The Impact of Mobility Patterns on the Spread of the COVID-19 in Indonesia

Syafira Fitri Auliya [1], Nurcahyani Wulandari [2] *

[1][2] *Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Indonesia*
Jalan Grafika No.2, Yogyakarta
[1]syafira.f.a@mail.ugm.ac.id, [2] nurcahyani.wulandari@ mail.ugm.ac.id

Abstract

Background: The novel coronavirus disease 2019 (COVID-19) has been spreading rapidly across the world and infected millions of people, many of whom died. As part of the response plans, many countries have been attempting to restrict people's mobility by launching social distancing protocol, including in Indonesia. It is then necessary to identify the campaign's impact and analyze the influence of mobility patterns on the pandemic's transmission rate.
Objective: Using mobility data from Google and Apple, this research discovers that COVID-19 daily new cases in Indonesia are mostly related to the mobility trends in the previous eight days.
Methods: We generate ten-day predictions of COVID-19 daily new cases and Indonesians' mobility by using Long-Short Term Memory (LSTM) algorithm to provide insight for future implementation of social distancing policies.
Results: We found that all eight-mobility categories result in the highest accumulation correlation values between COVID-19 daily new cases and the mobility eight days before. We forecast of the pandemic daily new cases in Indonesia, DKI Jakarta and worldwide (with error on MAPE 6.2% - 9.4%) as well as the mobility trends in Indonesia and DKI Jakarta (with error on MAPE 6.4 - 287.3%).
Conclusion: We discover that the driver behind the rapid transmission in Indonesia is the number of visits to retail and recreation, groceries and pharmacies, and parks. In contrast, the mobility to the workplaces negatively correlates with the pandemic spread rate.

Keywords: Business Intelligence, COVID-19, Human Mobility Habit, LSTM, Prediction

Article history: Received 3 February 2021, first decision 25 February 2021, accepted 4 March 2021, available online 28 April 2021

I. INTRODUCTION

The novel Coronavirus or COVID-19 has been spreading rapidly. Firstly, detected on 31 December 2019, this virus was reported to have contaminated 7,818 people worldwide within a month. As of 28 November 2020, there were 61,299,371 confirmed COVID-19 cases, including 1,439,784 deaths globally [1]. Accordingly, WHO gradually increased the risk alert from 'very high' in China only in the third week of January to 'high' at the global level on 30 January 2020, before reaching 'a pandemic' alert on 11 March 2020 [2]. Numerous studies have concentrated on this issue. Despite the development of knowledge about COVID-19, effective treatment has yet to be made available [3]. Therefore, to reduce the transmission of COVID-19, researchers are also working extensively to find the variable that affects the transmission. Few of the variables analyzed are meteorological factors [4], transport accessibility [5], demographic condition [6] and mobility trend [7–9].

As the virus is believed to spread through the cough and sneeze droplets, altering people's mobility such as by implementing social distancing protocol is one way to minimize COVID-19's spread rate. As a consequence of this policy, education institutions are closed, and works are done from home. In brief, people's mobility trend is notably affected—either by a complete lockdown or lesser travel restriction.

Some studies have observed the effects of social distancing measures in minimizing the COVID-19's spread rate, such as in China [7, 9] and Italy [8]. Nonetheless, we believe that the campaign's outcome in Indonesia will be different from either China or Italy because of two aspects. *First*, the data accessible in Indonesia is, in general, less ideal than in China or Italy. In some cases, few additional actions are required to handle the data because of, e.g., missing data or incorrect recordings. *Second*, the COVID-19 testing rate in Indonesia is considerably lower than in China or Italy. On 23 November 2020, Indonesia only conducted 19,444 tests in one million populations (rank 159th in the world)

_____

* Corresponding author

while China and Italy perform 111,163 tests and 337,412 tests in one million populations, respectively [1]. In addition, the testing rate in Indonesia is vastly diverse across the nation. Some provinces with high mobility trends may have low testing rates. In this case, the reported COVID-19 positive cases cannot represent the actual condition, and the COVID-19 prediction or model constructed based on the mobility trends in that areas requires improvement to make the data more accurate.

All things considered, we study two research questions: *First, how does social restriction implementation affect people's movement in Indonesia? Second, how does mobility affect the transmission rate and how fast is the transmission detected?* The transmission period determines the time when a new positive case is confirmed from the contagion, which is potentially longer than the associated 14-day quarantine period.

The remainder of the paper is structured as follows. In Section 2, we discuss related studies about factors affecting the spread of COVID-19 and its prediction methods, as well as the pandemic situation in Indonesia. In Section 3, we introduce our datasets and methods. In Section 4, we present our results. The discussion is presented in Section 5 and is then followed by the conclusion in Section 6.

II. RELATED WORKS

A. Factors that Influence the COVID-19 Spread Rate

The factors that influence the transmission of COVID-19 have become an interest of many stakeholders, including epidemiologists and policymakers [10] [11]. Several studies have proposed that a few of the variables considered to affect COVID-19 spread rate are meteorological factors [4], transport accessibility [5], demographic condition [6] and mobility trend [7-9].

Lin et al. [4] conclude that both temperature and relative humidity influence the transmission of COVID-19. High temperature mitigates virus transmission. Meanwhile, high relative humidity increases the transmission when the temperature is high but decreases the transmission when the temperature is low. Using data from China, Hong Kong, Singapore and few other regions, they used an extended SEIR model to describe the transmission process of the virus, including the pre-symptomatic and the transmission process among patients. Cartenì et al. [5] propose that the greater the transport accessibility in an area, the easier the virus reaches the population. They used a multiple linear regression model by linking the total number of the pandemic cases in Italy to transport accessibility variables, including car/rail accessibility, the average number of daily trips, and average car/rail travel time. This research resulted in 40% in weight. Lulbadda et al. [6] claim that temperature, population size and median age have a positive correlation with the COVID-19 transmission. Using the data from 58 countries covering the initial 60 days from each country's first reported case, they utilized the negative binomial regression model and Pearson Chi-square fit test. The combination of the three variables significantly affected the number of COVID-19 cases.

Fang et al. [7], Cartenì et al. [8] and Aleta et al. [9] suggest that mobility trend is a factor influencing COVID-19 transmission. Fang et al. [7] used a set of difference-in-differences (DID) estimations and revealed that the lockdown in Wuhan decreased people movement to 54.15-76.64%. If Wuhan had not been locked down, the pandemic cases in 363 nearby cities would be 52.64-64.81% higher. Cartenì et al. [8] estimated a model through a multiple linear regression model linking the number of COVID-19 daily cases to some variables, including mobility habits (e.g., daily number of people who commute, transport accessibility, and distance from the main Italian clusters). They found that the number of daily COVID-19 new cases was related to the Italians commuting activities in the previous 21 days (they called it 'positivity detection time'). Aleta et al. [9] constructed an epidemic metapopulation model to compare two radically different scenarios: China without a travel ban in 2019 and China with a travel ban in 2020. They concluded that a travel ban is only effective in the short term but cannot eliminate the pandemic. They argue that even with a travel ban, it is impossible to prevent the virus from spreading to other regions entirely.

B. Predicting COVID-19

In order to predict the trend of COVID-19, some research also studied the COVID-19 data pattern, such as by using the linear regression method [12], the Topological Weighted Centroid (TWC) algorithm [8] and various other machine learning methods [13–16]. Machine learning is widely used in research on predictions because it enables computers to access hidden insights.

Yang et al. [13] used a machine learning approach (especially the LSTM time series model) to train the 2003 SARS data and predict China's epidemic. They predicted that COVID-19 cases would hit a peak in China before gradually declining. They also simulated a five-day delay in control measures (e.g., travel restriction and lockdown) implementation. They concluded that mainland China's epidemic size could have increased three-fold. Golestaneh et al. [14] performed logistic modeling on a cohort of 505,992 ambulatory care patients during pre-COVID and COVID periods. The modeling showed that whites' and blacks' odds of hospitalization are statistically equivalent, but the

mortality rate was significantly higher on black patients. Peipei et al. [15] used LSTM to project the new infections over time for global data, including Brazil, Russia, India, Peru and Indonesia. Using a logistic growth-forecasting model, they estimated that the outbreak would peak globally in late October and infected 14.12 million people. All these forecasts have indeed produced beneficial insight. However, Holmdahl et al. [16] highlight the importance of asking five questions about the model results: its purposes, basic assumption, uncertainty, dataset and context of the model before interpreting the findings. They admitt the usefulness of data-driven forecasting models to make predictions to simulate virus transmission.

C. COVID-19 in Indonesia

COVID-19 was first detected in Indonesia on 2 March 2020, after a dance instructor and his mother were infected following a cross-cultural dance party [17]. Eight months afterward, the COVID-19 new cases and death rate in Indonesia continued to rise, with no sign to diminish just yet [18, 19]. However, the Indonesian government did not follow many countries' mitigation efforts to implement a national lockdown. Until the end of 2020, they only approved large-scale social restrictions (Indonesian: Large-Scale Social Restrictions, abbreviated as PSBB) for several districts and cities with high contamination rates, such as the capital province of DKI Jakarta [20]. Later, the government also started implementing the new normal campaign and classifying green and yellow zones with lower positive cases to reduce the public's anxiety. This policy received much criticism and was considered a 'disaster' because, subsequently, the number of COVID-19 cases in Indonesia continued to increase [21]. In the public's sentiment, although most Indonesians are satisfied with the government approach in dealing with the COVID-19's economic impacts, they criticize the government's overall performance in handling the pandemic [22].

III. METHODS

A. The Dataset

In order to analyze the effect of Indonesian mobility trends on the transmission rate of COVID-19, we used COVID-19 daily new cases data from BNPB and mobility trend dataset from Apple and Google. BNPB (Indonesian National Board for Disaster Management) is one of the core members of Indonesia Satgas COVID-19 (Response Acceleration Task Force). Since the first reported COVID-19 case was on 2 March 2020 [17], to predict the pandemic new cases, we use the data from 2 March to 30 September 2020 for our training dataset and the data during 1-10 October 2020 for the testing (prediction) period. Meanwhile, to obtain an overview of the mobility changes before and during the COVID-19 pandemic, we use the first available data on Google and Apple that are also their baseline period [23, 24], which are 15 February 2020 and 13 January 2020.

1) Indonesia COVID-19 Dataset

The daily outbreak case data were collected from the COVID-19 dashboard on BNPB official site [25]. In fact, there were some disputes and data disparity between BNPB and Indonesia's Ministry of Health. BNPB acknowledged that the COVID-19 daily case that the ministry announced to the public did not match data aggregated by BNPB at the regional level [26]. We still decide to retrieve the COVID-19 data from BNPB because of four reasons summarized in Table 1.

TABLE 1
COVID-19 DATA SOURCES

Source	Concerns
The Ministry of Health	Data was lastly updated on 21 October 2020
Satgas COVID-19	Access to download the raw data (on CVS/EXCEL/JSON format) is not available
KawalCOVID-19	Data on 2 March 2020 is aggregated
WHO	Data is only on a national scale
BNPB	Incorrect and missing data

First, presently, the COVID-19 dashboard at the Ministry of Health's official site [27] was lastly updated on 21 October 2020. Thus, we cannot use that outdated data. Second, while the Satgas COVID-19 also released their dashboard with the ministry supplying the data, the access to download the raw data (on CVS/EXCEL/JSON format) is not available on the site [28]. Third, even though the pandemic's public initiative dashboard KawalCOVID-19 [29] contains complete and up-to-date national COVID-19 data, it lacks detailed data on the provincial level. The data between 2 March and 17 July is only as "< July 18". Meanwhile, we need this data for our training set. Forth, another reliable source, WHO [30], only published data on the national scale. However, COVID-19 data obtained from BNPB also as flaws such as missing data and incorrect values. How we handle these two concerns then will be discussed in Section 3.3.

2) *Indonesia Mobility Trend Dataset*

We obtain our human mobility data from the COVID-19 Mobility Trends of Apple Inc. [23] and the COVID-19 Community Mobility Report of Google LLC [24]. Apple's Mobility Trends depicts mobility trends based on requests for directions in Apple Maps. It shows a relative volume of the requests compared to a baseline volume on 13 January 2020 (before the COVID-19 outbreak started). The categories of the data are (by) driving and (by) walking. The value on the data represents the ratio with the baseline. For example, on 25 January 2020 in the Indonesia dataset, the data is written as "144,2", which means Indonesian people use Apple Maps 44,2% more frequently than on baseline 13 January 2020. Apple provides a dataset for Indonesia (nationwide), Bali and DKI Jakarta.

Meanwhile, Google's Community Mobility Report shows how visits to places are changing since the pandemic started compared to the baseline. The baseline is the median value during the five weeks, from 3 January to 6 February 2020. Google collects data from users who have turned on the Location History setting. The data categories are retail and recreation, groceries and pharmacies, parks, transit stations, workplaces and residential; showing how visiting these places have changed compared with the baseline period. For example, on 15 February 2020, in the Indonesia dataset, the data is written as "-8" in 'park' column; this means Indonesian went to parks much less than between 3 January and 6 February 2020. Google provides a dataset for Indonesia (nationwide) and all 34 provinces.

B. *The Methods*

We first conducted the pre-processing of our dataset. Subsequently, using the variable of mobility trend, we determined the positivity detection time (pdt) that shows the time with the highest possibility of a new positive case being confirmed after he/she is infected. After that, we predicted COVID-19 daily new cases in the next 10 days from the last day at our training dataset.

1) *Pre-Processing*

As mentioned above, BNPB's COVID-19 dataset has missing data and value incorrectness. For handling the missing data, we constructed (1). While we use the logic in (2) to detect incorrect value.

$$Tc_t \begin{cases} Tc_t & , \text{if } day_{(t+1)} = day_t + 1 \\ \frac{Tc_{anom}}{nm+1} \end{cases} \tag{1}$$

$$dnc_t \neq Tc_{t+1} - Tc_t \tag{2}$$

Tc_{t+1} and Tc_t are respectively the value of daily new cases at $day_{(t+1)}$ and day_t. Tc_{anom} is the nearest anomaly value of daily new cases. nm is the days missing nearly. dnc_t is the daily new case at day (t).

We choose the dataset in the capital province, DKI Jakarta, to explain our methods. For 238 days between 2 March and 6 November 2020, 8 days are missed in the dataset. For instance, after 28 July , the next available data is from 30 July (skipping July 29). We assume that this missing data is caused by non-technical reasons such as trouble in gathering data from local government or local hospitals by the central government. The case report thus accumulated from the day after or the day prior. Therefore, we can fill the missing data's value by dividing the accumulation cases in the nearest day with the amount of skipped day(s)+1. Then for the case of 29 July 2020, we can find its value by dividing the value on 30 July by 2.

Meanwhile, the examples for minor incorrect value occurred 57 times where dnc_t was different with $Tc_{t+1} - Tc_t$ (dnc_t is respectively 32, 52, 89,.. while the $Tc_{t+1} - Tct$ is respectively 30, 44, 80, …). One major difference is also present on 3 August 2020, when the daily new cases is recorded to reach 20,036 while the calculation from $Tc_{t+1} - Tc_t$ is 472 (Tc_{t+1} is 23,026 and Tc_t is 22,616). We replace the value of the dnc_t for these days with 472 ($Tc_{t+1} - Tc_t$).

2) *Positivity Detection Time*

We examine the impact of human mobility on COVID-19 daily new cases. As the virus has an incubation period and testing the specimen also takes time, it is reasonable to assume that the COVID-19 daily new cases is related to the mobility trend several days before—a period introduced as "positivity detection time" in [8]. After pre-processing the dataset as mentioned above, we need to find out the most probable positivity detection time by calculating the correlation between COVID-19 daily new cases and every category on Apple's and Google's mobility trends. To achieve it, we modify the Pearson Correlation Coefficient formula into our formula:

$$r_{xi,yi} = \frac{\sum_{k=0}^{14}(x_{i-k} - \bar{x})(y_{i-k} - \bar{y})}{\sqrt{\sum_{k=0}^{14}(x_{i-k} - \bar{x})^2}\sqrt{\sum_{i=0}^{14}(y_{i-k} - \bar{y})^2}} \tag{3}$$

$$x_i = y_i - \text{pdt} \tag{4}$$

x_i is the mobility value from i to the 14 days before. $\bar{x}$ is the mean of the mobility trend for the last 14 days. y_i are the COVID-19 daily new case value from i to the 14 days before; with y_i starts from pdt days before x_i. pdt stands for positivity detection time that will be explained below. $\bar{y}$ is the mean of the COVID-10 daily new case for the last 14 days.

Equation (3) is used to calculate the correlation between COVID-19 daily new cases at day (i) and the value of mobility trends on day (y_i - pdt) for every mobility trend in the dataset during the period between 6 March and 30 September 2020. There are eight categories of mobility trend: two categories "driving" and "walking" from Apple's Mobility Trends; six categories "retail and recreation", "groceries and pharmacies," "parks", "transit stations", "workplaces", and "residential" from Google's Community Mobility Report.

For finding the value of pdt, we construct (5). Pdt represents the "x" day before the day when all eight-mobility categories give the highest accumulation correlation values with COVID-19 daily new cases.

$$CV_{\text{pdt}} = \sum_{k=1}^{8} \sum_{i=0}^{9} \text{freq}(x_{i,j})(5 - i) \tag{5}$$

$\text{freq}(x_{i,j})$ is the frequency value within range i appear at category j. i is the determined correlation value range, with the determined range: {1.00–0.76, 0.75-0.51, 0.50-0.26, 0.25-0.00, (0.01-(-0.25)), ((-0.26)-(-0.50)), ((-0.51)-(-0.75)), ((-0.76)-(-1.00))}. j represents the mobility category. pdt is the day that gives the highest Correlation Value (CV). The pdf chosen is the one that gives the highest CV_{pdt}.

3) Prediction

This study presents the prediction of COVID-19 daily new cases, Indonesian mobility trends, and the correlation between them using the LSTM algorithm. LSTM is one of the algorithms in Neural Network (NN). NN is one of the most prominent ways to predict time-series data because of the mechanism to update its weight value. It also uses backpropagation algorithms to model and extract unseen relationships and features. Because of these mechanisms, the decision in time (y) can be affected by the decision at time step (t-1). Nevertheless, NN has a problem with the vanishing gradient. It uses an activation function to scale the output between 0-1. Hence, when the value is near the border, the change of gradient output is insignificant. RNN then uses a memory mechanism to store information from the previous iteration to overcome this problem [31]. However, RNN can only consider data provided by the previous stage of iteration. Therefore, it has difficulty in learning long-term dependencies. Later, to solve this problem, LSTM adds a Forget Gate to decide the previous state's information that should be forwarded, deleted or modified [32, 33]. The framework of LSTM model is shown in Fig. 1.
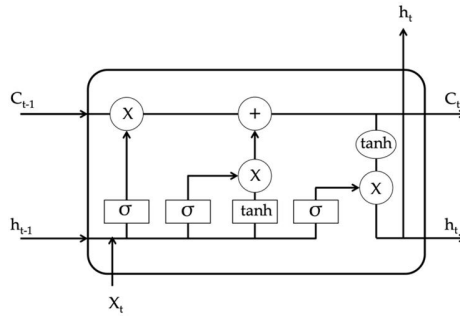

Fig. 1. LSTM model structure. Ct-1 is the previous cell state. Ct is this cell state ht-1 is the hidden state from previous state. ht is the hidden state from this state. σ is the sigmoid activation function. tanh is the tanh activation function. Xt is the input vector. ⊗ represents the element-wise product. ⊕ represents concatenation operation.

First, the decision to discard information from the previous cell state (Ct-1) is made by Forget Gate. The output, a vector ranging from 0 (completely dropping the information from Ct-1) to 1 (keeping the whole information from Ct-1), is decided according to ht-1 and Xt. Second, the new information that will be stored in the cell state is decided. A sigmoid layer determines the updated values, while a tanh layer creates a new candidate value. These two values will be combined to update the state. Third, Output Gate determines the output by capturing the previous hidden state's information (ht-1) and input vector Xt.

For our training data, we have 208 days, from 2 March to 30 September 2020. The testing data is for 10 days that are

predicted using the model constructed in the training process. We set the value for window size as two, which means the model is built by looking for the previous two days as the references. We cannot set higher window size because the number of COVID-19 daily new cases fluctuates and multiply several times in a short duration. We then compare the predictions with the real data from the updated same dataset resources and calculate the prediction error using Mean Absolute Percentage Error (MAPE) as in (6). x is the value of the data on that specific day, n=10 since the prediction is made for 10 days long.

$$\text{MAPE} = \frac{1}{n}\sum_{n=0}^{10}\left|\frac{X_{real} - X_{prediction}}{X_{real}}\right|) \qquad (6)$$

IV. RESULTS

Using the methods mentioned in Section 3, we obtained the result for: positivity detection time (pdt), prediction of COVID-19 daily new cases, prediction of mobility trend, and the mobility patterns that influence COVID-19 daily new cases most significantly in Indonesia.

A. Positivity Detection Time

Utilizing (5), we calculated pdt to determine the "x" day before, where eight-mobility categories give the highest accumulation correlation values with COVID-19 daily new cases, as seen in Fig. 2.
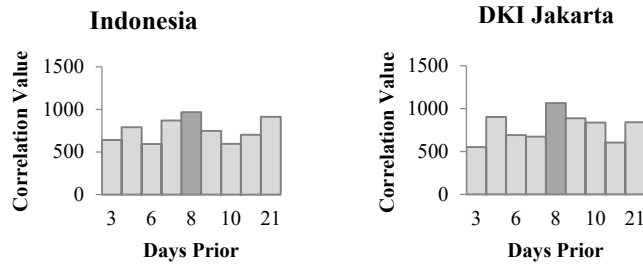

Fig. 2. pdt (days prior when mobility trends influence the transmission the most) in Indonesia and Jakarta

As seen in Fig. 2, the highest value of pdt is at n = 8 in Indonesia (as national wide) and DKI Jakarta. It means that by calculating the correlation between COVID-19 daily new cases and the mobility trends during the period of 6 March to 30 September 2020, we discovered that the COVID-19 cases in Indonesia and DKI Jakarta was mostly connected with the mobility done in the past eight days. Using the concept of business intelligence, we also visualized the more detailed correlation on pdt=8 in Fig. 3.
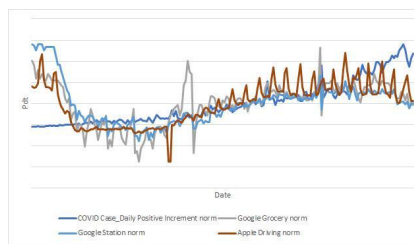

Fig. 3. Correlation Value on pdt=8 in DKI Jakarta

B. Prediction

1) COVID-19 Daily New Cases

We used COVID-19 Daily New Cases in Indonesia (national-scale) and DKI Jakarta as our examples. Jakarta was chosen because it was the epicenter in Indonesia. We focused on finding the lowest loss value and ignored the execution time constraints. We then got the best LSTM performance using a learning rate of 0.005 and epoch 50. We obtained a 0.050 loss of value for the Indonesia dataset and a 0.055 loss of value for the DKI Jakarta dataset from these variables. As a comparison, using the same algorithm, we also trained and predicted the pandemic on a global scale with a result of 0.018 loss of value.

TABLE 2
LOSS AND MAPE OF COVID-19 DAILY NEW CASES

	Loss Avg (Training)	MAPE (Prediction)
Indonesia	0.050	6.2%
DKI Jakarta	0.055	9.4%
Worldwide	0.018	7.1%

TABLE 3
COMPARISONS BETWEEN PREDICTION AND ACTUAL COVID-19 DAILY NEW CASES IN INDONESIA, DKI JAKARTA AND WORLDWIDE

| Date	Indonesia		DKI Jakarta		Worldwide	
	Prediction	Actual	Prediction	Actual	Prediction	Actual
October 1	3945,6	4232	1180,0	1225	292306,7	316025
October 2	3996,6	4232	1177,9	1225	313602,7	323180
October 3	4140,0	3988	1192,1	1265	307499,8	318812
October 4	4117,7	3987	1098,8	1398	294435,2	301978
October 5	4164,6	3610	1129,3	1022	289167,1	281076
October 6	4153,0	4051	1205,0	1107	276862,5	270276
October 7	4278,9	4528	1198,7	1211	290895,3	326777
October 8	4399,0	4848	1159,2	1182	310013,9	353626
October 9	4473,7	4080	1113,8	943	313728,2	363056
October 10	4404,7	4385	1149,1	1324	307009,2	349258
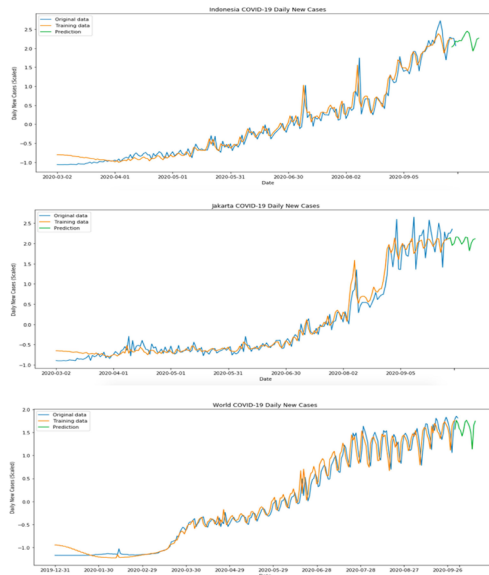
Fig. 4. The visualization of COVID-19 daily new cases prediction result in respectively Indonesia, Jakarta, and worldwide using LSTM

TABLE 4
LOSS IN TRAINING AND MAPE IN PREDICTION PROCESS FOR MOBILITY DATA

Categories	Loss in Training		MAPE in Prediction	
	Indonesia	DKI Jakarta	Indonesia	DKI Jakarta
Google				
Retail & Recreation	0.103	0.098	17.4%	6.9%
Groceries & Pharmacies	0.242	0.256	287.3%	17.5%
Parks	0.175	0.063	83.4%	13.3%
Transit Stations	0.058	0.054	11.0%	19.4%
Workplaces	0.501	0.453	67.1%	38.4%
Residential Area	0.222	0.224	44.4%	23.2%
Apple				
Driving	0.138	0.142	18.2%	9.3%
Walking	0.076	0.165	11.8%	9.5%

As seen in Fig. 5, the training value (orange color) depicts the real (blue color) trends. The model constructed from that training process was thus used to construct the prediction (green color). Using that results, we calculated that the

error rate (MAPE) of our method are: 6.2%, 9.4%, and 7.1% for respectively Indonesia (national-scale), DKI Jakarta and the worldwide dataset. Following the categorization of Lewis [34], the forecasts (MAPE <10%) are interpreted as highly accurate. The summary of the results obtained can be seen in Table 2 and Table 3. The visualizations are present in Fig. 4.
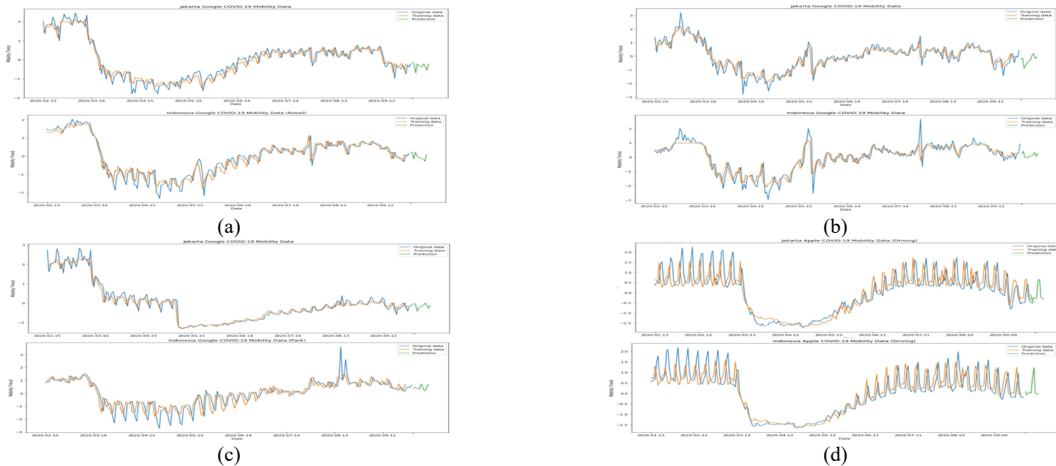

(a)   (b)

(c)   (d)

Fig. 5. The visualization of mobility trend prediction result in DKI Jakarta (above image) and Indonesia (below image) using LSTM on the selected categories: (a) Retail and Recreation, (b) Groceries and Pharmacies, (c) Park, and (d) (by) Driving

2) Mobility Trends

With Indonesia (national scale) and DKI Jakarta as our examples, we predict their mobility data using Google and Apple dataset. Using the same LSTM algorithm and variable mentioned above, we obtain the losses and MAPE as seen in Table 3 and Table 4. We calculated that the error on MAPE of our method was at 6.4% - 287.3%, which highly varies. We analyze this result in Section 5. The visualizations are present in Fig. 5.

From eight mobility trend categories, we chose four categories to be put in Fig. 5, with the exact value is revealed in Table 4. As seen in Fig. 5, in general, the mobility trend in both Indonesia (national wide, below images in each category) and DKI Jakarta (above images in each category) show a similar pattern.

The mobility trend dramatically decreased in March when the president of Indonesia announced the first COVID-19 positive cases, and some cities finally began to restrict their citizens' movement. However, information about the danger of COVID-19 was not disseminated well enough in Indonesia. As seen in Fig. 5, slowly, people (especially in cities outside DKI Jakarta) started to mobile normally and often without adequate health measurement, which increased the virus transmission rate. We will elaborate on these findings in Section 5.
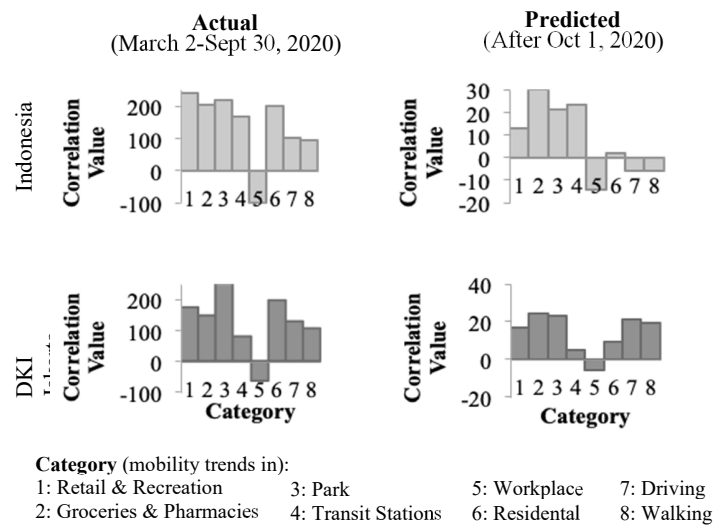

Category (mobility trends in):
1: Retail & Recreation   3: Park   5: Workplace   7: Driving
2: Groceries & Pharmacies   4: Transit Stations 6: Residental   8: Walking

Fig. 6. Correlation value among eight mobility trends

C. The Most Influential Features

We calculated the correlation values for eight mobility trends using (5) (for k = 1, 2, …8, which k represents a category of mobility trends) with the results in Fig. 6. As seen in Fig. 6, in general, the mobility in workplaces (category number "5") correlates negatively with COVID-19 daily new cases. Meanwhile, other categories correlate positively, although there is discrepancy between the actual data and the prediction result. The most influencing mobility categories in Indonesia (nation- wide) and DKI Jakarta are also not identical. We will discuss this result more in Section 5.

V. DISCUSSION

We inferred some analysis from the results above to answer our research questions.

First, how does social restriction implementation affect people's movement in Indonesia?

Fig. 5 illustrates that Indonesia and DKI Jakarta people dramatically decreased their mobility started from the beginning of March. In the following seven months, the majority of Indonesians stayed at home. However, starting from October, the trends were altered, especially in the categories of groceries and pharmacies (both in Indonesia and DKI Jakarta), park (in Indonesia), workplaces (both in Indonesia and DKI Jakarta) and residential area (in Indonesia). In other words, starting in October, many Indonesians have started to ignore social distancing to buy their groceries and to go to parks and their office. Meanwhile, people, especially in DKI Jakarta, generally followed the protocol to stay home. Some exceptions are that many people in DKI Jakarta have started to pick up their groceries and go to the office in person. Nevertheless, perhaps because people in DKI Jakarta are more accustomed to online grocery shopping and the multi/national companies in DKI Jakarta have more resources to make the online working successful, the movement is less than that of in other provinces.

Second, how does mobility affect the transmission rate and how fast is the transmission detected?

Fig. 1 shows that the eight-mobility categories give the highest accumulation correlation values with COVID-19 daily new cases eight days before. It means that the pandemic daily new cases in Indonesia and DKI Jakarta are mostly related to the Indonesian mobility trend conducted in the past eight days. This result is in line with the sum of this virus's 5-6 days average incubation periods (according to WHO [35] and the spokesperson of Indonesia COVID-19 task force [36]) and 2-4 days required to test the specimen in the labs [37, 38]. However, we released that this analysis might be inaccurate since (1) many Indonesians, even having COVID-19 symptoms, with various reasons (most likely worried about negative stigma), avoid being tested. So even if a significant number of people went to crowded places (that increase the mobility trend) and infected with the virus there, the COVID-19 daily new cases might not be affected (since they are not tested); (2) the testing duration varies across the country. In some more isolated areas, it takes 14 days to confirm the status of the suspected-specimen [39]. As this work uses Indonesia (most detected cases are from big cities) and DKI Jakarta (that has sufficient laboratory facilities) as our examples, further study using other local areas dataset might result in different trends and insights than ours.

In addition, from Fig. 6, we infer that the increase of people visits to retail and recreation, groceries and pharmacies, and parks constantly become the most influential features affecting the pandemic spread in Indonesia. These three areas have more people gathering and interacting together, so the risk of transmission is higher here. Meanwhile, the amount of people coming to work has a negative correlation with COVID-19 spread. We cannot find a reasonable explanation supporting this counterintuitive result with our current data, so further specific and comprehensive studies are required to explain this.

VI. CONCLUSIONS

This study investigates the influence of mobility trends on the spread of the COVID-19 pandemic in Indonesia and DKI Jakarta. We hypothesize that the number of COVID-19 daily new cases should be related to commuting activities accomplished several days before. Using a modified Pearson's correlation formula into mobility trend dataset from Google and Apple, we found that all eight-mobility categories result in the highest accumulation correlation values between COVID-19 daily new cases and the mobility eight days before. We called these eight days the 'positivity detection time'. Using Long Short-Term Memory (LSTM) algorithm, we also made forecasts of the pandemic daily new cases in Indonesia, DKI Jakarta and worldwide (with error on MAPE 6.2% - 9.4%) as well as the mobility trends in Indonesia and DKI Jakarta (with error on MAPE 6.4 - 287.3%).

We discover that for the first seven months starting from March 2020, people in Indonesia followed the social distancing protocol by staying at home. However, starting in October 2020, people began to mobile. We also discover that the increase in the number of visits to retail and recreation, groceries and pharmacies, and parks are the most influencing factor of COVID-19 transmission rate. Therefore, we suggest that the government of Indonesia and all related stakeholders put in place more stringent measures in these three places. Visits to workplaces has a negative

correlation with the COVID-19 transmission rate. We cannot apprehend why activities in the workplace area do not positively influence the COVID-19 spread rate, so further studies are required to address this phenomenon.

Author Contributions: *Syafira Fitri Auliya*: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing—Original Draft Preparation, Writing—Review and Editing, Supervision, Project Administration, Funding Acquisition. *Nurcahyani Wulandari*: Conceptualization, Software, Validation, Resources, Writing—Review and Editing, Visualization, Funding Acquisition

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1]   Worldometer, "Worldometer COVID-19 coronavirus pandemic." [Online]. Available: https://www.worldometers.info/coronavirus/. [Accessed: 23-Nov-2020].

[2]   WHO, "Archived: WHO Timeline Covid-19", 2020. [Online]. Available: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19. [Accessed: 10-Nov-2020].

[3]   S. Roy and P. Ghosh, "Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking," PLoS One, vol. 15, no. 10, pp. 1–18, 2020, doi: 10.1371/journal.pone.0241165.

[4]   J. Lin et al., "Containing the spread of coronavirus disease 2019 (COVID-19): Meteorological factors and control strategies," Sci. Total Environ., vol. 744, no. December 2019, p. 140935, 2020.

[5]   Cartenì, L. Di Francesco, and M. Martino, "The role of transport accessibility within the spread of the Coronavirus pandemic in Italy," Saf. Sci., vol. 133, no. July 2020, p. 104999, 2021.

[6]   K. T. Lulbadda, D. Kobbekaduwa, and M. L. Guruge, "The impact of temperature, population size and median age on COVID-19 (SARS-CoV-2) outbreak," Clin. Epidemiol. Glob. Heal., no. August, pp. 0–1, 2020.

[7]   H. Fang, L. Wang, and Y. Yang, "Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China," J. Public Econ., vol. 191, p. 104272, 2020.

[8]   Cartenì, L. Di Francesco, and M. Martino, "How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study," Sci. Total Environ., vol. 741, p. 140489, 2020.

[9]   Aleta, Q. Hu, J. Ye, P. Ji, and Y. Moreno, "A data-driven assessment of early travel restrictions related to the spreading of the novel COVID-19 within mainland China," Chaos, Solitons and Fractals, vol. 139, p. 110068, 2020.

[10]   W. Liu et al., "Analysis of factors associated with disease outcomes in hospitalised patients with 2019 novel coronavirus disease," Chin. Med. J. (Engl)., vol. 133, no. 9, pp. 1032–1038, 2020, doi: 10.1097/CM9.0000000000000775.

[11]   US Pharm, "Factors Affecting COVID-19 Transmission," 2020, [Online]. Available: https://www.uspharmacist.com/article/factors-affecting-covid19-transmission. [Accessed: 10-Nov-2020].

[12]   K. Gowthami and M. R. P. Kumar, "Study on Business Intelligence Tools for Enterprise Dashboard Development," Int. Res. J. Eng. Technol., vol. 4, no. 4, pp. 2987–2992, 2017, [Online]. Available: https://www.irjet.net/archives/V4/i4/IRJET-V4I4721.pdf. [Accessed: 10-Nov-2020].

[13]   Z. Yang et al., "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," J. Thorac. Dis., vol. 12, no. 3, pp. 165–174, 2020, doi: 10.21037/jtd.2020.02.64.

[14]   L. Golestaneh et al., "The association of race and COVID-19 mortality," EClinicalMedicine, vol. 25, p. 100455, 2020, doi: 10.1016/j.eclinm.2020.100455.

[15]   P. Wang, X. Zheng, J. Li, and B. Zhu, "Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics," Chaos, Solitons and Fractals, vol. 139, p. 110058, 2020, doi: 10.1016/j.chaos.2020.110058.

[16]   Holmdahl, I. and Buckee, C., 2020. Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. New England Journal of Medicine.

[17]   R. Nuraini, "Kasus Covid-19 Pertama, Masyarakat Jangan Panik." [Online]. Available: https://indonesia.go.id/narasi/indonesia-dalam-angka/ekonomi/kasus-covid-19-pertama-masyarakat-jangan-panik. [Accessed: 10-Nov-2020].

[18]   B. Nugraha, L. K. Wahyuni, H. Laswati, P. Kusumastuti, A. B. Tulaar, and C. Gutenbrunner, "COVID-19 pandemic in Indonesia: Situation and challenges of rehabilitation medicine in Indonesia," Acta Med. Indones., vol. 52, no. 3, pp. 299–305, 2020.

[19]   D. M. Purnamasari, "Capai 657.948, Covid-19 di Indonesia belum menurun." [Online]. https://nasional.kompas.com/read/2020/12/20/06490421/capai-657948-covid-19-di-tanah-air-belum-menurun. [Accessed: 10-Nov-2020].

[20]   Pemda DKI Jakarta, "JAKARTA TANGGAP COVID-19," 2020. [Online]. https://corona.jakarta.go.id. [Accessed: 10-Nov-2020].

[21]   J. P. BEans, "Indonesia's 'new normal' a disaster in the making," 2020. [Online]. https://asiatimes.com/2020/06/indonesias-new-normal-a-disaster-in-the-making/. [Accessed: 23-Nov-2020].

[22]   Prastyo, P. H., Sumi, A. S., Dian, A. W., & Permanasari, A. E. (2020). Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel. Journal of Information Systems Engineering and Business Intelligence, 6(2), 112-122.

[23] Apple Inc., "Mobility Trends," 2020. [Online]. Available: https://covid19.apple.com/mobility. [Accessed: 06-Nov-2020].

[24] Google LLC., "Community Mobility Report," 2020. [Online]. Available: https://www.google.com/covid19/mobility/. [Accessed: 06-Nov-2020].

[25] Gugus Tugas Percepatan Penanganan COVID-19 Republik Indonesia, "Indonesia COVID-19 Hub Site", 2020. [Online]. Available: https://bnpb-inacovid19.hub.arcgis.com. [Accessed: 06-Nov-2020].

[26] M. K. Indonesia, "Beda Data BNPB vs Kemenkes," 2020.

[27] Kementerian Kesehatan Republik Indonesia, "Dashboard Data Kasus COVID-19 di Indonesia," 2020. [Online]. Available: https://www.kemkes.go.id/article/view/20031900002/Dashboard-Data-Kasus-COVID-19-di-Indonesia.html. [Accessed: 06-Nov-2020].

[28] Satuan Tugas Penanganan COVID-19. "Peta Penanganan,". 2020. [Online]. Available: https://covid19.go.id/peta-sebaran. [Accessed: 06-Nov-2020].

[29] Kawal COVID-19. "Informasi Terkini COVID-19 di Indonesia,". 2020. [Online]. Available: https://kawalcovid19.id. [Accessed: 06-Nov-2020].

[30] World Health Organization. "WHO Coronavirus Disease (COVID-19) Dashboard,". 2020. [Online]. Available: https://covid19.who.int. [Accessed: 06-Nov-2020].

[31] S. Hochreiter, "the Vanishing Gradient Problem during Learning Recurrent Neural Sets and Problem Solutions," Int. J. Uncertainty, Fuzziness, Knowledge-Based Syst., vol. 6, no. 2, pp. 107–116, 1998.

[32] S. Hochreiter. "Long short-term memory". Neural Comput. 1997, 9, 1735–1780

[33] Gers, F.A.; J. Schmidhuber.; F. Cummins. "Learning to Forget: Continual Prediction with LSTM". Neural Comput. 2000, 12, 2451–2471.

[34] Lewis, C.D. (1982). "Industrial and business forecasting methods". London: Butterworths.

[35] World Health Organization. "Transmission of SARS-CoV-2: implications for infection prevention precautions". 2020. [Online]. Available: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions. [Accessed: 10-Jan-2021].

[36] Berita Satu. "Rata-rata Masa Inkubasi Covid-19 di Indonesia, 5-6 Hari". 2020. [Online]. https://www.beritasatu.com/kesehatan/619139/ratarata-masa-inkubasi-covid19-di-indonesia-56-hari. [Accessed: 10-Jan-2021].

[37] Detik News. "Berapa Lama Pemeriksaan Spesimen Virus Corona? Ini Kata Kemenkes". 2020. [Online]. https://news.detik.com/berita/d-4930237/berapa-lama-pemeriksaan-spesimen-virus-corona-ini-kata-kemenkes. [Accessed: 10-Jan-2021].

[38] Koran Tempo. "Durasi Pemeriksaan Covid di DKI Makin Cepat". 2020. [Online]. https://koran.tempo.co/read/metro/454483/durasi-pemeriksaan-covid-di-dki-makin-cepat. [Accessed: 10-Jan-2021].

[39] Republika. "Lamanya Hasil Swab Test Jadi Masalah di Daerah". [Online]. https://republika.co.id/berita/q98fb5328/lamanya-hasil-tes-emswabem-jadi-masalah-di-daerah. [Accessed: 10-Jan-2021].