Visualization of Prediction of The Spread of Covid-19 in Indonesia using Cellular Automata

Safira Aurellia Azzahra
Informatics Engineering
Universitas Trisakti
Jl. Kyai Tapa, No. 1, Indonesia
Email: Safira064001700017 [AT] std.trisakti.ac.id

*Dian Pratiwi
Informatics Engineering
Universitas Trisakti
Jl. Kyai Tapa No. 1, Indonesia
Email: dian.pratiwi [AT] trisakti.ac.id

Syandra Sari
Information System
Universitas Trisakti
Jl. Kyai Tapa No. 1, Indonesia
Email: syandra_sari [AT] trisakti.ac.id

Abstract—**The first case of COVID 19 was detected in Indonesia in early March 2020. One way to assist the government in making decisions to deal with COVID-19 is to create a map of the distribution of COVID-19 patients based on which can only be accessed by people who have an interest through the website. The data used in this study is the period, location, total cases. After getting the data, the data is then processed to get weekly rules. After getting the weekly rules, the data is entered into the calculation of the Moore scheme to get the prediction results for the next week. Then the prediction results are poured in the form of a map. The prediction process using CA neighbors is carried out using Moore's formula, a formula that applies the adjacent neighbors of 8 neighbors. The accuracy level of Cellular Automata with Moore's neighbors reaches 431.1466353% using MAPE. The error value in the Cellular Automata method is quite high due to several factors that make the prediction results with the original data results different, but this method can be used to research cases of covid 19**

Keywords; cellular automata; covid-19, moore, virus

I. INTRODUCTION

Coronavirus disease 2019 or commonly known as COVID-19 is a disease caused by a new type of coronavirus, namely Sars-CoV-2. Coronavirus is a very large virus family. This virus was first discovered in Wuhan, China at the end of 2019. As of December 19, 2020, COVID-19 cases in the world reached 77,801,721. On December 19, 2020, the largest number of COVID-19 cases was recorded in the United States with 18,188,744 cases, 321,631 deaths, and 55,192 recovered patients. The first case of COVID-19 was detected in Indonesia in early March 2020, Jakarta is the epicenter of the spread of COVID-19 cases in Indonesia, followed by other big cities. COVID-19 has an impact on human life in the world, including Indonesia. In addition to the impact on health, COVID-19 has an impact on the economy, so this problem has prompted the government to rack its brains to find new policies to deal with problems from the impact of COVID-19. Policies that optimize state and regional budgets for the need of preventing the spread of COVID-19 and optimizing local economic potential to meet community needs sustainably [1] [2]. If the COVID-19 problem continues to be ignored, Indonesia's condition will get worse, especially the Indonesian economy. Since the pandemic, many Indonesians have lost their jobs due to the company's lack of income from their production so companies cannot pay salaries to their employees, especially laborers. Therefore, the COVID-19 case needs to be resolved, not only medical personnel who are always on the cutting edge who are always sacrificed by the community because of their duties, but citizens need to take part in eradicating this covid 19 case using social distancing.

In addition to social distancing, another way to assist the government in making decisions to deal with COVID-19 is to create a map of the distribution of COVID-19 patients that can be accessed only by people who have an interest through the website. This map of the distribution of covid 19 can be made by the community, including students majoring in informatics. They can make a map using predictions for the spread of COVID-19 cases in Indonesia. To make predictions that match the spatial and temporal conditions of Indonesia, a cellular geographic model approach is needed using the Cellular Automata method. The cellular automata model is considered better than other scientific studies, even according to Floreani and Mattiussi cellular automata have proven reliable in presenting spatial and temporal dynamics which makes it interesting to study because it can explore complex dynamics knowledge with simple computational formulas [3] [4].

Cellular Automata (CA) was first introduced by John Von Nouman and Stain Slaw MarchinUlam and named cellular space in 1948 as a simple model for investigating the behavior of complex systems in general and studying biological processes. The concept of cellular automata has been developed since 1940 by scientists, including in 1970 CA was applied to a popular game called Game of Life. Then CA was applied in the field of geography by Tobler in 1979. Since CA was applied in the field of geography by Tobler then many studies on CA in geography, including predicting the spread of disease in an area [5][6]. Therefore, this study chose to use this method to predict the spread of COVID-19 in Indonesia. After doing the calculations using the CA entered into the PHP programming language.

II. LITERATURE REVIEW

A. Cellular Automata

Cellular automata were first introduced by John Neuman and Stain SlawaMarchinUlam under the name cellular space in 1948 as a simple model for investigating the behavior of complex systems in general and studying biological processes.

Cellular automata were used by Ulan and Von Neuman as the basis for observing complex phenomena. Cellular automata is an example of a genetic algorithm, which means that cellular automata are part of AI. Cellular automata are also mathematical modeling of a physical system, where the space and time of the interrelated system are made discrete. In other words, cellular automata are defined as a dynamic system that has discrete properties that are discrete from a collection of identical cells [4] [5] [7].

Cellular automata (CA) works in discrete time stages, where the value of the variable in each cell is influenced by the value of the variable in its neighboring cell in the previous time step. Prof. Andreas said that cellular automata have 3 main principles. The first principle is that CA has a spatial principle. The CA concept is likened to a city that has many buildings, if you only look at one particular building, you will also see the closest neighboring buildings. The second principle is local interaction, this shows that one individual in automata cells can interact with neighboring cells. Local interaction itself is a limit to the extent to which the neighboring cell and the ignored cell can be considered as 'nearest neighbors. The last principle is the change of choices made by the main cell.

According to Liu cellular automata has 5 elements, namely [4]:

a) Cell

Cells are arranged in a spatial tessellation, a two-dimensional grid of cells that is the most common form of CA used in urban growth models. CA is written in graph form and the cells are vertices, then neighbor relations are represented by edges.

The model of Cellular Automata has four interacting components, which can be denoted as follows:

$$CA = \{S, T, Tr, N\} \quad (1)$$

where:

S: A spatial space that contains a grid in which each cell has the value of one or more specified variables.

T: The observation time-space where the grid can change its value from one time to another

Tr: The Transitional Rule is an engine that drives grid change

N: Neighborhood scheme

The change in cell (S) from the initial state (St) at time t to (St+1) at time t+1 is the surrounding condition (N) and a certain transition principle (T). Mathematically the change function can be denoted as follows:

$$St + 1 = f(St, N, T) \quad (2)$$

b) Condition (state)

The state is an attribute of a system. Each cell is only allowed to take one condition from a set of conditions at any given time.

c) Neighborhood (neighborhood)

The neighborhood is a component that is one of the characteristics of Cellular Automata. The value of a cell will change at a certain time as the influence of several cells that are neighbors of the cell (Neighborhood Cell). The number of neighboring cells that is commonly used is the neighbor of 4 (called the Von Neuman neighborhood) or 8 (called the Moore neighborhood).

Figure 1. Neighbor Configuration on Cellular Automata; Von Neumann Neighborhood (left); Moore Neighborhood (right)

Moore Neighborhood is an extended neighbor of Von Neuman Neighborhood which also has diagonal cells.

d) Transition rules

Transition rules are a definition of the changing response of a cell in response to a current state and neighboring conditions.

e) Time (time-step)

The time-step variable determines the time dimension used during the calculation process in the CA process.

B. Covid-19

COVID-19 or Coronavirus disease 2019 is a disease caused by a new type of coronavirus, namely Sars-CoV-2. This virus was first discovered in Wuhan, China at the end of 2019. Coronaviruses are a very large virus family. WHO said the increase in the number of COVID-19 cases was going quite quickly and spread to other areas outside Wuhan and other countries. On January 30, 2020, WHO declared it a Public Health Emergency of International Concern (PHEIC). As of May 6, 2020, active cases of covid reached 2,226,335 patients with details of 2,177,007 patients with mild condition status and 49,328 patients with severe condition status. On May 6, the United States recorded the most cases of COVID-19 with 1,234,592 cases, 72,054 deaths, and 199,151 recovered patients. The first case of covid 19 was detected in Indonesia in early March 2020, Jakarta is the epicenter of the spread of covid 19 cases in Indonesia, followed by other big cities. Until now, there is no cure or vaccine for COVID-19 [2] [5].

Many studies have examined this covid case, one of which is "The Role of Geography in the Analysis of the Distribution of Covid-19", this study explains that COVID-19 has spread throughout the world, in this case, it is not only the role of medical personnel needed during this pandemic period in terms of caring for COVID-19 patients, the role of a geographer is very important in analyzing and making policies to handle this pandemic and is assisted by UI, UGM, and the task force to assess the need for geospatial information. In addition to this research, there are many other studies examining COVID-19 cases, one of which is the "Alternative Strategy for Handling the Economic Impact of COVID-19 East Java Regional Government in Agropolitan Areas", this study explaining how the economic impact that occurred in East Java which is an agricultural area, livestock, fisheries, plantations, and forestry.

C. Data Visualization

Data visualization is a form of communication of information by presenting it in the form of graphics or images. An image can be understood faster than writing. Interpretation of an object occurs in parallel in a system of human perception, while the speed of analyzing information from writing is limited by the speed of its reading. Images can also be understood by people using different languages because images do not depend on local languages. Examples are traffic signs, railroad maps, weather charts, or weather symbols. The benefit of using data visualization is to understand the information conveyed by the data and to make a decision [8].

Data visualization is very helpful in research because this data visualization can display the results of a researcher's analysis, for example, research on the "Simulation System for the Pattern of Spread of Septicemia Epizootica Disease Using the Cellular Automata Method in NTT" before simulating the results of predictive calculation data using CA the results are visualized using a cartogram which is a chart in the form of a map, this cartogram is useful for displaying which areas are affected by cow disease, namely septicemia epizootic. Furthermore, there is another study that uses data visualization in displaying their research results, namely the research "Parameters Estimation of Generalized Richards Model for COVID-19 Cases in Indonesia Using Genetic Algorithm". Line charts [4].

D. Geospatial Information System (GIS)

Geographic information system or in English the Geographic Information System is abbreviated as GIS. Is a special information system that manages data that has spatial information or in a narrower sense is a computer system that has the ability to build, store, display and manage geographic information to support decision making in the management and design of an area [7].

III. ANALYSIS

In making a Visualization of the Prediction of the Spread of COVID-19 in Indonesia, we first collect data on the number of COVID-19 cases in Indonesia per province. The Visualization of the Spread of COVID-19 Cases in Indonesia provides a feature that can see predictions of the spread of COVID-19 in the following weeks by looking for patterns from the previous weeks.

A. Data Preparation

Research data is taken from the kaggle.com site, this site is a site to get data that will be used in data science, on this site there are 50,000 public datasets, and one of them is a dataset regarding COVID-19. The data obtained for this study started from March 3, 2020 to September 12, 2020. The COVID-19 dataset consists of data on the date of the incident, the location of the incident (province), new cases, new deaths, recovered patients, active cases, total cases, total deaths, total patients recovered, total active cases, level of location, province, country, continent, island, time zone, special status, total district, total city, total district, total village, number of villages pendesa, area, area population, population density, longitude, latitude, new cases per million, total new cases per million, case recovery rate. With the word in one date there are 37 attributes. Meanwhile, the number of data is 6084 cases. Then the data is processed into data that is only needed.

B. Research Flowchart

The following is a flowchart in carrying out the research stages:

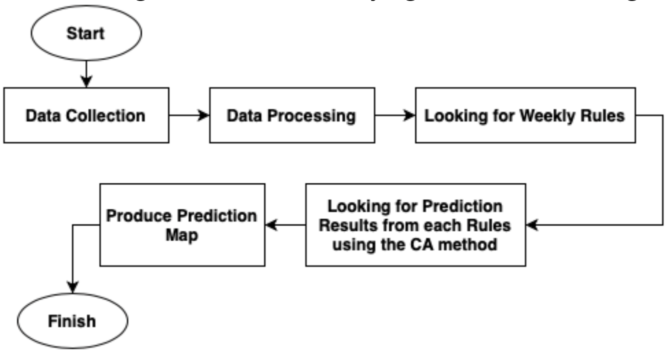

Figure 2. Research Flowchart

At this visualization stage, visualization of data that has been processed is carried out in order to make it easier to find out the spread of COVID-19 cases in Indonesia, data on COVID-19 cases that have occurred are visualized on a map using the QGIS application, on this map will show a color change with each increment. the number of cases of COVID-19, the more cases, the color will be red, if there are still a few cases, the color will be shown is pink to white. Following are the results of the spread of COVID-19 cases for 7 weeks that were visualized using the QGIS application.

In the first week, from March 3, 2020 to March 9, 2020, it was explained that the total number of COVID-19 cases did not reach 500 COVID-19 cases per week, therefore the color on the map was pink.

Figure 3. Map of The First Week (3 - 9 March 2020)

Justlikethepreviousweek, in thesecondweek, from March 10, 2020 to March 16, 2020, itwasexplainedthatthe total numberof COVID-19 casesdid not reach 500 cases per week, thereforethecoloronthe map ispink.

Figure 4. Map of The 2nd Week (10 – 16 March 2020)

To find out whether the results of the application calculations are appropriate or not, researchers can conduct trials by doing manual calculations and comparing the results between the manual and the application results. The data used by researchers on the Covid-19 case is the latest active case data. Then to predict, researchers used 16 weeks before the week they wanted to predict. The following is an example of manual calculations from week 28 in 2021, namely on July 12 to July 18, 2021, researchers use data on new active cases on March 22, 2021-16 May 2021 for the X coefficient and data on new active cases on 17 May – 11 July 2021 as the Y coefficient. Then the data is used as a matrix and then entered into the formula [4]:

$$\beta = (X'X)^{-1}X'Y \quad (3)$$

Where:
B = Rules
X = Covid-19 case data 16-9 weeks before prediction week
X' = data in the X matrix transposed,
$(X'X)^{-1}$= the product of X multiplied by X' then inverse
Y = Covid-19 case data 10-1 week before prediction week.
Which will be described as follows:

TABLE I. DATA MATRIX X ON WEEK 12-19 IN 2021

H/M	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*
1	17	1	10	5	-89	31	41	-24
2	6	9	-19	-7	-67	60	84	66
3	12	28	18	5	-52	-40	67	93
4	1	10	34	-62	-20	23	68	3
5	-1	3	21	-99	80	-6	-5	-16
6	13	0	12	-79	-66	3	91	43
7	-7	2	7	-58	-89	8	-71	-41

TABLE II. DATA MATRIX X' ON WEEK 12-19 IN 2021

H/M	*1*	*2*	*3*	*4*	*5*	*6*	*7*
20	17	6	12	1	-1	13	-7
21	1	9	28	10	3	0	2
22	10	-19	18	34	21	12	7
23	5	-7	5	-62	-99	-79	-58
24	-89	-67	-52	-20	80	-66	-89
25	31	60	-40	23	-6	3	8
26	41	84	67	68	-5	91	-71
27	-24	66	93	3	-16	43	-41

After getting the matrix X (in TableI) and X' (in TableII), the matrix is multiplied andwill form a matrix-like matrix below:

TABLE III. DATA MATRIX X MULTIPLIED BY X' THEN INVERSE

689	400	392	-481	-2874	419	3758	1969
400	979	760	-951	-2286	-321	3196	3074
392	760	2575	-5268	-968	-802	2822	175
-481	-951	-5268	23349	3460	-1998	-6840	262
-2874	-2286	-968	3460	34191	-6549	-14208	-7651
419	-321	-802	-1998	-6549	6799	4930	-538
3758	3196	2822	-6840	-14208	4930	31197	17899
1969	3074	175	262	-7651	-538	17899	17376

After that, determine the Y data be used as a matrix. Data Y used in this study was week 20 to 27 in the province of Aceh, the data are as follows:

TABLE IV. DATA MATRIX Y ON WEEK20-27 IN 2021

H/M	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
1	64	-98	-31	-67	113	-165	60	6
2	50	25	1	84	-134	-26	36	100
3	76	225	98	97	74	-74	-29	78
4	194	95	159	204	48	-44	35	105
5	65	107	161	190	81	40	-173	-41
6	120	160	81	-95	31	-15	-143	0
7	6	150	55	115	70	15	4	0

After that, multiply the results of the Y matrix by the X' matrix like the matrix below:

TABLE V. DATA MATRIX Y MULTIPLY X'

3947	2202	1321	-1497	1885	-4233	-791	1784
4789	7998	4905	6245	1842	-2761	-589	4017
10501	11042	11579	10071	9203	-3219	-4737	2273
-27941	-37363	-35058	-31061	-15631	-1930	25925	-2731
-20132	-21903	-2945	2446	-7683	24010	-11702	-16670
6424	-7315	-1447	3270	-6226	-4892	6626	5727
35277	22992	18852	6920	-955	-19529	-6511	21217
13288	24230	9053	4945	-7363	-6670	-5201	14681

Then to get the value of the rules as the result of multiplying

$(X.X')^{-1}$ with (Y.X') multiplied, later these results will produce values a to h in the Moore schema formula. The results of the rules are as below:

TABLE VI. DATARESULTS MATRIX RULES (B)

-150.5
103
17
5.25
-30,625
-6
19.5
-19

After getting the rules, then these rules are entered into the calculation of the Moore scheme to get the prediction results for the 28th week of 2021 in Aceh Province, in the following way [4]:

$$S_{ij}(t+1) = (a.S_{i-ij}(t) + b.S_{i-1j}(t) + c.S_{i-1j}(t) + d.S_{ij-i}(t) + e.S_{ij}(t) + f.S_{ij}(t) + g.S_{i+1j-1}(t) + h.S_{i+j}(t) + k.S_{i+1j+1}(t) \quad (4)$$

Where:
S_ij (t+1) = Prediction of the next week in one province
S_(i-ij) (t) to S_(i+1j+1) (t) = Active case data from Monday to Sunday of the previous week at the same time and in the same province.
After getting the rules, the rules are then multiplied by the total data of active cases in week 27. The total data for week 27 cases is 3737. The total active cases are multiplied by the matrix results from the rules previously calculated using the Moore schema formula above. This results in a total of 3703 active case predictions at week 28.
After knowing the prediction results for the 28th week of 2021, the researcher can make the accuracy level of the prediction results by using MAPE, with the formula as below [9]:

$$MAPE = \frac{1}{n}\sum \frac{(|Actual - Forecast|)X\ 100}{Actual} \quad (5)$$

IV. RESULT AND DISCUSSION

The following is a table of data comparisons from predicted data with real data from week 28 and calculations from (y_i-y_1)/y:

TABLE VII. PREDICTED VALUE

Province	*Total Real Data of New Active Cases*	*Total New Active Case Prediction Data*	$\frac{y_i - \hat{y}_1}{y}$
Aceh	-42	3703	8916.666667 %
Bali	2851	6278	120.2034374 %
Banten	20546	10069	50.99289399 %
Bengkulu	1118	2100	87.83542039 %
Jakarta	20054	100053	398.9179216%
Jogja	5888	16316	177.1059783%
Jambi	1584	1733	9.406565657%
Jawa Barat	24859	85677	244.6518364%
Jawa Tengah	9803	49080	400.6630623%
Jawa Timur	27420	15774	42.4726477%
Kalimantan Barat	2387	2027	15.0816925%
Kalimantan Selatan	1823	1311	28.08557323%
Kalimantan Tengah	1371	6170	350.0364697%
Kalimantan Timur	6426	7895	22.86025521%
Kalimantan Utara	742	1746	135.309973%
Bangka	1617	2062	27.52009895%
Kepuluauan Riau	2009	6065	201.8914883%
Lampung	1445	4176	188.9965398%
Maluku	927	2432	162.3516721%
Maluku Utara	648	1734	167.5925926%

NTB	358	1510	321.7877095%
NTT	3704	5052	36.39308855%
Papua	1627	9624	491.5181315%
Papua Barat	734	3059	316.7574932 %
Riau	2355	4919	108.8747346%
Sulawesi Barat	331	529	59.81873112%
Sulawesi Selatan	2135	3588	68.05620609%
Sulawesi Tengah	1165	1258	7.982832618%
Sulawesi Tenggara	184	1783	869.0217391%
Sulawesi Utara	1283	1327	3.429462198%
Sumatera Barat	2269	5576	145.7470251%
Sumatera Selatan	2983	2490	16.52698626 %

Based on the test results, it was found that the predicted values from several regions were still very different. In table VII it can be seen that there are differences in the values that reach more than 400% error rate. And if you calculate the MAPE value obtained, it is 431.1466353%.This value is still quite a large error, when compared with other research that uses Cellular Automata [10]. However, if analyzed further, this error rate could be caused by the use of very different data. In the research of Hanafi, et al [10] the data is land cover data which does not have significant daily changes, in contrast to this research which uses Covid-19 data which tends to change the incidence value and has very different values in each area.Therefore, it can be concluded that the cellular automata method for mapping the spread pattern of the Covid-19 disease is not suitable to be applied because the error rate is still quite high.

V. Conclusion

Based on this research, it can be concluded that the results of the prediction accuracy test using Moore's neighbor get an error rate using MAPE of 431.1466353%. This error value is obtained quite high because there are several things that cannot be predicted by cellular automata but become one of the significant determinants for showing the number of cases of covid-19, these factors are influenced by several cases such as the decision on social restrictions by the government so that it can reduce the number of cases. covid so that the number of covid cases did not match the prediction calculations to be wrong and the presence of other factors such as flights from abroad and the existence of positive data that was not recorded, the implementation of PPKM became one of the factors that made the number of cases of this covid decrease. Therefore, the cellular automata method using a system of linear regression equations is not suitable for use in research using COVID-19 data because the data is always changing drastically

References

[1] KEMENKES, "Hindari Lansia dari Covid-19," *Kementerian Kesehatan Republik Indonesia*, 2020. http://www.padk.kemkes.go.id/article/read/2020/04/23/21/hindari-lansia-dari-covid-19.html (accessed Oct. 23, 2020).

[2] S. Fadjarajani, "Peranan Geografi dalam Analisis Sebaran Covid-19," *Pros. Semin. Nas. Hardiknas 1*, vol. 1, pp. 71–78, 2020.

[3] D. M. U. Atmaja, A. R. Hakim, D. Haryadi, and N. Suwaryo, "Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Pengelompokkan Tingkat Risiko Penyebaran Covid-19 Jawa Barat," *SNTEM*, vol. 1, no. November, pp. 1218–1226, 2021, doi: 10.53026/sntem.v1i2.591.

[4] L. Muslimah and D. Pratiwi, "Outspread Pattern Simulation System Of Septicaemia Epizootica Disease Using Cellullar Automata Method In Nusa Tenggara Timur," *Int. J. Technol. Bus. Accord.*, vol. 2, no. 2, pp. 127–138, 2018.

[5] O. Hasan, S. Sentinuwo, and A. Sambul, "Pemodelan dan Simulasi Penyebaran Penyakit Demam Berdarah Dengue (DBD) dengan Menggunakan Model Cellular Automata," *J. Tek. Inform.*, vol. 10, no. 1, 2017, doi: 10.35793/jti.10.1.2017.15636.

[6] N. A. Pratomoatmojo, "Permodelan Perubahan Penggunaan Lahan Berbasis Cellular Automata dan Sistem Informasi Geografis dengan Menggunakan LanduseSim," *J. Penataan Ruang*, vol. 13, no. 1, p. 26, 2018, doi: 10.12962/j2716179x.v13i1.7064.

[7] A. Shalihah, F. T. Elektro, U. Telkom, and M. Matematika, "Pemodelan Dan Simulasi Penyebaran Penyakit Demam Berdarah Dengue (Dbd) Di Kota Bandung Menggunakan Cellular Automata Modeling and Simulation of Spread Dengue Fever Disease (Dhf)," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 6259–6266, 2018.

[8] E. Nurhidayati, I. Buchori, and M. Mussadun, "Cellular Automata Modeling in the Built Up Areas Within Urban Development At Pontianak, West Borneo, Indonesia," *Geoplanning J. Geomatics Plan.*, vol. 4, no. 2, p. 201, 2017, doi: 10.14710/geoplanning.4.2.201-212.

[9] A. Kumar, R. Agrawal, and C. Chattopadhyay, "Weather based forecast models for diseases in mustard crop," *Mausam*, vol. 64, no. 4, pp. 663–670, 2013, doi: 10.54302/MAUSAM.V64I4.749.

[10] F. Hanafi, D. P. Rahmadewi, and F. Setiawan, "Land Cover Changes Based on Cellular Automata for Land Surface Temperature in Semarang Regency," *Geosfera Indones.*, vol. 6, no. 3, p. 301, 2021, doi: 10.19184/geosi.v6i3.23471.