Applied Information System and Management (AISM)
Volume 7, (1) 2024, p. 9-16
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v7i1.34088

Correlation Analysis and Prediction of Confirmed Cases of Covid 19 and Meteorological Factor

Ni Wayan Priscila Yuni Praditya[1], Annisa Khodista Syaka[2], Regina Anggraini[3]

***Abstract*—The spread and resilience of Covid 19 in an environment depend on meteorological factors. The relationship between the covid case and meteorological factors needs to be examined more deeply. This study aims to determine the relationship between confirmed cases of Covid 19 and meteorological factors, namely temperature and humidity levels. The data used in this study are the number of confirmed cases of Covid 19, the average temperature, and the average humidity level in five states in the United States. Data were obtained from 22 January - 30 September 2020. This study used the Pearson and Spearman Correlation Analysis to find an effect from temperature and humidity levels to increase the number of confirmed cases of Covid 19. The Business Intelligence approach with LSTM is also carried out by predicting a multivariate time series in confirmed cases of Covid 19. Based on the results of Pearson Correlation and Spearman Correlation, it is stated that humidity and temperature have a correlation that affects the spread of Covid 19. The use of multivariate time-series can predict cases of confirmed Covid with meteorological factors such as temperature and humidity levels. The prediction results show that an increase in Covid-19 cases in the States of California, Texas, Florida, Illinois, and Georgia can still occur.**

***Index Terms*—COVID 19, humidity, temperature, meteorology, pearson, spearman, business intelligence, LSTM, prediction.**

I. Introduction

Corona virus disease (COVID-19) is a disease caused by the SARS-CoV-2 virus. Covid cases were first reported in December 2019 in Wuhan, China [1]. The spread of the Corona virus is increasingly widespread, until finally the WHO (World Health Organization) decided establish Covid-19 as a world pandemic. The number of confirmed cases of Covid is increasing worldwide. WHO reports that the number of confirmed cases of Covid as of September 31, 2020 has been more than 33 million cases worldwide [2].

The development of confirmed cases of Covid 19 is increasingly widespread and increasing throughout the world. This cannot be separated from the uncontrolled spread of the Covid virus, but now, namely entering 2023, the government stated that the Covid 19 virus which was previously declared a pandemic has turned into endemic, which means a situation where something permanently exists in society and is found somewhere. particular place or population. Even so, the Covid virus still exists and lives side by side with the community and you still need to be vigilant because the Covid virus continues to develop and has various types.

The spread of Covid-19 occurs mostly through droplets of saliva or fluids from nose when people who are infected with this virus cough or sneeze. Covid-19 virus infection can occur when the virus enters through the mouth, nose or eyes. This transmission can occur when there is close contact (less than 1 meter) with a person with Covid-19 [3].

The transmission of Covid-19 apart from close contact with sufferers in through aerosols [4]. Some things that increase the risk of transmission through aerasols are crowded spaces with inadequate ventilation. Liquit droplets that are spread by someone with Covid symptoms in an environment can pose a greater risk of transmission, this is because the liquid will stick to surrounding objects and even be left in a room with poor air circulation.

The spread and resistance of viruses in an environment depends on several aspects, one of which is meteorogical factors. Several previous state [5], show that there are studies that temperature and humidity levels affect the spread and resistence of the 229E, SARS-CoV, and MERS-CoV coronaviruses. This is because temperature and humidity levels affect the evaporation of droplets released by Covid sufferers. In addition, the level of effectiveness of a person's respiratory system in fighting infections caused by viruses is also influenced by these meteorogical factors.

Based on previous research regarding the relationship between Covid cases and meteorological factors [5]. This study

Received: 29 July 2023; Revised: 8 August 2023; Accepted: 19 September 2023
*Corresponding author
[1]N. W. P. Y. Praditya, Faculty of Computer Science, Universitas Indo Global Mandiri (e-mail: niwayanpris@uigm.ac.id).
[2]A. K. Syaka, Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada (e-mail: anissasyaka@ugm.ac.id).
[3]R. Anggraini, Faculty of Computer Science, Universitas Indo Global Mandiri (e-mail: regina@uigm.ac.id).

proposes to examine more deeply about these two aspects. The meteorological factors that will be used in this study are temperature and humidity, this is because the temperature and humidity in each region are different. The process of looking for the relationship between the two variables required adequate data. The research process uses daily data of confirmed Covid cases, average temperature, and average humidity levels in five states in the United States, namely California, Florida, Illinois, Georgia, and Texas. Data collected from 22 January 2020 to 30 September 2020 was obtained from the STC Center [6].

This study aims to determine the relationship between confirmed cases of Covid-19 and meteorological factors, namely temperature and humidity levels. In this study, we adopted two analysis method, namely the pearson correlation and the spearmen correlation so that the results of the analysis can be compared and produce more accurate data. The pearson correlation itself functions to measure the linear relationship between two variables which describes the direction and degrre of one variable that is linearly related to other variables, while the spearmen correlation is used to test the significance of the associative hypothesis if there are ordinal-shaped variables linked [7].

After the correlation analysis was carried out, we conducted an experiment with a business inteligence approach. The approach taken with business intelligence in this study adopts artificial intelligence data processing and modeling methods. This research uses implementation deep learning in data time series. We applied a multivariate time series with the LSTM model to predict confirmed cases of Covid using data on confirmed cases, average temperature, and average humidity levels.

This research is expected to help formulate: (a) Meteorological factors that have a correlation with the number of cases, (b) Identify the results of the pearson correlation analysis, (c) Identify the results of the spearman correlation analysis, and (d) Predict cases of Covid confirmation using the model LSTM in the data multivariate time series.

This paper consists of several compilations, namely: Part II describes the literature review which is used as a basis for literature studies of previous research. Part III is the research method used. Part IV describes the analysis of the experiments that have been carried out in this study. The last is part V which contains brief conclutions regarding the research results that have been obtained by including proposal for futher research development. All literature used in this research in mentioned in the Bibliography Section.

II. Related Work

Analysis of spreading cases of Covid19 against temperature and humidity has been carried out by several previous studies. Some methods used for this case are SEIR (Susceptible Exposed Infectious Recovered), statistical analysis, and approaches using artificial intelligence models. The following is a summary of the relationship between the Covid 19 case and several factors causing its spread with several models or methods used by previous researchers.

Research by adapting the SEIR model is [8]. The SEIR Model can be used to estimate the extent of spreading Covid 19 cases distribution and the effective number. Then a discussion can be held to measure the effect of the spread of Covid 19 in different countries when the weather in northern areas, namely Cina, Singapore, and Hong Kong starts to rise.

Apart from using The SEIR Model, statistical models can also be used in analyzing cases of the spread Covid 19 against temperature and humidity levels. Research [8] used statistical analysis in the form of correlation analysis to identify factors that influence the spread of Covid 19 on several meteorological factors, population, and population density of an area. The correlation analysis applied can obtain variables that have a correlation with the spread of Covid 19..

Similar to the previous study [8], study in [9] also used correlation analysis to determine the relationship between two variables. The application of correlation analysis in this study uses environmental factors and confirmed cases of Covid 19. The results of the correlation analysis conducted showed that the average humidity and maximum temperature had the highest impact on confirmed cases of Covid 19.

SIR and non-SIR approaches based on Gauss-error-function and Monte Carlo simulations cannot predict the spread of Covid 19 in India. Therefore Ghosh [10] estimates a mathematical model to predict the spread of Covid 19 in India. The mathematical model used is based on the 'change-factor' or 'rate-of-change'. The data used is using Covid 19 data in India until July 16, 2020. This research has obtained accurate results of 90.36% since the first discovery of Covid 19 in India and can be added by 96.67% after passing April.

After that, there is also research [11] to analyze about the correlation between the weather and the increase in Covid 19 in Jakarta. The data used are sourced from Covid 19 data in Jakarta from January 2020 to March 29, 2020. The analysis in this study uses the Spearman rank to see the effect of the weather on the increase in Covid 19 every day. The weather calculation in this study uses five parameters, namely the average maximum temperature, minimum temperature, average temperature, and the level of rainfall. After that, the data from the five parameters were analyzed using the Spearman rank to find out what parameters have a relationship in influencing the rate of increase in Covid 19 in Jakarta. The result shows that the average temperature in Jakarta has the highest correlation among other parameters.

On the other hand, there is [12] discussed the spread of Covid 19 in Brazil with the influence of meteorological conditions from aircraft traffic and population density in 27 cities in Brazil during 1 month affected by the outbreak. The data used were Covid 19 data from 26 February-26 March 2020, population density based on the Brazilian Institute for Geography and Statistics (IBGE), flight traffic data based on the Brazilian Agency for Civil Aviation (ANAC), and meteorological data based on the Brazilian Institute for Meteorology (INMET). The prediction method in this study uses the Generalized Linear Mixed Model (GLMM). The

Applied Information System and Management (AISM)
Volume 7, (1) 2024, p. 9-16
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v7i1.34088

results of this study indicate that temperature is one of the factors affecting the rate of increase in Covid 19. The best estimate in this study is that when the temperature increases by 1 oC it can reduce as much as 8% the rate of increase in confirmed cases of Covid 19.

The method with the artificial intelligence approach has been carried out by several previous studies [8], [9], [13], [14]. The artificial intelligence approach used in research related to Covid 19 is classified into three types, namely artificial neural network [13], machine learning [9], and deep learning [8]. The Artificial Neural Network was used by [13]. Analysis of the role of the specific environmental parameters to the Covid 19 pandemic was carried out by applying the ANN-PSO (Particle Swarm Optimization) method and the DE (Different Evolution) algorithm. This study proves that the ANN-PSO Models has better performance, so that this study is able to predict confirmed cases of Covid 19 with two main environmental variables.

Another artificial intelligence method, namely the fuzzy method can also be used in this case. Research [14] points to the factors that influence the spread of Covid 19, among climatic and atmospheric conditions. The Fuzzy method used is the Fuzzy Mamdani method. This method using fuzzy linguistic variables and parameters used in the system. The experiments conducted were able to show that temperature had a greater effect on the number of people infected with Covid 19 compared to atmospheric conditions in the form of UVI levels.

The application of the machine learning model in the Covid 19 case has been carried out by [9]. This study uses a binary classification model to predict confirmed cases of Covid 19. The results of this study indicate that the binary classification model can be used to predict confirmed cases with good performance. In addition to classifying, this study applies regression analysis to compare trends in confirmed cases with increases in daily weather parameters.

Mechanical learning algorithms can also be applied as an algorithm to predict the influence of weather on the dissemination of Covid 19. Regression can be used for extracting the correlation between various factors and the level of spread of Covid 19 while mechanical learning algorithms are used to estimate the impact of weather variables such as temperature and humidity on Covid 19 transmission by extracting the relationship between the number of confirmed cases and the variable of weather in a certain area. After implementation, the results show that temperature and humidity are the important variables in predicting mortality rate, and it is indicated that the higher the temperature value, the lower the number of infection cases [15].

The deep learning model is also used by several studies on the Covid 19 case. One of them is a study [8] which uses the LSTM Model to predict confirmed cases and deaths due to Covid 19. The implementation is carried out on univariate time-series data. The predictions made with these models have successfully supported the hypothesis of self-isolation of social distancing as the main criterion against the spread of Covid 19.

The Bi-LSTM method has also been applied in decision making to stop the spread of the Covid 19 pandemic. The data used are categorized by utilizing the K-Means clustering algorithm to identify groups of countries that have socio-economic properties and the health sector of countries in the world. By utilizing Bi-LSTM and lockdown information in forecast data, the proposed approach achieves a significant increase in the performance [16].

The LSTM method can also be combined with SEIRD (Susceptible Exposed Infectious Recovered Dead) and the GWR (Geographically weighted regression) model to see the suitability and comparison in predicting Covid 19 occurring in China. This study uses data from January to March 2020 sourced from the National Health Commission of China. The results of this study, the SEIRD model has an error proportion of less than 5% in each province. Meanwhile, LSTM had an error proportion of less than 5.2% on February 3 and less than 0.64% on February 14. The last model is GWR, an experiment on GWR using data on February 2 and February 3. The error percentage of the experiment has different results from February 2 and February 3, such as Beijing has 11.67% and 3.95%, Shanghai has 2.24% and -5.88%, Xiaogan has -1.27% and 1, 70, whereas in Wuhan it has 0% and 14.57% [17].

In addition, there are also those who use LSTM, Gated recurrent unit (GRU), and Bi-LSTM which are combined into a deep learning model to predict Covid 19. Three categories must be predicted, namely, confirmed cases, patients who died, and patients who recovered. The data used were confirmed cases until June 27, 2020. In this study, 110 days of Covid 19 data were used as training samples and 48 days to be tested from 10 countries. The results sought are based on the best values of the mean absolute error (MAE), root means square error (RMSE), and the value of r2. The best results were MAE 0.0070 and RMSE 0.0077 for the Chinese mortality rate and an r2 value of 0.9997 on the cure rate for patients in China [18].

Other studies that aimed to determine the correlation between climatic factors such as air temperature, humidity, wind speed, sunlight, and the dissemination of Covid 19 in geographic physical areas has utilized an algorithm that serves to identify the climate and bioclimatic determinants of the Covid 19 case, namely the Random Forest algorithm. This study also utilized the final map created by importing the final R output in ArcGIS to generate a Covid 19 climate suitability map. The results of this study indicated that the effect of temperature and sunlight has a significant negative relationship between the intensity of Covid 19 while temperature and wind speed becomes the most significant effect. It was concluded in this study that the Covid 19 cases and mortality rates were affected by temperature fluctuations [19].

In predicting the upcoming number of Covid 19 case, various techniques and methods can be applied. RNN and the

VMD algorithm have also been widely used. In this study the RNN was created to predict the number of signals of future confirmed cases based on the previous mortality rate while the VMD algorithm aims to break down non-stationary non-linear signals into a structive mode. Transmission speed and meteorological parameters (temperature and humidity) are used as features to train a series of repetitive neural networks (RNNs) to estimate the number of cases in the future. The proposed procedure can provide insight into the systematic estimation of the number of Covid 19 cases in the future by considering other factors that affect it in society [20].

Other researches that aimed to explore the effect of meteorological parameters on Covid 19 mortality rate used the general additive model (GAM). Data obtained from the Shanghai Meteorological Bureau and the data center of the Ministry of Ecology and Environment of the People's Republic of China were data on average daily temperature, journal range temperature (DTR) and relative humidity. The absolute humidity calculation was measured by vapor pressure (VP) while the GAM statistical method was used to analyze the relationship between meteorological factors (temperature, DTR, relative humidity, and absolute humidity) and the number of daily deaths from COVID-19. The results showed that 19 daily Covid deaths were positive over DTR, but negative on absolute humidity. Temperature and humidity can also be important factors in the spread of Covid 19 [21].

III. Research Method

This study uses three analytical approaches, namely: correlation analysis using two methods, namely Pearson Correlation Analysis and Spearman Correlation Analysis. Correlation analysis was used to identify the relationship between confirmed cases of Covid 19 and meteorological factors, such as average temperature and average humidity levels. The final analysis used in this study uses artificial intelligence approach, namely LSTM. The development of prediction of confirmed cases of Covid 19 was made with LSTM using multivariate time-series data.

The methodology in this research consists of data collection, data processing, business intelligence, Pearson correlation analysis, Spearman correlation analysis, and LSTM modelling (Fig. 1).

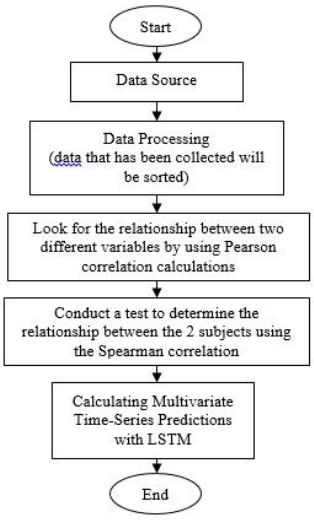


Fig. 1. Research Flow

A. Data Source

This study uses data obtained from the STC Center [6] to perform correlation analysis and predictive modeling using LSTM. The data used are the number of confirmed cases of Covid 19, the average temperature, and the average humidity level. The time span of the data used was 22 January 2020 to 30 September 2020. The data provided by the STC Center [6] consists of several countries in the world, we use data from five states of the United States of America. Using data from the United States because of The very high number of confirmed cases of Covid 19 in the United States. The five states selected were also the states with the highest number of cases among other the United States, namely Georgia, Illinois, Florida, California, and Texas.

B. Data Processing

The data that has been collected is then processed first before being used in this study. The data collected will be labeled according to the state and daily date. Downloading data from the STC Center does not have lost data, however the data is still in the several different folders. We use data processing techniques by combining data that was previously separated in each state, into a single dataset in a csv file.

This study uses SPSS in Pearson and Spearman Correlation Analysis. Prediction modelling of confirmed case data for Covid 19 uses “Python 3.x” for data processing of this research. We also use Google Collab environment to develop python-based deep learning applications. The library used is “Keras” with “TensorFlow” as a backend in the development of the LSTM model. The prediction results are then displayed with the python library, namely “matplotlib”.

C. Business Intelligence

Data mining is one of the techniques used in Business Intelligence or artificial intelligence that is able to classify and cluster data based on the nature and correlation of the data set used. The methods commonly used in data mining are C45, K-Means, Apriori, KNN, LSTM, Naive Bayesian, and many other [22].

Applied Information System and Management (AISM)
Volume 7, (1) 2024, p. 9-16
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v7i1.34088

This study used the LSTM method and the classification of The Pearson correlation and The Spearman correlation to determine the relationship between humidity, temperature, and the Covid 19 cases in various US states.

D. *Pearson Correlation*

Pearson Correlation is a form of formula used to find the correlation between two different variables, the independent variable and the dependent variable [7]. The basis of decision making in the Pearson correlation is: [23]

- Rule of decision making

Pearson linear correlation value ±0,1 indicates small effect.
Pearson linear correlation value ±0,3 indicates medium effect.
Pearson linear correlation value ±0,5 indicates large effect).

This correlation test aims to determine the level of accuracy of the correlation between variables expressed by the correlation coefficients, and the type of relationship between variables is usually determined by Y and X which can be positive or negative [23].

E. *Spearman Correlation*

Spearman Correlation to find the relationship between two or more subject data with Ordinal scale variables. In general, basic to make a decision and guideline to know the correlation level on the Spearman Correlation is the same as the Pearson Correlation.

The data used is in the form of average data each month on the average temperature and humidity data every day. The data is analyzed to see its attachment to the confirmed Covid 19 cases every day.

F. *LSTM (Long Short Term Memory)*

LSTM is a deep learning model with a distinctive RNN (Recurrent Neural Network) structure. LSTM has a chain-like structure, which consists of (cell, input gate, output gate, and forget get). What distinguishes the LSTM model from the standard feedforward neural network model is that the training model has a different structure, namely a feedback connection [8]. These elements, called gates, control the incoming information on each LSTM unit. The application of this gate can make the learning training process more effective because it can forget learning which is considered to have no effect on the model.

The problems that occur in the RNN model in the form of exploding gradient and vanishing gradient can be solved by the LSTM model with a structure that utilizes the forget gate. This LSTM model can be used on data that has long-term dependence, such as time series prediction, handwriting recognition, voice detection, and anomaly detection in network traffic. Several types of LSTM models are univariate, multivariate, multi-step, and multivariate multi-step) which can be used in several types of time series data prediction problems [8].

The LSTM model has an effective structure on data that has long-term dependence, such as time series. The forget gate structure can solve the problems caused by the RNN model. The data used in this study is time series data which has a dependency from 22 January 2020 to 30 September 2020. The multivariate LSTM model can also be used for prediction on several variables, namely confirmed cases of Covid 19, average temperature, and average humidity. Based on this, the LSTM model was chosen by us as a model for predicting multivariate time series data.

This study uses the LSTM model to multivariate time series. The multivariate type is used to show the relationship between existing data, so that if applied to this study it can show confirmed case data with temperature and humidity levels.

IV. Result and Discussion

A. *Implementation Pearson Correlation*

In this study, Pearson Correlation was used to determine the relationship between humidity and temperature in five countries with the highest rates of Covid-19 cases, which were California, Texas, Florida, Illinois and Georgia in the US [24]. Figure 2 illustrates a graph of the results of the Pearson correlation which has been made based on each country.

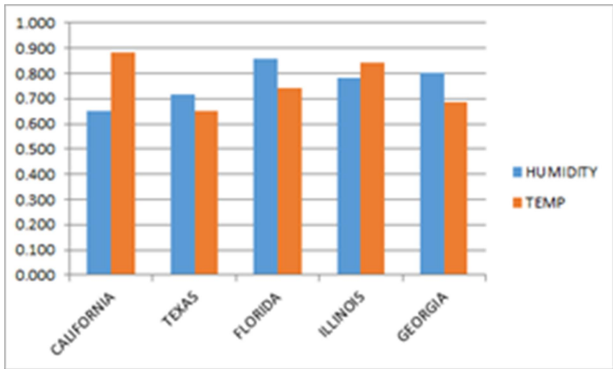


Fig. 2. Pearson correlation graph.

Based on the picture above, it can be seen that there are three countries (Texas, Florida and Georgia) where the spread of Covid-19 cases based on humidity is higher than based on temperature while the other two countries (California and Illinois) have a higher spread of Covid-19 cases based on temperature than based on humidity.

The correlation between humidity and temperature in the spread of the Covid-19 case in five countries based on the results of the Pearson Correlation calculation has the following values:

- California, the value of the correlation between humidity and Covid-19 case was 0.651, and the value of the correlation between temperature and the Covid-10 case was 0.881.
- Texas, the value of the correlation between humidity and Covid-19 case was 0.716, and the value of the correlation between temperature and the Covid-10 case 0.653.

- Florida, the value of the correlation between humidity and Covid-19 case was 0.856, and the value of the correlation between temperature and the Covid-10 case 0.741.
- Illinois, the value of the correlation between humidity and Covid-19 case was 0.780, and the value of the correlation between temperature and the Covid-10 case 0.845.
- Georgia, the value of the correlation between humidity and Covid-19 case was 0.802, and the value of the correlation between temperature and the Covid-10 case 0.685.

B. Implementation Spearman Correlation

At this stage using the Spearman Correlation to find the correlation temperature and humidity level for the confirmed Covid 19 case. Based on the highest Covid 19 cases from 5 US states, namely Georgia, Illinois, Florida, California, and Texas. The results of the Spearman Correlation can be seen in Fig. 2.

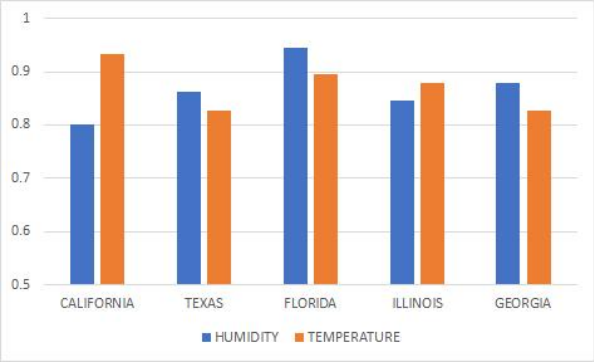


Fig. 3. Spearman correlation graph.

The results of the experiment using SPSS found that there was a correlation of the Covid 19 data which was confirmed by temperature and humidity. The correlation coefficient value of the number of confirmed Covid 19 cases with humidity and temperature in each state exceeds 0.80. Therefore, the level of correlation of the number of confirmed Covid 19 cases with humidity and temperature has "large effect".

C. Multivariate Time-Series Prediction with LSTM

The business intelligence approach used in this study is the implementation of the LSTM model. The model is used to predict confirmed cases of Covid 19 using a dataset consisting of date, state name, state code, total daily confirmed cases, average daily temperature, and average humidity level. The data consist of five states in the United States, namely California, Texas, Florida, Illinois, Georgia.

We performed a predictive analysis using the LSTM model. The design of the model is done so that it can use multivariate time series data. We process daily data from 22 January 2020 to 30 September 2020. Each county will be processed in turn on LSTM modeling. The stages in predicting a confirmed case of Covid 19 in each State of the United States are as follows:

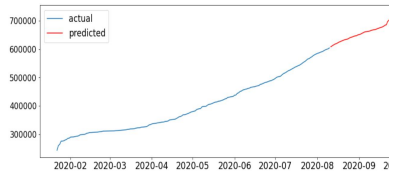
- The data were divided into two groups, train data and test data with a ratio of 80:20. Data sharing cannot be done randomly because we will apply time series analysis.
- Data is converted into dependent and independent variables. This study uses the variable case confirmed by Covid 19 as the dependent variable and other variables such as average temperature and average humidity level as independent variables.
- The data is then converted into the appropriate dimensions so that it can be assigned to the LSTM network.
- Declare a callback, namely: Check Point Model, Early stopping, and Reduce learning rate. Callback is used so that the training process will be more efficient later.
- After the data has been processed, the model training process can be carried out. We used the Keras API in the Tensorflow library to build this model. The architectural details of the model are:
  a. The Hidden nodes uses 250 layer
  b. The Dropout used is 50% and 20%
  c. Activation function using Relu
  d. Accuracy using Mean Absolute Error
  e. Loss function using Huber
  f. Optimization function using Adam
- The following are the results of the training carried out:

Table 1.
Results of LSTM Model Training

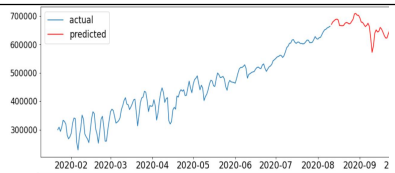
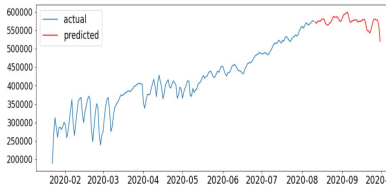
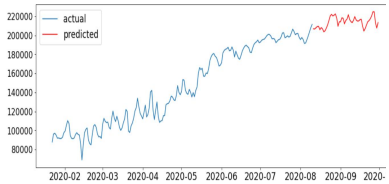
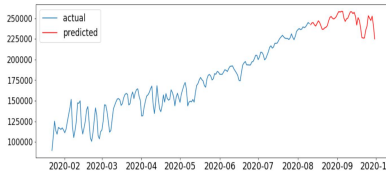
States	MAE	Val_MAE	Loss	Val_Loss
California	0.1052	0.0522	0.0083	0.0016
Texas	0.1767	0.1539	0.0242	0.0205
Florida	0.1725	0.0689	0.0232	0.0037
Illinois	0.1786	0.1677	0.0228	0.0219
Georgia	0.1799	0.1090	0.0249	0.0109

Based on Table 1, it can be seen that the performance of the training process carried out on the train data and test data. Training is carried out at 100 epochs using early stopping, so that the training process will stop automatically when there is no increase in performance. The training process in all states is performing well. The training process in all states is performing well. This was obtained based on the MAE value used was less than 10% and the loss value had a small value. After the training process is complete, the next process is to predict and compare values. Previously, the data that has been normalized before model training will be converted back into the value it should be. Prediction value is carried out on training data and testing data. The prediction process carried out by each of the five countries. The following is the prediction results in the form of a graph for each country using the previously trained LSTM model.

Table 2.
Prediction Results of Multivariate LSTM Models

States	Graph	Explanation
California		Increased

Applied Information System and Management (AISM)
Volume 7, (1) 2024, p. 9-16
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v7i1.34088

Texas		Increased
Florida		Increased
Illinois		Increased
Georgia		Increased

Based on Table 2 above, it can be seen that the prediction of confirmed cases of Covid 19 in five states in the United States uses the Time Series Multivariate LSTM Model. Confirmed cases of Covid 19 in these five states are predicted to increase. The prediction process involves variable cases of confirmed Covid 19 on 22 January 2020 - 30 September 2020 with other meteorological factors such as data on average temperature and average humidity levels.

D. Discussion

This research was conducted to determine the relationship between meteorological factors, namely the temperature and humidity of an area with cases of contracting Covid 19. This was done so that the local community could be more vigilant and wise in responding to this phenomenon. Even though the Covid 19 virus has been declared endemic, the virus continues to grow and coexist with society.

In previous research, many have raised the Covid case as research by utilizing various types of models and methods, both used for medical needs and artificial intelligence.

Based on the results of the implementation that was carried out in this study by utilizing calculations using several models, namely Pearson correlation, Spearman correlation, and LSTM, it appears that an increase in the spread of the Covid-virus can occur in all five samples or data. still going through this meteorological factor. The data used to test this correlation uses data from five states in the United States, namely California, Texas, Florida, Illinois, and Georgia.

Theoretically, this research can add insight and study regarding the relationship between the Covid 19 case and the factors that allow the spread of the virus to become more widespread. Considering that the Covid 19 virus used to be very dangerous [2] and is still around today, this virus can coexist with humans.

As for the impact of this research, namely, academically, researchers can find out the correlation between temperature and temperature with cases related to Covid 19. Then the non-academic impact, the community can be more aware of the spread of the Covid 19 virus which can spread through droplets, and meteorological factors can also affect the spread of Covid 19.

The limitations of this study are that researchers used confirmed data from Covid 19 in 2020. Future researchers can use the latest data with the latest current cases (endemic era) by utilizing better models or methods and being able to compare them with previous studies.

V. CONCLUSION

Based on the results of research and results of calculations that have been carried out using the Pearson correlation, Spearman correlation, and LSTM, it is stated that; the results of the pearson correlation that has been implemented, it states that the relationship between humidity and cases in the five predetermined countries has a correlation that affects the spread of Covid 19. As with humidity, temperature also has a correlation with the spread of Covid cases in that country (California, Texas, Florida, Illinois, and Georgia).

After using the Spearman Correlation, it can be concluded that the relationship of the number of confirmed Covid 19 cases with humidity and temperature has a big influence. The level of correlation can be seen in Figure 2. The correlation coefficient value of the number of Covid 19 cases confirmed with an average humidity and temperature exceeds 0.80.

The business intelligence approach in this study uses the LSTM Model to predict confirmed cases of Covid 19 in five states in the United States with a multivariate time-series. Confirmed cases of COvid 19 were used as the dependent variable and data on variable and data on average temperature and average humidity level as independent data. The use of a multivariate time-series can predict confirmed cases of Covid 19 with meteorological factors such as temperature and humidity levels. The prediction results show that the increase in Covid 19 cases in the State of California, Texas, Florida, Illinois, and Georgia can still occur.

Then, a suggestion for further research is that researchers can use other factors as variables that can affect the spread of the Covid 19 virus, can also adapt models or other methods that are more accurate. In addition, further research can also further examine the current condition of the Covid 19 virus in the future.

REFERENCES

[1] WHO, "Coronavirus disease (COVID-19)," 19 Desember 2019. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answer-hub/q-a-detail/coronavirus-disease-covid-19.. [Accessed 31 September 2020].

[2] WHO, "WHO Coronavirus Disease (COVID-19) Dashboard," 31 September 2020. [Online]. Available: https://covid19.who.int/. [Accessed 31 September 2020].

[3] WHO, "Coronavirus disease (COVID-19): How is it transmitted?," 2019. [Online]. Available: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted. [Accessed 31 September 2020]

[4] WHO, "Transmission of SARS-CoV-2: implications for infection preventation precautions," 2019. [Online]. Available: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implication-for-infection-prevention-precautions. [Accessed 31 September 2020].

[5] J. Lin, W. Huang, M. Wen, D. Li, S. Ma, J. Hua, H. Hu, S. Yin, Y. Qian, P. Chen, Q. Zhang, N. Yuan and s. Sun, "Containing the spread of coronavirus disease 2019 (COVID-19): Meteorological factors and control strategies," *Science of The Total Environment,* vol. 744, pp. 1-7, 2020.

[6] STC Center, "COVID-19-Data," 30 September 2020. [Online]. Available: https://github.com/stccenter/COVID-19-Data. [Accessed 31 September 2020].

[7] S. D. Bolbaoca and L. Jantschi, "Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationship of Biologic Active Compounds," *Leonardo Journal of Sciences,* vol. *5,* no. 9, pp. 179-200, 2006.

[8] A. Chatterjee, M. W. Gerdes and S. G. Martinez, "Statistical Explorations and Uvivariate Timeseries analysis on COVID-19 Datasets to Understand the Trend of Disease Spreading and Death," *Sensors,* vol. 20, no. 11, pp. 1-27, 29 May 2020.

[9] B. Pirouz, S. S. Haghshenas, S. S. Haghshenas and P. Piro, "Investigating a Serious Challenge in the Sustainable Development Process: Analysis of Comfirmed cases of Covid-19 (New Type of Coronavirus) Through a Binary Classification Using Artificial Intelligence and Regression Analysis," *Sustainability,* vol. 12, no. 6, pp. 1-21, 20 March 2020.

[10] S. Ghosh, "Predictive model with analysis of the initial spread of COVID-19 in India," *International Journal of Medical Informatics,* vol. 143, pp. 1-4, 2020.

[11] R. Tosepu, J. Gunawan, D. S. Effendy, L. O. I. Ahmad, H. Lestari, H. Bahar and P. Asfian, "Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia," *Science of The Total Environment,* vol. 725, pp. 1-4, 2020.

[12] P. Pequeno, B. Mendel, C. Rosa, M. Bosholn, J. L. Souza, F. Baccaro, R. Barbosa and W. Magnusson, "Air transportation, population density and temperature predict the spread of COVID-19 in Brazil," *PeerJ Journal,* vol. 8, pp. 1-12, 2020.

[13] S. S. Haghshenas, B. Pirouz, S. S. Haghshenas, B. Pirouz, P. Piro, K.-S. Na, S.-E. Cho and Z. W. Geem, "Prioritizing and Analysis the Role of Climate and Urban Parameters in the Confirmed Cases of COVID-19 Based on Artificial Intelligence Applications," *International Journal of Environment Research and Public Health,* vol. 17, no. 10, pp. 1-21, 2020.

[14] M. A. Chowdhury, Q. Z. Shah, M. A. Kashem, A. Shahid and N. Akhtar, "Evaluation of the Effect of Environmental Parameters in the Spread of COVID-19: A Fuzzy Logic Approach," *Advances in Fuzzy Systems,* vol. 2020, pp.1-5, 2020.

[15] Z. Malki, E.-S. Atlam, A. E. Hassanien, G. Dagnew, M. A. Elhosseini and I. Gad, "Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches," *ELSEVIER, Chaos, Solitons and Fractals,* vol. 138, pp. 1-10, 2020.

[16] A. B. Said, A. Erradi, H. A. ALy and A. Mohamed, "Predicting COVID-19 cases using bidirectional LSTM on multivariate time series," *Springer,* vol. 28, pp. 56043-56052, 2021.

[17] F. Liu, J. Wang, J. Liu, D. Liu, J. Tong, Z. Li, D. Yu, Y. Fan, X. Bi, X. Zhang and S. Mo, "Predicting and analysis the COVID-19 epidemic in China: Based on SEIRD, LSTM and GWR models," *Plos One Journal,* vol. 15, no. 8, pp. 1-22, 2020.

[18] F. Shahid, A. Zameer and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons and Fractals,* vol. 140, pp. 1-9, 2020.

[19] M. Pramanik, P. Udmale, P. Bisht, K. Chowdhury, S. Szabo and I. Pal, "Climatic factors influence the spread of COVID-19 in Russia," *International Journal of Environmental Health Research,* vol. 32, no. 4, pp. 723-737, 2020.

[20] M. Mousavi, R. Salgotra, D. Holloway and A. H. Gandomi, "COVID-19 Time Series Forecast Using Transmission Rate and Meteorological Parameters as Features," *IEEE Computational Intelligence Magazine,* vol. 15, no. 4, pp. 35-50, 2020.

[21] Y. Ma, Y. Zhao, J. Liu, X. He, B. Wang, S. Fu, J. Yan, J. Niu, J. Zhou and B. Luo, "Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China," *Science of the Total Environment,* vol. 724, pp. 1-7, 2020.

[22] A. I. Burhanuddin, M. N. Massi, Marsuki, et al., "Merajut Asa di Tengah Pandemi COVID-19 (Pandangan Akademisi UNHAS)," Deepublish, 2020.

[23] P. U. Gio and R. E. Caraka, "Pedoman dasar Mengelola Data dengan Program Aplikasi Statistika Statcal," Inarxiv papers, 4 June 2018. [Online]. Available: https://osf.io/preprints/inarxiv/796th/. [Accessed 2 Oktober 2020]

[24] Y. Ma, Y. Zhao, J. Liu, X. He, B. Wang, S. Fu, J. Yan, J. Niu, J. Zhou and B. Luo, "Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China," *Science of the Total Environment,* vol. 724, pp. 1-7, 2020.