128+1, which is 129. So the total parameters trained in this LSTM method amount to 66.560+131.584+131.584+129 i.e. 329.875. Model21 architecture as can see in Fig. 10.

```
Model: "sequential_20"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_60 (LSTM)              (None, 5, 128)            66560

 dropout_60 (Dropout)        (None, 5, 128)            0

 lstm_61 (LSTM)              (None, 5, 128)            131584

 dropout_61 (Dropout)        (None, 5, 128)            0

 lstm_62 (LSTM)              (None, 128)               131584

 dropout_62 (Dropout)        (None, 128)               0

 dense_20 (Dense)            (None, 1)                 129

=================================================================
Total params: 329,857
Trainable params: 329,857
Non-trainable params: 0
_________________________________________________________________
```

Fig. 10. Model21 architecture

3.4.5. Deployment

The last process is forecasting for the next 7 days outside of the available datasets. The process carried out is using the LSTM method and the last 5 days of data testing for forecasting. The results of forecasting the mobility of the people of Denpasar City as cas see in Table 5 and Fig. 11.

Table.5 Denpasar City Community Mobility Forecasting Results

Date	Prediction
2021 – 11 – 1	715,793214
2021 – 11 – 2	1170,089534
2021 – 11 – 3	1445,007874
2021 – 11 – 4	2849,400251
2021 – 11 – 5	1732,756458
2021 – 11 – 6	558,7630896
2021 – 11 – 7	1081,554515

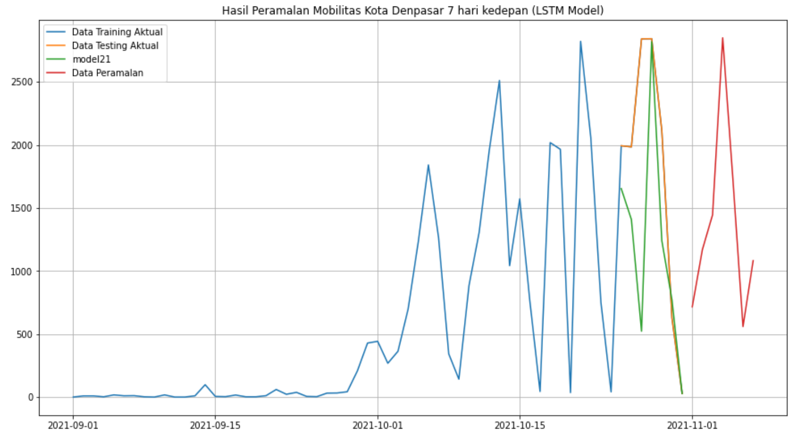

Fig. 11. Denpasar City Community Mobility Forecasting Data Visualization

Forecasting results show that the mobility of the people of Denpasar City will increase with a peak on November 4, 2021 of 2849 as PTM is implemented. There have been several times that community mobility experienced a decline, the lowest was touching 558 on November 6, 2021, which could be due to several factors in the field, such as weekends.

3.4.6. Discussion

The forecasting carried out resulted in a fluctuating level of mobility by reaching the highest point on November 4, 2021 at 2849 and the lowest point on November 6, 2021 at 558. The accumulation of community mobility in Denpasar City which was predicted in the period from November 1, 2021 to November 7, 2021, reached 9,550 in total check-ins. Visualization of forecasting data when compared with actual in the period November 1, 2021 to November 7, 2021, showed in Table 6 and Fig 12.

Table.6 Comparison Of Forecasting Results With Actual Data

Date	Prediction	Actual
2021 – 11 – 1	716	1752
2021 – 11 – 2	1170	1858
2021 – 11 – 3	1445	2603
2021 – 11 – 4	2849	2116
2021 – 11 – 5	1733	1783
2021 – 11 – 6	559	345
2021 – 11 – 7	1082	29

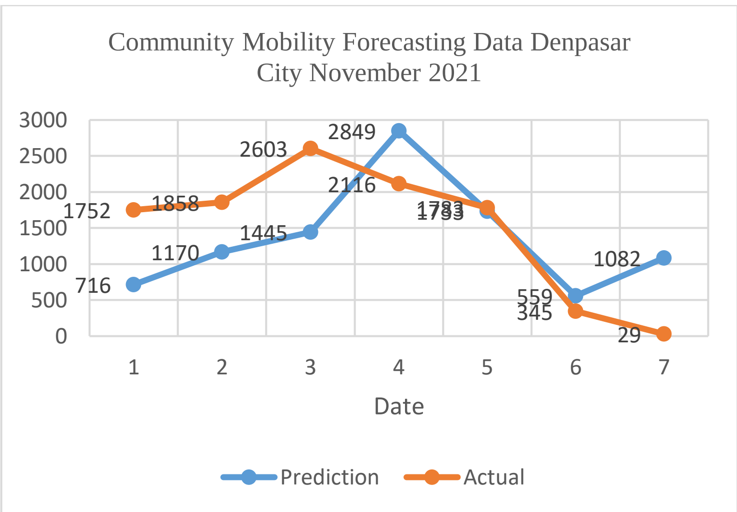

Fig. 12. Comparison Of Forecasting Results With Actual Data

4. Conclusion

Based on the results of the analysis, it can be concluded that the Denpasar City community mobility data that has been collected during the period September 1, 2021 to October 31, 2021, has a random pattern with a tendency at the end of the period to have a seasonal trend. The mobility of the Denpasar City community tends to be higher in the junior high school education sector. This relates to the return of limited face-to-face meetings. The best LSTM method for forecasting the mobility data of the Denpasar City community is a combination of hyperparameters with 128 LSTM_units, 0.1 Dropout, and using the Adam optimizer. The LSTM method that has been produced has an evaluation value of Root Mean Square Error of 971,438687.

References

[1] WHO, "Coronavirus disease 2019 (COVID-19)," World Health Organization, 2020. [Online]. Available at: https://www.who.int/health-topics/coronavirus#tab=tab_1.

[2] M. U. G. Kraemer *et al.*, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science (80-.).*, vol. 368, no. 6490, pp. 493–497, May 2020, doi: 10.1126/science.abb4218.

[3] WHO, "WHO, 'SARS-CoV-2 transmission: implications for infection prevention precautions,'" 2020. [Online]. Available at: https://cdn.who.int/media/docs/default-source/searo/indonesia/covid19.

[4] KEMENKES, "COVID-19 Situation by Regency/City," *Kementerian Kesehatan Republik Indonesia*, 2021.

[5] S. F. Pane and E. V. Rahcmadani, *Big Data : Forecasting Menggunakan Python*, 1st ed. Bandung: Kreatif Industri Nusantara, 2020.

[6] A. Bezerra, I. Silva, L. A. Guedes, D. Silva, G. Leitão, and K. Saito, "Extracting value from industrial alarms and events: A data-driven approach based on exploratory data analysis," *Sensors*, vol. 19, no. 12, p. 2772, 2019, doi: 10.3390/s19122772.

[7] B. M. Konopka, F. Lwow, M. Owczarz, and Ł. Łaczmański, "Exploratory data analysis of a clinical study group: Development of a procedure for exploring multidimensional data," *PLoS One*, vol. 13, no. 8, p. e0201950, 2018, doi: 10.1371/journal.pone.0201950.

[8] C. Deming, S. Dekkati, and H. Desamsetti, "Exploratory Data Analysis and Visualization for Business Analytics," *Asian J. Appl. Sci. Eng.*, vol. 7, no. 1, pp. 93–100, 2018, doi: 10.18034/ajase.v7i1.53.

[9] K. Chen, X. Jiang, Y. Li, and R. Zhou, "A stochastic agent-based model to evaluate COVID-19 transmission influenced by human mobility," *Nonlinear Dyn.*, vol. 111, no. 13, pp. 12639–12655, Jul. 2023, doi: 10.1007/s11071-023-08489-5.

[10] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Deep learning with long short-term memory for time series prediction," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 114–119, 2019, doi: 10.1109/MCOM.2019.1800155.

[11] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.

[12] J. Gonzalez and W. Yu, "Non-linear system modeling using LSTM neural networks," vol. 51, no. 13, pp. 485–489, 2018, doi: 10.1016/j.ifacol.2018.07.326.

[13] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, no. July 2019, p. 106435, 2020, doi: 10.1016/j.cie.2020.106435.

[14] S. Wang, J. Zhao, C. Shao, C. D. Dong, and C. Yin, "Truck Traffic Flow Prediction Based on LSTM and GRU Methods with Sampled GPS Data," *IEEE Access*, vol. 8, pp. 208158–208169, 2020, doi: 10.1109/ACCESS.2020.3038788.

[15] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining Practical Machine Learning Tools and Techniques*, 4th ed. Cambridge: Todd Green, 2017. [Online]. Available at: https://openlibrary.telkomuniversity.ac.id/pustaka/165648/data-mining-practical-machine-learning-tools-and-techniques.html.

[16] M. Alazab, S. Khan, S. S. R. Krishnan, Q.-V. Pham, M. P. K. Reddy, and T. R. Gadekallu, "A multidirectional LSTM model for predicting the stability of a smart grid," *IEEE Access*, vol. 8, pp. 85454–85463, 2020, doi: 10.1109/ACCESS.2020.2991067.

[17] K. Gupta, N. Jiwani, and N. Afreen, "Blood pressure detection using CNN-LSTM model," in *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, 2022, pp. 262–366, doi: 10.1109/CSNT54456.2022.9787648.

[18] L. Irawan, F. Fauzi, and D. Andwiyan, "Prediction Of Incoming Orders Using The Long Short-Term Memory Method At Pt. Xyz," *JISA(Jurnal Inform. dan Sains)*, vol. 4, no. 1, pp. 80–89, Jun. 2021, doi: 10.31326/jisa.v4i1.902.

[19] J. Luo, Z. Zhang, Y. Fu, and F. Rao, "Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms," *Results Phys.*, vol. 27, p. 104462, 2021, doi: 10.1016/j.rinp.2021.104462.

[20] A. Moghar and M. Hamiche, "Stock Market Prediction Using LSTM Recurrent Neural Network," *Procedia Comput. Sci.*, vol. 170, pp. 1168–1173, 2020, doi: 10.1016/j.procs.2020.03.049.

[21] R. Saleem and S. Shaukat, "Denormalization To Enhance Effciency In Data Mining," *Int. J. Sci. Eng. Res.*, vol. 7, no. 9, pp. 1409–1413, 2016, [Online]. Available at: https://www.ijser.org/researchpaper/DENORMALIZATION-TO-ENHANCE-EFFCIENCY-IN-DATA-MINING.pdf.

[22] W. Wang and Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 324, p. 012049, Mar. 2018, doi: 10.1088/1757-899X/324/1/012049.

[23] V. Plotnikova, M. Dumas, and F. P. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," *Data Knowl. Eng.*, vol. 139, p. 102013, 2022, doi: 10.1016/j.datak.2022.102013.

[24] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[25] S. Al Gharbi, A. Al-Majed, A. Abdulraheem, S. Patil, and S. Elkatatny, "Using data-mining CRISP-DM methodology to predict drilling troubles in real-time," p. 10, 2021, doi: 10.2118/201026-MS.

[26] F. Martínez-Plumed *et al.*, "CRISP-DM twenty years later: From data mining processes to data science trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, 2019, doi: 10.1109/TKDE.2019.2962680.

[27] S. F. Nissa and A. Haryanto, "Implementation of Face-to-Face Learning During the Covid-19 Pandemic," *J. IKA PGSD (Ikatan Alumni PGSD) UNARS*, vol. 8, no. 2, p. 402, 2020, doi: 10.36841/pgsdunars.v8i2.840.

[28] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Procedia*, vol. 158, pp. 2922–2927, 2019, doi: 10.1016/j.egypro.2019.01.952.

[29] R. Y. Hayuningtyas, "Inventory forecasting using the weighted moving average method and the double exponential smoothing method," *J. PILAR Nusa Mandiri*, vol. 13, no. 2, pp. 217–222, 2017, [Online]. Available at: https://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/240.

[30] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/neco_a_01199.