8 (2) July 2022 142-155 ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
Register: *Jurnal Ilmiah Teknologi Sistem Informasi*
Contents lists available at www.journal.unipdu.ac.id
Register
Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

Predicting the Number of Passengers of MRT Jakarta Based on the Use of the QR-Code Payment Method during the Covid-19 Pandemic Using Long Short-Term Memory

Riyanto Jayadi [a,*], *Taskia Fira Indriasari* [b], *Charis Chrisna* [c], *Putri Natasya Fanuel* [d], *Rayhana Afita* [e]

[a,b,d,e] *Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia*
[c] *Department of Ocean Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
email: [a,*] riyanto.jayadi@binus.edu, [b] taskia.indriasari@binus.ac.id, [c] karisma.17030214035@mhs.unesa.ac.id, [d] atikwintarti@unesa.ac.id
* *Correspondence*

ARTICLE INFO

Article history:
Received 1 September 2021
Revised 7 December 2022
Accepted 13 December 2022
Available online 26 December 2022

Keywords:
prediction
machine learning
LSTM
CNN
QR code

Please cite this article in IEEE style as:
R. Jayadi, T. F. Indriasari, C. Chrisna, P. N. Fanuel and R. Afita, " Predicting the Number of Passengers of MRT Jakarta Based on the Use of the QR-Code Payment Method during the Covid-19 Pandemic Using Long Short-Term Memory," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 8, no. 2, pp. 142-103, 2022.

ABSTRACT

The trend of using public transportation has been rising over the last several decades. Because of increased mobility, public transportation has now become more crucial. In modern environments, public transportation is not only used to carry people and products from one location to another but has also evolved into a service company. In Jakarta, Mass Rapid Transit Jakarta (MRTJ) started to operate in late 2019. Recently, they updated their payment gateway system with QR codes. In this study, we predicted the hourly influx of passengers who used QR codes as their preferred payment method. This research applied machine learning to perform a prediction methodology, which is proposed to predict the number of passengers using time-series analysis. The dataset contained 7760 instances across different hours and days in June 2020 and was reshaped to display the total number of passengers each hour. Next, we incorporated time-series regression alongside LSTM frameworks with variations in architecture. One architecture, the 1D CNN-LSTM, yielded a promising prediction error of only one to two passengers for every hour.

1. Introduction

MRT Jakarta has been in operation since March 2019 and is owned by the Provincial Government of DKI Jakarta. The company was established on June 17th, 2008, as a limited company with shares held by the Provincial Government of DKI Jakarta. This company is responsible for a wide range of operations, including the operation and development of the MRT infrastructure and amenities and maintaining the surrounding area [1], [2].

As a critical component of a public transportation system, the metro system has received increasing academic and functional attention because of its characteristics of high capacity, high speed, and high dependability. In order to accommodate this rapid expansion, the urban rail system has undergone several challenges, including oversaturation in the train, station, and other underground infrastructure [3]. In 2019, MRT Jakarta was used by 19.9 million passengers, accounting for an average daily passenger load of approximately 83,516 passengers and a peak daily load of 93,165 [4].

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

The Quick Response Code (QR code) has recently been utilised by MRT Jakarta as a payment method. The introduction of new payment methods and the expansion of Mobile Payment can potentially dematerialise payment methods [5][6]. Compared to other widely used payment methods, this technique offers more convenience and advantages than those offered by other methods [7][8]. Especially, QR codes can be read by a smartphone device, which is currently very popular among dwellers in the capital city who can generally afford it.

After January 1st, 2020 the QR code standardisation known as the QR Code Indonesian Standard (QRIS) was launched by the central bank of the Republic of Indonesia or Bank Indonesia [9]. The use of QR codes as a payment method also plays an essential part in the economy. This assertion is supported by the fact that QR Code payment users in China have reached approximately 70 per cent of the population[10]. To maximise the usage of QR Codes, it is recommended that the payment method is used in the Jakarta Transportation Sector [11].

Improving the use of QR Codes can benefit the economy of Indonesia. Indonesia's economy has tremendous potential to profit from the digitisation of the world economy. Being the fourth most densely-populated country in the world and a demographic dominated by generations Y and Z, Indonesia has emerged as one of the most promising consumer markets to benefit from the wave of digitisation currently underway. With these opportunities, it is only natural that online platform companies thrive in Indonesia, especially those working in the financial technology and e-commerce sectors [11].

MRT Jakarta has collaborated with several payment providers to provide QR-Code payment services. For example, OVO, GoPay, LinkAja, and Dana are the companies involved. Customers who purchase their tickets via MRT Jakarta Mobile Application (MRT-J) will have the QR Code automatically generated, which will be scanned at the entrance gate. The gate machine records the date and time of the transaction, the tapping time, and the User ID. Obtaining this data enables us to predict the number of people using MRT Jakarta to enhance service quality. This upgrade is part of MRT Jakarta's attempts to implement a marketing strategy to increase passengers by conducting a promotional campaign and perhaps optimising operations and personnel planning.

However, the daily average number of passengers who use the QR code as a payment method media only reached 267 people per day. Even in the present epidemic period, as we are minimising interaction with other people, the usage of QR codes in MRT Jakarta is still relatively low compared to other payment methods, such as debit cards, single-trip cards, and multi-trip cards. Therefore, this study aims to solve this problem by estimating the number of passengers in MRT Jakarta.

A number of relevant research projects have investigated the application of predictive systems in public transportation using machine learning [12]–[20]. Using machine learning appropriately may allow businesses to gather data to gain competitive advantages [12]. Research in [13] mainly deals with predicting the demand for taxis in specific areas using machine learning algorithms based on climate, location, time of the day, and other parameters affecting customer waiting time imbalance. The authors of [15] proposed a pattern to forecast the total number of passengers who seek to travel from one point to another. Meanwhile, in [16], the researchers sought to estimate the total number of passengers entering each station and boarding at each stop. The study in [17] made use of the retrieved serial set to construct a BRT passenger flow forecast model. To train the model parameters, it used a greedy layer-wise method. Experimental results demonstrate that this approach performs well when it comes to forecasting short-term BRT passenger flow. Another research focusing on predicting user behaviour was conducted in [18]. It revolves around using metro card data from Shenzhen, China, to identify individual mobility patterns and forecast future movements of passengers. The research in [19] studies the route choice behaviour of passengers from Auto Fare Collection, timetable, and train loading data using Bayesian inference and the Metropolis-Hasting sampling method. Another study in [20] offers a technique for estimating, using data from smart cards, and how much public transportation users'

 http://doi.org/10.26594/register.v8i2.2546

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

behaviour has changed over time in Canada's transportation system. However, there are not many studies on how to predict the number of passengers of an MRT during the pandemic time.

Thus, in this paper, we proposed a machine learning approach to train a predictive model for predicting passengers who use the QR code within a specific time in MRT. 1D CNN-LSTM -based time series prediction is proposed to conduct the prediction. Having the prediction can improve the efficiency of marketing campaigns and bolster decision-making to do promotions on particular hours and/or days. In addition, it can be used as a forecasting method of railroad demands, machines used to facilitate operation, and human resources planning. The attribute in the dataset in this study is the number of users, Time Stamp, provider, tap-in time, tap-out time, and the number of transactions. Using a machine learning algorithm on time-series analysis, this research proposes an approach to predict the number of passengers that use QR Codes as their preferred payment method when using MRT Jakarta.

The rest of this paper is organised as follows. Section II reviews and discusses the literature related to the topic of the current research. Section III presents the methods applied in this study. Section IV consists of the results of this research, and Section V presents the conclusion of this study.

2. Materials and Methods

2.1. Previous Studies

A number of research projects have investigated the application of payment systems in public transportation using machine learning. For instance, the study conducted by [13] mainly deals with predicting the demand for taxis in specific areas using machine learning algorithms based on climate, location, time of the day, and other parameters affecting customer waiting time imbalance. The study allows users to use the prediction model on a web-based platform. The users need to specify the input of various parameters required by the model, such as climate, time of the day, location, and the model attempting to predict the number of taxis needed to be distributed to that specific location based on past data.

Meanwhile, the paper by [15] proposes a pattern to forecast the total number of passengers who travel from one location to another. To meet the research goal, the researchers generated the company's vehicle event files and obtained complementary data to be used as information sources. The data were then classified into their respective use cases: temporal (time of the trip, type of the day, month and season), geographic, as well as demographic (departure and destination bus stop, type of bus stop and zip codes of the origin and destination bus stop).

In addition, in the study conducted by [16], the researchers sought to estimate the total number of passengers entering each station and boarding at each stop. Using machine learning algorithms, forecasting models for both the long and short term were created. In addition to calendar models, the algorithms implemented in this research are Random Forests (RF), Long-Short Term Memory (LSTM) neural networks, and other types of neural network architectures. Depending on the model (long-term or short-term), forecasting may be done up to a year in advance for train stations, tram stops, and bus stops. For the short-term, prediction can forecast within the next 15 minutes for bus stops and within the next 30 minutes for buses, which can benefit both passengers and operators. To conduct the tests, the Transport Organization Authority of Ile-de-France supplied a genuine 2-year smart card dataset used in the trials. For this study, the area of La Defense, a well-known crucial commercial centre in the Paris Metropolitan Area, and its 145 stations and stops are discussed in detail. Using the available data and machine learning models, the researchers were able to demonstrate the efficacy of their predicting methods.

Besides, the study by [17] utilised the retrieved serial set to construct a BRT passenger flow forecast model. It uses a greedy layer-wise method to train the model parameters. Experimental results demonstrate that this approach performs well when it comes to forecasting short-term BRT passenger flow.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

Another research focusing on predicting user behaviour was conducted by [18]. The study focuses on the usage of metro card data from Shenzhen, China, to identify individual mobility patterns and forecast future movements of passengers. These findings indicate that it is possible to anticipate travel patterns and enhance conventional pattern identification and prediction techniques for modelling urban mobility. The evidence also lends credence to an explanation of the structural characteristics of human behaviour in metropolitan transportation networks.

In the paper by [19], the researchers studied the route choice behaviour of passengers from Auto Fare Collection, timetable, and train loading data using Bayesian inference and Metropolis-Hasting sampling method. Then, an investigation of a case study of the Beijing subway was conducted to verify the validity of the model and algorithm.

Last but not least, a study by [20] offers a technique for estimating how much public transportation users' behaviour has changed over time in Canada's transportation system using data from smart cards. The study technique involves dividing users' smart card data into a series of periods separated by a period of time. The researcher measured the dissimilarity between each pair of time windows, which may be interpreted as an indication of the user's proclivity to adjust to changes in the transportation system, and then they averaged the results.

2.2. Methods

This study follows the CRISP-DM methodology. Section I presents the description of the business. This section describes the data, data preparation, and modeling of the CRIPS-DM methodology. Meanwhile, Section IV presents a discussion of the evaluation and implementation. This study incorporates the Jupyter Notebook environment via Google Colaboratory without GPU integration. Several software packages were utilised, including Python (3.7.11); Pandas (1.1.5) for the standard data exploration and cleaning module; TensorFlow (2.6.0) for the standard deep learning module; Matplotlib (3.2.2) as the standard visualisation module for data analysis; and NumPy (1.19) as the standard module for scientific computing.

2.2.1. Understanding the Data

This research used the data from MRT Jakarta Passengers' QR codes from June 1st, 2020, until June 22nd, 2020. This period was during the first wave Covid-19 pandemic in Indonesia. A total of 7760 ridership instances were used in the dataset, which consists of numerical and nominal data. The attributes in the dataset consist of Date, User Number, Payment Method, Origin Station, Destination Station, Tap-in Time, Tap-out Time, as shown in Table 1. Meanwhile, Table 2 shows the sample of the dataset.

The data with the Timestamp consists of time per hour during the opening hours of MRT Jakarta, which are "05.00", "06.00", "07.00", "08.00", "09.00", "10.00", "11.00", "12.00", "13.00", "14.00", "15.00", "16.00", "17.00", "18.00", "19.00", "20.00". The passenger data consist of the number of passengers for each hour which has a minimum value of 0 and a maximum value of 128, as shown in Figure 1.

The statistical parameters derived from Figure 1 yield a mean of 24.32 and a standard deviation of 24.8. Figure 1 also shows that the number of passengers using the QR code payment method is around 10 to 20 at most hours. However, at the tail end, there are multiple instances where the number of passengers reached 100 or more in an hour. Further analysis finds that the maximum passenger of passengers in a particular hour reached 128 passengers.

In addition, the duration travelled by each passenger was also analysed to better understand the average time taken from each station. Figure 2 shows that the average duration peaks around 30 minutes before eventually increasing to 40 to 60 minutes. The visual analysis also indicates that travel distances resemble a lognormal distribution, with the right tail mainly representing sparse outliers – travel periods of approximately 100 minutes. We consider those outliers interpretable and will still include them in further analysis and modelling.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

Table 1. Metadata of Dataset before Pre-Processed

Attribute	Type	Description
Date	Date time	Varchar of the Transaction date
User No.	Integer	The User Number generated by the system
Payment Method	Polynomial	Varchar of the Payment Method ("LinkAja", "OVO", "Go-Pay", "Dana")
Origin Station	Polynomial	Varchar of the Departure Station
Destination Station	Polynomial	Varchar of the Destination Station
Tap-in Time	Polynomial	Varchar of the time when the user taps the code while entering the station
Tap-out Time	Polynomial	Varchar of the time when the user taps the code while exiting the station

Table 2. Sample of Dataset before Pre-Processed

No.	Date	Origin Station	Destination Station	Payment Method	Tap-in at	Tap-out at
1	6/1/2020	Dukuh Atas BNI	Blok M BCA	OVO	6:07:08 am	6:24:26 am
2	6/1/2020	Blok M BCA	Bundaran HI	OVO	7:04:20 am	7:30:15 am
3	6/1/2020	Lebak Bulus Grab	Blok M BCA	OVO	7:55:29 am	8:13:07 am
4	6/1/2020	Cipete Raya	Lebak Bulus Grab	OVO	8:36:26 am	9:11:16 am
5	6/1/2020	Lebak Bulus Grab	Cipete Raya	OVO	9:59:20 am	10:36:52 am
...	...	...	...	...	...	...
7757	6/22/2020	Cipete Raya	Lebak Bulus Grab	GO PAY	8:32:09 am	8:48:28 am
7758	6/22/2020	Senayan	Lebak Bulus Grab	OVO	8:37:07 am	9:01:11 am
7759	6/22/2020	Bendungan Hilir	Fatmawati	GO PAY	8:37:13 am	8:58:58 am
7760	6/22/2020	Senayan	Cipete Raya	DANA	8:37:47 am	8:54:50 am

Finally, the hourly average number of passengers across all observed days was analysed. This was done to anticipate rush hours, better enhance promotion strategies, and reveal the root cause of a potential breakdown in the system during such hours. The line plot in Figure 3 shows that the distribution is bimodal, with each modal at the peak busy hours of the day (hours 6-8 and 16-19). The average daily passenger count for each aggregated hour is shown Table 3 as follows.

Table 3. Average Passengers of Each Working Hour

Hour of the Day	Average Passengers
5	3.1875
6	19.95455
7	37.27273
8	37.40909
9	20.68182
10	14.85714
11	11.14286
12	11.55
13	11.65
14	11.90476
15	20.8
16	44.22727
17	55.54546
18	46.11111
19	19.66667
20	9.75

147
ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

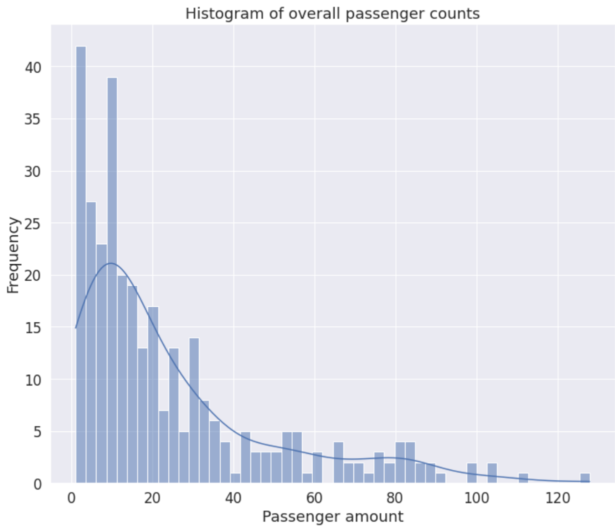

Fig 1. Histogram Visualisation of Passenger Counts per hour

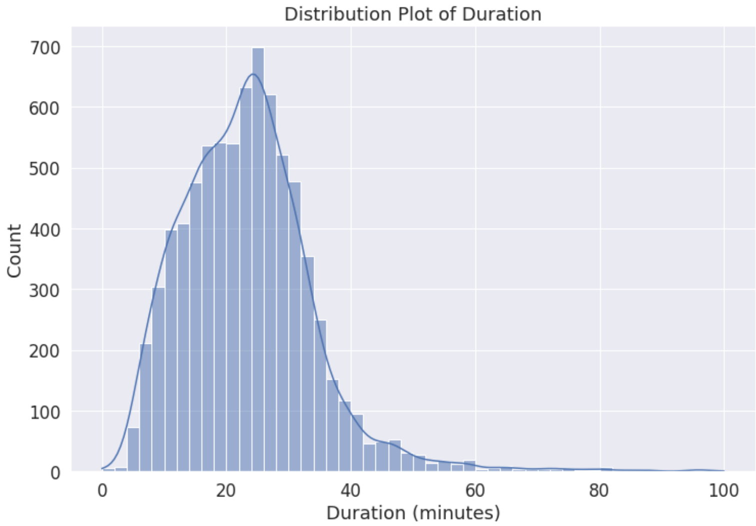

Fig 2. Distribution Visualisation of Travel Duration

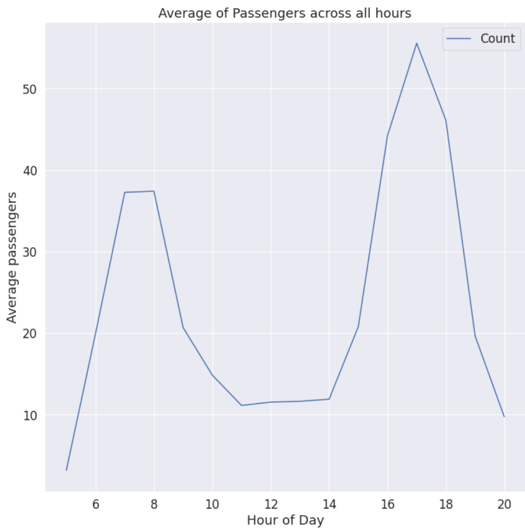

Fig. 3. Distribution Visualisation of Average Passengers across All Hours

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

2.2.2. Data Preparation

a) Aggregation

To conduct the data modelling for predicting hourly number of passenger according to the time series, we collected the data of 7760 individual ridership instances for each hour between 5:00 am to 20:00 pm (15 hours). The data collection resulted in 319 aggregated instance hours. Data transformation into an aggregated number of passengers per hour was done. The pre-processed dataset of 319 data points with additional variables are shown in Table 4 and Table 5.

Table 4. Metadata of Dataset after Aggregation

Attribute	Type	Description
Timestamp	Date time	Varchar of classification time in MRT Jakarta, grouped by hour of the day.
Aggregated Number of Passengers	Integer	The total number of passengers for each hour.

Table 5. Sample of Dataset after Aggregation

No.	Timestamp	Count
1	6/1/2020 6:00	1
2	6/1/2020 7:00	2
3	6/1/2020 8:00	1
4	6/1/2020 9:00	1
5	6/1/2020 10:00	2
6	6/1/2020 11:00	2
...	...	...
316	6/22/2020 17:00	128
317	6/22/2020 18:00	83
318	6/22/2020 19:00	33
319	6/22/2020 20:00	20

b) Normalisation

Normalisation was carried out to compress the input integers so that the deep learning model could process the data more efficiently. In this research, normalisation scales the feature set with a minimum of 0 and a maximum of 1. Table 6 shows the dataset after rescaling.

Table 6. The Sample of Dataset after Rescaling

Time	Passengers
6/1/2020 6:00	0
6/1/2020 7:00	0.007874
6/1/2020 8:00	0
6/1/2020 9:00	0
6/1/2020 10:00	0.007874

c) Lagging

This research also incorporates a time-lag of 5 hours as a dataset. Table 7 shows the final dataset sample, including lagging. In the table, t represents the number of passengers at the timestamp hour being predicted as the target or label variable. Meanwhile, t-1 until t-5 are the number of passengers at the timestamp hour, minus 1 hours until minus 5 hours, respectively, as the predicting variables. The first 5 rows and last 5 rows of the dataset were truncated due to time series lagging for the first 5 hours and the last 5 hours.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

Table 7. Sample of Dataset after Pre-Processed

Time	*t*	*t-1*	*t-2*	*t-3*	*t-4*	*t-5*
6/1/2020 5:00	0.0	?	?	?	?	?
6/1/2020 6:00	0.0078125	0.0	?	?	?	?
6/1/2020 7:00	0.015625	0.0078125	0.0	?	?	?
6/1/2020 8:00	0.0078125	0.015625	0.0078125	0.0	?	?
6/1/2020 9:00	0.0078125	0.0078125	0.015625	0.0078125	0.0	?
6/1/2020 10:00	0.015625	0.0078125	0.0078125	0.015625	0.0078125	0.0
6/1/2020 11:00	0.015625	0.015625	0.0078125	0.0078125	0.015625	0.0078125
6/1/2020 12:00	0.03125	0.015625	0.015625	0.0078125	0.0078125	0.015625
6/1/2020 13:00	0.015625	0.03125	0.015625	0.015625	0.0078125	0.0078125
…	…	…	…	…	…	..

d) Training-Test Splitting of Dataset

In this research, the subsets used were training and test sets without a validation set due to the use of time-series data. The data needed to be sequentially ordered for an algorithm to be correctly classified. Here, the partition used for splitting was 70% for the training set and 30% for the test set. Thus, the days taken into consideration for the training set were June 1st to June 16th, 2021. Meanwhile, the days for the test set are June 16th to June 21st.

2.2.3. Modelling

We mainly proposed different variations of LSTM algorithm[21] for solving this time series problem, i.e., Two-stacked LSTM layers, Bidirectional LSTM, and 1D-CNN LSTM Model. We also proposed several other traditional algorithms, such as the Generalized Linear Regression Model [22], Gradient Boosted Trees [23] and Random Fores t[24], as a baseline comparation.

Across all LSTM architectures, we incorporated the following configuration. Adam optimiser, is 0.005. Early stopping callback is implemented to prevent overfitting for monitoring training loss with a patience distance of 10 epochs. Kernel initialiser was implemented as random normal, specified with a fixed seed to ensure reproducibility of results if applicable to a particular hidden layer. The inputs were reshaped from a one-dimensional column to a three-dimensional array of shapes (samples, 1, time-lags).

a) Two-stacked LSTM

The following is the architecture of the two-stacked LSTM layers on top of one another. Two 32-unit feed-forward LSTM layer with one fully connected regressor is implemented into the output.

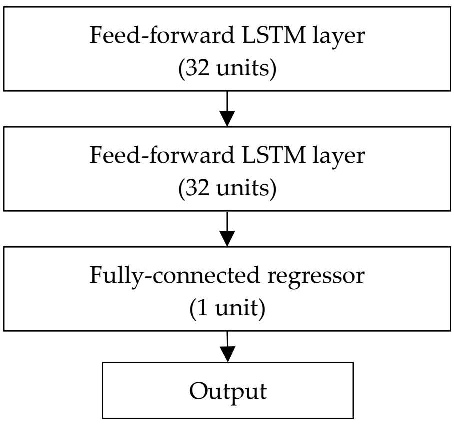

Fig 4. The Two-stacked LSTM Model

b) Bidirectional LSTM

Meanwhile, the bidirectional LSTM was implemented with 50 nodes. The bidirectional LSTM incorporates the same idea as vanilla LSTM. However, this variation allows the network to learn the

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

input sequence forward and backward and concatenate both interpretations (e.g. looking back and peeking forward). While this algorithm is mainly used for sequence classification – more specifically on text data – we observed its performance on this univariate time-series dataset. The model architecture can be seen in Figure 5.

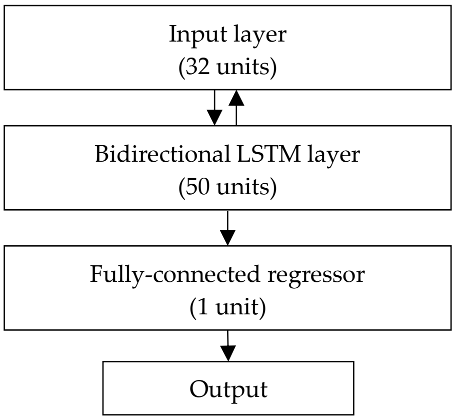

Fig 5. The bidirectional LSTM Model

c) Encoder-Decoder LSTM Model

The encoder-decoder model involves two encoder and decoder layers, respectively, to construct a simple sequence to sequence model—the lag between timesteps are two hours from the past and two hours from the future. While mainly sequence to sequence models are incorporated to tackle machine translation tasks, in this research, we evaluated its performance when given a standard univariate time-series dataset. The architecture for this model can be seen in Figure 7.

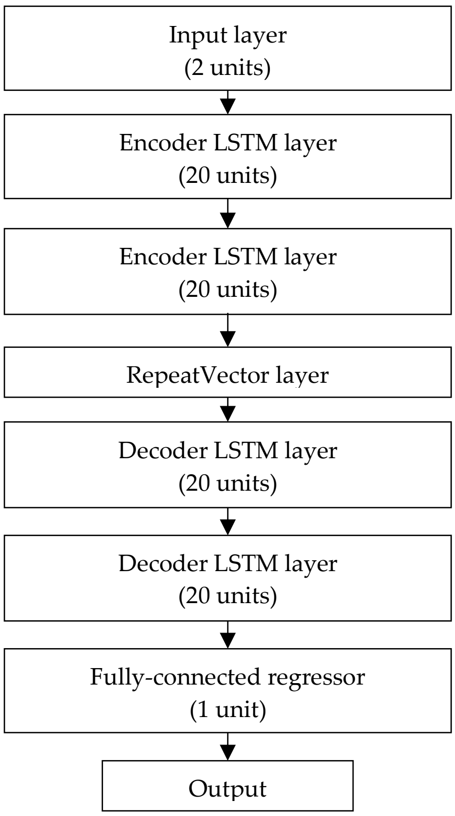

Fig 7. Encoder-decoder LSTM Model

d) 1D-CNN LSTM Model

The 1D-CNN LSTM Model is implemented as convolutional networks paired with LSTM. While often used for image classification tasks, convolutional networks can also be incorporated for sequence-based tasks because of their ability to detect local patterns. Here, we implemented a one-dimensional

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

convolutional layer to detect our dataset's local trends in the first place (by striding the kernels) before eventually passing the data into a pooling layer – in this case, one-dimensional maximum pooling. However, the input sequences were first reshaped into a 4-dimensional array representing samples, subsequences, timesteps, and features. These are subsequences that this study CNN model can process. Because of this, we tuned the time-lag to be 5 hours for the matrices to be divisible. The overall model architecture can be seen in Figure 6 as follows.

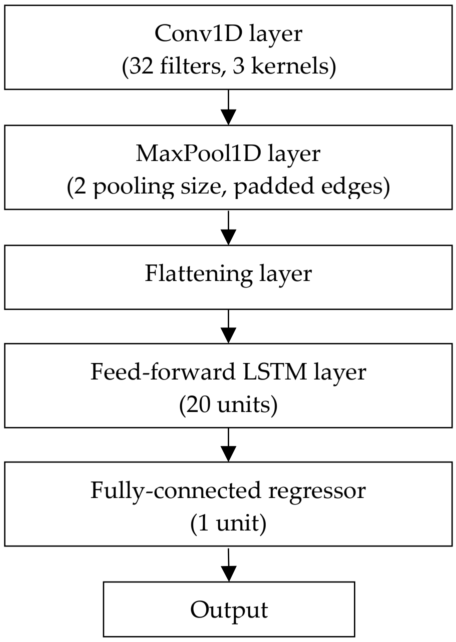

Fig 6. 1D-CNN LSTM Model

3. Results and Discussion

The metrics used to evaluate the performance of models in this research were Mean Squared Error, and Root Mean Squared Error in reverted unscaled data, that is, how many passengers the model(s) predicted. These metrics measure the average value of the errors in a set of predictions, therefore penalising the model for large errors more than other regression metrics such as Mean Absolute Error.

We present Figures 8 to 11 to illustrates how the model performs against the actual value for both training dataset and test dataset. Because of the rescaling of the original dataset, we unscaled (i.e., conducting an inverse transform) the rescaled features to resemble the original data.

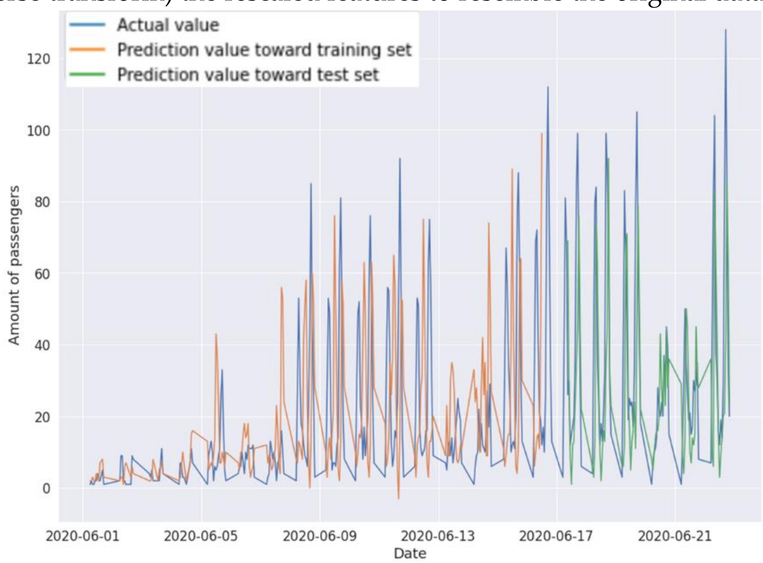

Fig 8. Two-stacked LSTM

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

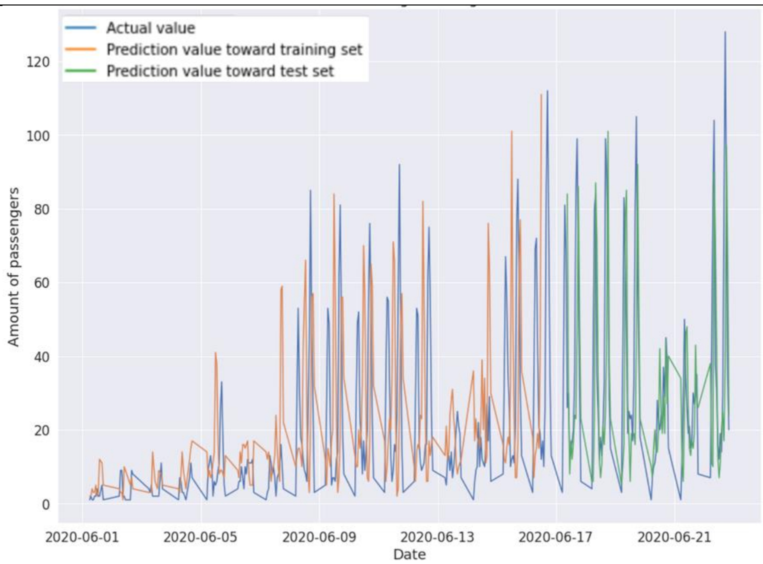

Fig.9. Bidirectional LSTM

Briefly, it can be noticed that the bidirectional LSTM output is not very different compared to the previous two-stacked LSTM. However, if we compare the resulting graphs side-by-side, it is apparent that our bidirectional implementation of LSTM fits the training data better. In the shifted training set, especially, it can be seen that it closely represents the original data.

Figure 10. 1D CNN-LSTM

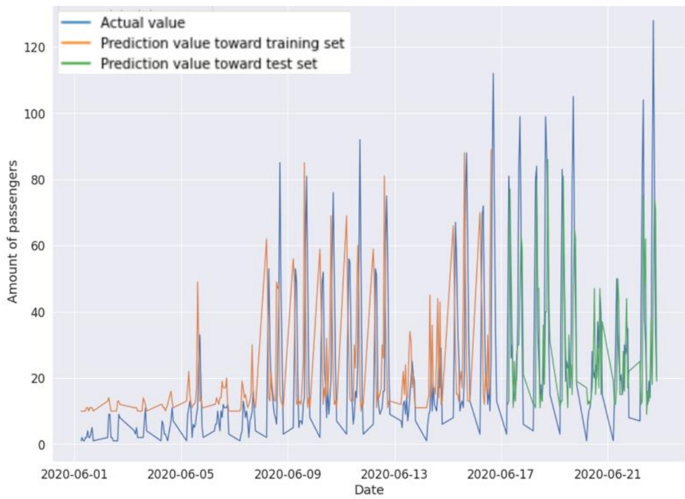

Figure 11. Sequence to Sequence Model

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

Even though the model architecture in this study is very complicated, the encoder-decoder model overshoots the prediction by around 10, making it unreliable for further usage. However, to clarify and further justify this finding, we present each model's performance against Training Set and Test Set in Table 8 and Table 9, respectively.

Table 8. Performance Metrics against Training Set

Performance Metric	Two-stacked LSTM	Bidirectional LSTM	Sequence to Sequence Model	1D CNN-LSTM
Mean Squared Error (MSE)	0.8005	0.8132	1.8551	0.7497
Root Mean Squared Error (RMSE)	10.0762	10.1652	15.3748	9.7713

Table 9. Performance Metrics against Test Set

Performance Metric	Two-stacked LSTM	Bidirectional LSTM	Sequence to Sequence Model	1D CNN-LSTM
Mean Squared Error (MSE)	2.2872	1.9187	4.1550	1.7408
Root Mean Squared Error (RMSE)	17.0648	15.6163	22.9860	14.8666

We also compared the performance of 1D-CNN-LSTM, the best LSTM-based algorithm in this case, with other traditional baseline algorithms, such as Generalised Linear Regression Model, Gradient Boosted Trees, and Random Forest as illustrated in Table 10. The result shows the LSTM-based algorithm outperforms other baseline traditional algorithms, up to 45 times better, in terms of its MSE.

Table 10. Performance Metrics against Test Set compare to Baseline Traditional Algorithm

Performance Metric	Gradient Boosted Tree	Random Forest	Linear Regression	1D CNN-LSTM
Mean Squared Error (MSE)	45.67	45.687	45.612	1.7408
Root Mean Squared Error (RMSE)	2085.873	2087.377	2080.5	14.8666

The model architecture that yields the best performance (with lowest error) is 1D-CNN LSTM, missing the next hour's prediction by 1 to 2 passengers at most. A 1-dimensional convolutional layer shifts multiple kernels across our dataset, resulting in a 1-dimensional per kernel feature map used for local pattern recognition. Coupled with the nature of 1D CNNs that can learn representations of time directly, the development does not necessarily require domain expertise – in this case, passenger trends of MRT Jakarta – to engineer input features from the raw dataset.

Other two implementations of LSTM, i.e., two-stacked LSTM and Bidirectional LSTM, yielded errors close to their 1D-CNN counterpart. Even though the data was transformed inversely and then rounded to resemble the nature of the input features, it is worth noting that these errors were first calculated against rescaled pre-processed data and may not yet represent the actual data.

It is possible to remove the LSTM layer entirely and implement only 1D CNN with a doubling dilation rate, as proposed in [25], in which each sequence is computed by (1), where d is the dilation rate and n is the layer number.

$$d = 2^{n-1} \quad (1)$$

The above step was done so that the first layer yields one dilation rate; the second layer yields 2; the third yields 4, and so on. Despite being claimed to be more robust and use significantly fewer

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

parameters, this model architecture version is more suitable for extensive data sequences such as audio-visual cases. Therefore, the model is not suitable for a small 300-datapoint dataset used in this research.

However, the results show that the sequence-to-sequence model used for machine translation did not perform well in this study. The model architecture achieved the highest error, missing overall predictions by 4 to 5 passengers. This fact indicates that the application of this model for time-series regression is unsuitable, so it is strongly recommended to avoid using this model for the same type of dataset in a future research.

Overall, in this time series of the hourly number of passenger cases, the LSTM-based algorithms have been shown to outperform traditional baseline algorithms, i.e., Generalised Linear Regression Model, Gradient Boosted Trees, and Random Forest. This result reaffirms that LSTM is suitable for forecasting time series problems. LSTM are good at forecasting future values based on prior sequential datasets. The LSTM-based model can train itself to successfully transfer a sequence of prior observations as input to an output observation compared to traditional baseline algorithms above.

4. Conclusion

This study has attempted to estimate the number of passengers using QR codes as payment methods on the MRT Jakarta system by using machine learning technologies. The model is able to determine the number of MRT Jakarta passengers in advance. This study's results can help current stakeholders by assisting them in their decision-making. For example, if the model is used correctly, it can help determine the appropriate time to carry out a promotional advertisement or predict passenger inflow during specific periods. This study uses data mining techniques to develop a prediction model for time series data using machine learning in conjunction with the CRISP-DM methodology. The data samples were 7760 instances aggregated per hour to be passed as input to the machine learning model. The final dataset consists of 319 instances. The performance was assessed for the passenger origin counts, incorporating the metrics Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

The model used in this study is mainly built from LSTM architecture. Based on the evaluation, we find that the most suitable model to implement is 1D CNN. As soon as all of the data was collected and the decision-making for the strategy process began, the number of passengers was determined at the conclusion of the machine learning procedure. The results of this study would also reduce the workload of workers responsible for monitoring passenger activity and mobility during certain hours since the system has taken over this function with the assistance of machine learning.

Furthermore, MRT Jakarta needs to be aware of the risk when deploying and using the model. The development should be tested carefully because any error or miscalculation in the algorithm could alter the result of the machine learning prediction. IT valuation also needs to be held after the deployment to monitor the result of machine learning applications in the sales and marketing department. This valuation could assess any adjustment or update needed for the proposed model to be eventually employed. Moreover, further advancements in the implementation of this model could be utilised to manage the resource allocation for operation management departments across various MRT Jakarta stations.

References

[1] M. Isradi, H. Dwiatmoko, M. D. R. Putri, R. Hidayatullah, and J. Prasetijo, "Analysis Of Effectiveness Service Of Public Transportation Mass Rapid Transit Or MRT Case Study Lebak Bulus–Bundaran HI."

[2] A. F. Dahlan and A. Fraszczyk, "Public Perceptions of a New MRT Service: a Pre-launch Study in Jakarta," *Urban Rail Transit*, vol. 5, no. 4, pp. 278–288, 2019.

[3] C. Zhong *et al.*, "Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data," *PLoS One*, vol. 11, no. 2, p. e0149222, 2016.

[4] W. Setyaningsih, "MRT Jakarta Passengers Reach More Than 19 Million," *BeritaJakarta.ID*, Jakarta, 29-Nov-2019.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (2) July 2022 142-155

[5] I. Atmawidjaja, "Bolstering financial inclusion in Indonesia How QR Codes can drive digital payments and enable financial inclusion," 2018.

[6] C. Shuran and Y. Xiaoling, "A New Public Transport Payment Method Based on NFC and QR Code," in *2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE)*, 2020, pp. 240–244.

[7] P. Suebtimrat and R. Vonguai, "An Investigation of Behavioral Intention Towards QR Code Payment in Bangkok, Thailand," *J. Asian Financ. Econ. Bus.*, vol. 8, no. 1, pp. 939–950, 2021.

[8] L.-Y. Yan, G. W.-H. Tan, X.-M. Loh, J.-J. Hew, and K.-B. Ooi, "QR code and mobile payment: The disruptive forces in retail," *J. Retail. Consum. Serv.*, vol. 58, p. 102300, 2021.

[9] K. Mulia, "Indonesia imposes standard QR code, fixed fees for e-wallets," *Tech in Asia*, 2020. [Online]. Available: https://www.techinasia.com/indonesia-imposes-standard-qr-code-fixed-fees-wallets. [Accessed: 08-Jan-2020].

[10] A. Kim, "RFi Group Insight - Asia: QR Code payments shaping into relevance," *RFi Group*, 2017. [Online]. Available: https://www.rfigroup.com/rfi-group/news/rfi-group-insight-asia-qr-code-payments-shaping-relevance. [Accessed: 10-Aug-2020].

[11] Bank Indonesia, "Indonesia Payment Systems Blueprint 2025 Bank Indonesia : Navigating the National Payment Systems in the Digital Era," 2019.

[12] S. Das, A. Dey, A. Pal, and N. Roy, "Applications of artificial intelligence in machine learning: review and prospect," *Int. J. Comput. Appl.*, vol. 115, no. 9, 2015.

[13] K. Agrawal, K. Maldanna, and G. N. Raj, "Taxi Demand Prediction System Using Machine Learning."

[14] Y. Liu, Z. Liu, and R. Jia, "DeepPF: A deep learning based architecture for metro passenger flow prediction," *Transp. Res. Part C Emerg. Technol.*, vol. 101, pp. 18–34, 2019.

[15] T. Cristóbal, J. J. Lorenzo, and C. R. García, "Using data mining to improve the public transport in Gran Canaria Island," in *International Conference on Computer Aided Systems Theory*, 2015, pp. 781–788.

[16] F. Toqué, M. Khouadjia, E. Come, M. Trepanier, and L. Oukhellou, "Short & long term forecasting of multimodal transport passenger flows with machine learning methods," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 560–566.

[17] X. U. Haitao, P. Jiaxue, N. Zheng, and G. Ying, "Short-term BRT Passenger flow prediction with a deep learning method," *Int. J. Simulation--Systems, Sci. Technol.*, vol. 17, no. 40, pp. 1–6, 2016.

[18] C. Yang, F. Yan, and S. V Ukkusuri, "Unraveling traveler mobility patterns and predicting user behavior in the Shenzhen metro system," *Transp. A Transp. Sci.*, vol. 14, no. 7, pp. 576–597, Aug. 2018.

[19] X. Xu, L. Xie, H. Li, and L. Qin, "Learning the route choice behavior of subway passengers from AFC data," *Expert Syst. Appl.*, vol. 95, pp. 324–332, Apr. 2018.

[20] C. Espinoza and B. Bustos, "Assessing the public transport travel behavior consistency from smart card data," *Transp. Res. Procedia*, vol. 32, pp. 44–53, Jan. 2018.

[21] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] P. McCullagh, "Generalised Linear Models," in *Breakthroughs in Statistics: Methodology and Distribution*, S. Kotz and N. L. Johnson, Eds. New York, NY: Springer New York, 1992, pp. 543–546.

[23] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[24] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[25] K. Simonyan, S. Dieleman, A. Senior, and A. Graves, "WaveNet," *arXiv Prepr. arXiv1609.03499v2*, pp. 1–15, 2016.