An Application of Markov Chain for Predicting Rainfall Data at West Java using Data Mining Approach

A Azizah¹, R WELastika¹, A Nur Falah², B N Ruchjana² and A S Abdullah³

¹Magister Student at Department of Mathematics, Universitas Padjadjaran
²Department of Mathematics, Universitas Padjadjaran
³Department of Computer Science, Universitas Padjadjaran
annisa17056@mail.unpad.ac.id

Abstract. Markov chain model is a stochastic process to determine the transition probability of a state space based on a previous state. We can use a stationary distribution of first order-Markov chain model to determine the long term probability rainfall phenomena. A rainfall data in West Java area has a big data because we can have a large rainfall data from many cities and regencies both of in spatial and time series observations. Furthermore, in this paper, we demonstrate an application of Markov chain using a Data Mining approach to get the knowledge as a pattern for description and prediction the monthly rainfall data in wet seasons December-January-February (DJF) using Knowledge Discovery in Database (KDD) method through pre-processing, data mining process and post-processing. We simulate the monthly rainfall data from the year 1981-2017 using four-state spaces: low (0), medium (1), high (2), and very high (4). The result of Markov chain shows that the probability of occurrence rainfall phenomena for four state spaces are: low (22.62%), medium (24.86%), high (25.46%), and very high is 27.05%. It means that West Java area over the long term condition will have a very high rainfall probability.

Keywords: Markov chain, Stationary Distribution, Data Mining, Rainfall.

1. Introduction
Java Island is the island with the biggest population in Indonesia and makes various development sectors centers on there. Java is also the largest island affected by climate phenomena compared to other islands in Indonesia, if we predict the rainfall data in unobserved locations, it will be obtained the rainfall data with a pattern similar to observed locations as its neighbor [1]. Rainfall is the height of rainwater that collects in a flat place, not evaporates, not penetrates, and not flow. 1 millimeter of rainfall, which means that in a solid square meter area a flat place is held up to one millimeter of water or can hold one liter or 1000 ml of water [2].

In West Java, rainfall is considered to be one of the major constraints of agriculture plans and policy decisions because of the position of West Java as one of the centers of food based on the production paddy in Karawang Regency. It supports the Sustainable Development Goals to cover social and economic development issues including poverty, hunger, health, global warming, water, etc. Rainfall is the most important part of tropics which influenced the paddy production at West Java. Based on this, it's needed a fairly accurate prediction method especially for West Java Province's rainfall because West Java is one of the regions with the largest agricultural sector [3].

The amount of rainfall that occurs at this time could have been influenced by the amount of rainfall one time before, and the amount of rainfall in the future can be affected by the current rainfall, etc. This phenomenon is a real-life example of the Markov chain event which is a method for modeling in
stochastic processes [4]. Markov chain model used to assist in estimating the changes that may occur in
the future, where the changes are represented in dynamic variables at certain times. The Markov chain
was first coined by Andrey Andreyev Markov (1856-1922)[5]. A stochastic process is said to include
the Markov chain if it fulfills the properties of Markov (Markovian property). The properties of Markov
stated that the probability of a future event, with known past events and present events, is not dependent
on past events and only depends on the present events [4, 6].

The Markov chain is generally classified into two, namely the Markov chain with discrete parameter
index and the Markov chain with continuous parameter index. The Markov chain is said to be a discrete
parameter index if the shift state occurs with a fixed discrete time interval. Whereas, the Markov chain
is said to be a continuous parameter index if the shift state occurs with a continuous time interval [7].
Rainfall data is a time series data that indicates the movement of state in a fixed discrete time interval.
Rainfall prediction in the future is needed to anticipate prevention if high rainfall intensity will occur
for a long time. Besides that, it indicates that we need to consider the other phenomena can make a
significant contribution to increasing the rainfall intensity [8]. Furthermore, in this paper was conducted
an analysis of a large database of rainfall data from 27 district/cities using the stationary distribution of
the Markov chain and then used to predict rainfall in West Java based on Data Mining approach using
KDD method.

2. Method

2.1. Stochastic Processes

A stochastic process \( \{X(t), t \in T\} \) is a collection of random variables. That is, for each \( t \) in the
index set \( T, X(t) \) is a random variable. If the time parameter \( T \) is a countable set \( T = \{0,1,2,...\} \), the process
\( \{X(n), n = 0,1,2,...\} \) is called a discrete-time stochastic process, and if \( T \) is a continuum, the process
\( \{X(t), t \geq 0\} \) is called a continuous stochastic process. For a stochastic process \( \{X(t), t \in T\} \), a set of
all values of \( X(t) \) is called a state space [9, 10].

2.2. Basic Concept of Markov Chains

A Russian mathematician, Markov, introduced the concept of a process in which a sequence or chain of
discrete states in time for which the probability of transition from one state to any given state in the next
step in the chain depends on the condition during the previous step [11]. A first order Markov chain is
a stochastic process having the property that probability for future events depends only on the present
event, in other words:

\[
Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = Pr(X_{n+1} = x | X_n = x_n)
\] (1)

For all states \( x_1, x_2, ..., x_n \) and all \( n \geq 0 \) such a stochastic process is known as a Markov chain [7, 12].

2.3. Discrete Time of Markov Chain

Suppose \( \{X(n), n = 0,1,2,...\} \) is stochastic process with discrete parameter index and state space \( i = 0,1,2,... \) unless otherwise specified. If

\[
P\{X(n+1) = j | X(0) = i_0, X(1) = i_1, ..., X(n-1) = i_{n-1}, X(n) = n\} = P\{X(n+1) = j | X(n) = 1\} = p_{ij}
\] (2)

for all \( i_0, i_1, ..., i_{n-1}, i, j \) and \( n \), then the process is called a discrete-time Markov chain, and \( p_{ij} \) is called
a transition probability. The value \( p_{ij} \) is called a (stationary) transition probability, it represents the
probability that the process will, when in state \( i \), next make a transition into state \( j \). The transition
between the states is described by the transition probability matrix, defined as:

\[
P = (p_{ij}) = \begin{pmatrix}
p_{00} & p_{01} & p_{02} & \cdots \\
p_{10} & p_{11} & p_{12} & \cdots \\
p_{20} & p_{21} & p_{22} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\] (3)

2
Since probabilities are non-negative and since the process must make a transition into some state, we have that \( p_{ij} \geq 0, i, j \geq 0 \) and \( \sum_{j=0}^{\infty} p_{ij} = 1, i, j = 0, 1, 2, ... \) [7, 13].

2.4. The n-step Transition Matrix

The one step transition probabilities \( P_{ij} \) is already defined. Now define the n-step transition probabilities \( P_{ij}^n \) to be the probability that a process in state \( i \) will be in state \( j \) after \( n \) additional transitions. Let \( A \) be an event. A convenient notation is \( P_i(A) = P(A|X_0 = i) \). For example

\[
P_i(X_1 = j) = p_{ij}.
\]

Given the initial distribution \( \lambda \), let us treat it as row vector. Then

\[
P(X_1 = j) = \sum_{i \in \Omega} \lambda_i P_i(X_1 = j) = \sum_{i \in \Omega} \lambda_i p_{ij}.
\]

Similarly,

\[
P(X_2 = j) = \sum_{i \in \Omega} P_i(X_1 = k, X_2 = j) = \sum_{i \in \Omega} p_{ik} P_{kj} = (P^2)_{ij}
\]

Thus, \( P(X_n = j) = (\delta_i P^n)_j = (P^n)_{ij} = p_{ij}^{(n)} \)

Continuing in this way,

\[
P(X_n = j) = \sum_{i_0, ..., i_{n-1} \in \Omega} \lambda_{i_0} p_{i_0 i_1} ... p_{i_{n-1} j} = (\lambda P^n)_j.
\]

Thus, \( P^{(n)} = (p_{ij}^{(n)}) \), the n-step transition matrix, is simply \( P^n \).

Also, for all \( i, j \) and \( n, m \geq 0 \), the (obvious) Chapman-Kolmogorov equations hold:

\[
P_{ij}^{(n+m)} = \sum_{k \in \Omega} P_{ik}^{(n)} P_{kj}^{(m)}
\]

Named for their independent formulation by Chapman, and Kolmogorov (1903-1987) [7].

The Chapman-Kolmogorov equation asserts that

\[
p^{(n+m)} = p^{(n)} \cdot p^{(m)}
\]

where the dot product represents matrix multiplication. Hence, in particular

\[
p^{(2)} = p^{(1+1)} = p^2
\]

and by induction

\[
p^{(n)} = p^{(n+1-1)} = p^{n-1} \cdot p = p^n
\]

That is, the n-step transition matrix may be obtained by multiplying the matrix \( p \) by itself \( n \) times [12, 14].

2.5. Stationary Distribution of Markov Chain

There exist a limiting probability that the process will be in state \( j \) after a large number of transitions, and this value is independent of the initial state. In other words, \( p_{ij}^n \) is converging to some value (as \( n \to \infty \)) which is the same for all \( i \). Theorem. If an irreducible Markov chain is positive recurrent and aperiodic, there exists the limiting probability
\[ \lim_{n \to \infty} p^n_{ij} = \pi_j > 0 \text{ where } (i, j = 0, 1, 2 \ldots) \] (14)

Which is independent of the initial state \( i \), where \((\pi_j, j = 0, 1, 2, \ldots)\) is a unique and positive solution to

\[ \pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} \text{ where } (j = 0, 1, 2, \ldots) \] (15)

\[ \sum_{j=0}^{\infty} \pi_j = 1 \] (16)

and it is called a stationary distribution for a Markov chain [10].

2.6. Markov Chain Model for Monthly Rainfall Data

Chain-dependent models treat the occurrence and intensity of daily rainfall events separately [15]. The term “chain dependence” reflects the statistical structure of the occurrence sequence. The monthly rainfall model based on a Markov chain can used to determine rainfall occurrence (i.e., the high or low rainfall conditions) based on transition probabilities [16]. The transition probabilities, estimated from the historic measurements, represent the probabilities of high to high, high to low, low to high, and low to low. If the next day the rainfall is high, then the rainfall intensity is given as a random variable following a probability density function.

The daily rainfall model based on the Markov chain can be explained as follows. First, define \( X_t \) as the high and low rainfall condition at the \( t \)-th day. That is,

\[ X_t = 0, \quad \text{if } t \text{ day rainfall is high} \]
\[ X_t = 1, \quad \text{if } t \text{ day rainfall is low} \]

Assuming that the occurrence probability of rainfall at present is dependent on the condition of the previous day, then \( X_t \) follows the first-order Markov chain, and then the transition probability of daily rainfall can be divided into the following four cases:

\[ P[X_t = 1 \mid X_{t-1} = 0] = p_{01} \]
\[ P[X_t = 1 \mid X_{t-1} = 0] = p_{11} \]
\[ P[X_t = 0 \mid X_{t-1} = 1] = p_{00} \]
\[ P[X_t = 0 \mid X_{t-1} = 1] = p_{10} \]

The above equations express the conditional probabilities of wet or dry on day \( t \) depending on the condition of wet or dry on day \( t - 1 \). Therefore \( p_{00} = 1 - p_{01} \) and \( p_{11} = 1 - p_{10} \). Also, these four probabilities constitute a transition probability matrix:

\[ P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \] (17)

To derive the number of wet days, we first need to define the \( n \)-step transition probabilities:

\[ p^n = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}^n \] (18)

The \( n \)-step transition probabilities converge to certain probabilities as \( n \) increases:

\[ \lim_{n \to \infty} p^n = \lim_{n \to \infty} \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}^n = \begin{pmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{pmatrix} \] (19)

These probabilities \( \pi_0 \) and \( \pi_1 \) are represent the mean occurrence probabilities of high and low rainfall conditions. These are called the stationary probabilities [16, 17].
2.7. Data Mining and Knowledge Discovery in Database

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can include databases, data warehouses, the web, other information repositories, or data that are streamed into the system dynamically. As a knowledge discovery process, it typically involves data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation [18]. Data mining also known as knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.

![Figure 1. Knowledge Discovery in Database (KDD) Process](image)

The Figure 1. above is a view from typical database systems and data warehousing communities. Data mining plays an essential role in the knowledge discovery process. Data mining has many successful applications, such as business intelligence, Web search, bioinformatics, health informatics, finance, digital libraries, and digital governments [18].

2.8. Data Mining for Rainfall Data in West Java

In this research we use the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data of rainfall from LAPAN Bandung. CHIRPS is a big data, because it a 30+ year quasi-global rainfall dataset. Spanning 50°S-50°N (and all longitudes), starting in 1981 to near-present. In this research we use a CHIPS data from the year 1981 until 2017. The procedure of Data Mining approach for CHIRPS data is shown as following:

- Data pre-processing is a data mining technique that involves transforming raw data into an understandable format [17]. Raw data is highly susceptible to noise, missing values, and inconsistency. The quality of data affects the mining results. In order to help improve the quality of data and consequently of the mining results, raw data is pre-processed so as to improve the efficiency and ease of the mining process. Data pre-processing methods are divided into four categories that is data cleaning, data integration, data transformation, and data reduction [19].

- In this research, rainfall data in West Java as a big data is a space-time data that consist of 432 monthly observation, in which 36 years during 1981-2017 in 27 district and city. Data size $432 \times 27$ is then carried out using the data cleaning method. Cleaning method is to fill the empty value of the data and to ignore the noisy data and to correct the inconsistencies data. Cleaning method is used by selecting data based on wet months, namely December, January and February (DJF) for each year which began in December 1981 to February 2017 so that the data size become more understandable and easier to process with data size $108 \times 27$. 
The data mining process using the Markov chain method starts with defining $K$ state spaces, determine transition frequency matrix with an order of $K \times K$, determine the transition probability matrix, determine the stationary distribution, calculate the long-term proportion of the Markov chain, description and prediction.

Post-processing component can be categorized into the following groups that is knowledge filtering, interpretation and explanation, evaluation, and knowledge integration [20]. In the case of application of Markov chain for predicting rainfall data in West Java using data mining approach, the knowledge results are to determine the long-term probabilities for rainfall in West Java using the Markov chain method.

3. Results
The process of this research in principle is divided into four parts, namely collecting data, data mining processes, processing data using Markov chains method, and analysis results. The flow of this research is as described in the following flowchart.

![Flowchart of Markov Process using Rainfall Data in 4 State Spaces](image)

Rainfall in a location are generally unpredicted, rainfall data can be daily, weekly, monthly, or yearly. The data that will be used in this research is rainfall data in 27 districts/cities in West Java Province on wet months of December, January, and February calculated from December 1981 to February 2017.

| No. | City/District  | Latitude | Longitude | Dec-81 | Jan-82 | … | Feb-17 |
|-----|----------------|----------|-----------|--------|--------|----|--------|
| 1   | Dist. Bandung  | -7.12    | 107.57    | 317    | 323    | … | 325    |
| No | Dist.          | Latitude | Longitude | Rainfall Value | Range | ... |
|----|---------------|----------|-----------|----------------|-------|-----|
| 2  | Dist. Bandung Barat | -6.87    | 107.41    | 305            | 231   | ... |
| 3  | Dist. Bekasi   | -6.23    | 107.15    | 291            | 403   | ... |
| 4  | Dist. Bogor    | -6.58    | 106.88    | 408            | 394   | ... |
| 5  | Dist. Ciamis   | -7.22    | 108.39    | 552            | 588   | ... |
| 6  | Dist. Cianjur  | -6.81    | 107.13    | 373            | 315   | ... |
| 7  | Dist. Cirebon  | -6.77    | 108.51    | 336            | 647   | ... |
| 8  | Dist. Garut    | -7.39    | 107.77    | 361            | 347   | ... |
| 9  | Dist. Indramayu| -6.45    | 108.16    | 232            | 404   | ... |
| 10 | Dist. Karawang | -6.29    | 107.41    | 298            | 367   | ... |
| 11 | Dist. Kuningan| -6.98    | 108.50    | 416            | 574   | ... |
| 12 | Dist. Majalengka| -6.86   | 108.22    | 482            | 535   | ... |
| 13 | Dist. Pangandaran| -7.61  | 108.50    | 344            | 361   | ... |
| 14 | Dist. Purwakarta| -6.60   | 107.47    | 477            | 318   | ... |
| 15 | Dist. Subang   | -6.56    | 107.1    | 444            | 390   | ... |
| 16 | Dist. Sukabumi | -6.85    | 106.96    | 528            | 326   | ... |
| 17 | Dist. Sumedang | -6.82    | 107.99    | 351            | 478   | ... |
| 18 | Dist. Tasikmalaya| -7.54   | 108.14    | 408            | 412   | ... |
| 19 | Bandung City   | -6.92    | 107.62    | 171            | 220   | ... |
| 20 | Banjar City    | -7.37    | 108.53    | 412            | 612   | ... |
| 21 | Bekasi City    | -6.27    | 106.97    | 309            | 435   | ... |
| 22 | Bogor City     | -6.59    | 106.80    | 411            | 357   | ... |
| 23 | Cimahi City    | -6.88    | 107.54    | 227            | 178   | ... |
| 24 | Cirebon City   | -6.73    | 108.56    | 360            | 608   | ... |
| 25 | Depok City     | -6.39    | 106.80    | 343            | 415   | ... |
| 26 | Sukabumi City  | -6.93    | 106.93    | 377            | 309   | ... |
| 27 | Tasikmalaya City| -7.35   | 108.23    | 466            | 474   | ... |
| 28 | West Java Average| -7.50   | 108.48    | 200            | 260   | ... |

To show that the data is normally distributed, shown on the histogram and QQ plot in the following figure:

![Figure 3. Histogram of Rainfall Data in West Java](image1)

![Figure 4. Normal Q-Q Plot of Rainfall Data in West Java](image2)

The Figure 3 above shows that the results of data cleaning for the average West Java are normally distributed which is characterized by a bell-shaped normal distribution curve that is quite symmetrical. As for Figure 4, we see the point is match up along a straight diagonal line which shows that the quantiles match and normally distributed [21]. The simulation steps using the Markov chain model are as follows [10]:

- Defining $K$ state spaces.
- Rainfall data is divided in 4 states, i.e. a low rainfall (0), medium (1), high (2), and very high (3). Rainfall is categorized as low (0) if the rainfall value $\leq$ lower quartile, categorized as medium (1) if the lower quartile $<$ rainfall value $<$ middle quartile, categorized as high (2) if
the middle quartile < rainfall value < upper quartile, and categorized very high (3) if the rainfall value ≥ upper quartile.

- Determine a matrix with an order of \( K \times K \), with the initial state as a row and the final state as a column, then specify the number of displacements for each state.

**Table 2.** Frequency of Transition Rainfall 4 State Spaces in West Java

| No. | City/District       | Transition state \( i \) to \( j \), \( i = 0,1,2,3 \) and \( j = 0,1,2,3 \) |
|-----|---------------------|-----------------------------------------------|
|     |                     | 00   | 01   | 02   | 03   | 10   | 11   | 12   | 13   | 20   | 21   | 22   | 23   | 30   | 31   | 32   | 33   |
| 1   | Dist. Bandung       | 7    | 5    | 7    | 8    | 4    | 6    | 8    | 7    | 10   | 5    | 7    | 5    | 6    | 9    | 5    | 7    |
| 2   | Dist. Bandung Barat | 4    | 5    | 8    | 9    | 7    | 5    | 9    | 6    | 9    | 8    | 5    | 7    | 9    | 4    | 7    |     |
| 3   | Dist. Bekasi        | 5    | 6    | 6    | 10   | 10   | 5    | 3    | 9    | 6    | 8    | 8    | 5    | 6    | 8    | 9    | 3    |
| 4   | Dist. Bogor         | 7    | 5    | 5    | 10   | 8    | 7    | 7    | 5    | 6    | 8    | 5    | 8    | 6    | 7    | 9    | 4    |
| 5   | Dist. Ciamis        | 6    | 5    | 8    | 8    | 8    | 6    | 5    | 7    | 9    | 7    | 7    | 4    | 4    | 9    | 7    | 7    |
| 6   | Dist. Cianjur       | 4    | 8    | 7    | 8    | 8    | 3    | 8    | 7    | 10   | 8    | 4    | 5    | 5    | 8    | 7    | 7    |
| 7   | Dist. Cirebon       | 4    | 9    | 6    | 8    | 8    | 3    | 8    | 7    | 10   | 5    | 5    | 7    | 4    | 10   | 8    | 5    |
| 8   | Dist. Garut         | 4    | 5    | 11   | 6    | 8    | 7    | 4    | 8    | 7    | 7    | 7    | 6    | 8    | 8    | 4    | 7    |
| 9   | Dist. Indramayu     | 6    | 8    | 2    | 11   | 5    | 4    | 11   | 7    | 10   | 7    | 5    | 4    | 10   | 5    | 8    | 9    |
| 10  | Dist. Karawang      | 5    | 4    | 10   | 8    | 7    | 6    | 5    | 9    | 7    | 9    | 5    | 6    | 8    | 8    | 6    | 4    |
| 11  | Dist. Kuningan      | 7    | 6    | 6    | 8    | 6    | 3    | 9    | 8    | 8    | 9    | 7    | 3    | 6    | 8    | 5    | 8    |
| 12  | Dist. Majalengka    | 8    | 3    | 8    | 8    | 9    | 5    | 4    | 8    | 5    | 9    | 7    | 6    | 5    | 10   | 7    | 5    |
| 13  | Dist. Pangandaran   | 5    | 7    | 5    | 10   | 9    | 9    | 5    | 4    | 7    | 6    | 8    | 5    | 6    | 5    | 9    | 7    |
| 14  | Dist. Purwakarta    | 5    | 4    | 8    | 10   | 9    | 5    | 7    | 5    | 6    | 9    | 7    | 5    | 9    | 7    | 5    | 6    |
| 15  | Dist. Subang        | 6    | 4    | 8    | 9    | 11   | 2    | 7    | 6    | 5    | 9    | 7    | 6    | 5    | 12   | 4    | 6    |
| 16  | Dist. Sukabumi      | 6    | 6    | 7    | 8    | 4    | 7    | 8    | 10   | 7    | 6    | 3    | 7    | 7    | 6    | 7    | 6    |
| 17  | Dist. Sun medan     | 5    | 5    | 7    | 10   | 11   | 2    | 5    | 8    | 5    | 10   | 9    | 3    | 6    | 9    | 6    | 6    |
| 18  | Dist. Tasikmalaya   | 4    | 4    | 10   | 9    | 11   | 8    | 2    | 5    | 6    | 8    | 7    | 6    | 6    | 7    | 8    | 6    |
| 19  | Bandung City        | 3    | 3    | 12   | 8    | 10   | 4    | 4    | 9    | 11   | 7    | 4    | 5    | 2    | 13   | 7    | 5    |
| 20  | Banjar City         | 5    | 4    | 8    | 10   | 12   | 6    | 4    | 5    | 5    | 10   | 5    | 6    | 5    | 7    | 9    | 6    |
| 21  | Bekasi City         | 5    | 7    | 7    | 8    | 9    | 4    | 4    | 10   | 9    | 4    | 7    | 4    | 12   | 8    | 2    |     |
| 22  | Bogor City          | 5    | 5    | 8    | 9    | 12   | 6    | 4    | 5    | 4    | 5    | 6    | 11   | 6    | 11   | 8    | 2    |
| 23  | Cimahi City         | 3    | 5    | 5    | 13   | 7    | 6    | 6    | 8    | 10   | 5    | 8    | 4    | 7    | 11   | 7    | 2    |
| 24  | Cirebon City        | 5    | 8    | 7    | 7    | 7    | 3    | 7    | 9    | 11   | 4    | 8    | 4    | 4    | 11   | 5    | 7    |
| 25  | Depok City          | 5    | 8    | 6    | 8    | 6    | 5    | 8    | 9    | 4    | 6    | 7    | 7    | 10   | 6    | 4    |     |
| 26  | Sukabumi City       | 5    | 1    | 10   | 11   | 8    | 7    | 6    | 5    | 9    | 11   | 3    | 4    | 5    | 8    | 7    | 7    |
| 27  | Tasikmalaya City    | 6    | 5    | 7    | 9    | 10   | 4    | 6    | 6    | 7    | 11   | 5    | 4    | 4    | 7    | 9    | 7    |
| 28  | West Java Average   | 17   | 5    | 5    | 0    | 5    | 12   | 7    | 3    | 4    | 8    | 8    | 7    | 0    | 2    | 7    | 17   |

- The transition frequency matrix in the previous step is used to determine the transition probability matrix by dividing each transition value with the number in each row (relative frequency). Transition probability matrix 4 space state for rainfall average in West Java

\[
P_{\text{West Java Average}} = \begin{bmatrix} 0.629626963 & 0.185185185 & 0.185185185 & 0 \\ 0.185185185 & 0.4 & 0.259259259 & 0.1 \\ 0.148148148 & 0.296296296 & 0.296296296 & 0.259259259 \\ 0 & 0.076923077 & 0.269230769 & 0.653846154 \end{bmatrix}
\]

The transition diagram for rainfall average in West Java with 4 state spaces as follows:
Figure 5. Transition Diagram of Rainfall Average in West Java with 4 State Spaces

- Calculate the long-term proportion of the Markov chain with the $K$ state denoted by $\pi_k$ using equation (15).

Next, with $K = 4$, then $\pi_0, \pi_1, \pi_2$ and $\pi_3$ are the long-term proportions for state 0, 1, 2 and 3. The following results of the stationary distribution in percentages are presented in Table 3

| No. | City/District       | Stationary Distribution (%) | Low (0) | Quite Low (1) | Medium (2) | High (3) |
|-----|---------------------|-----------------------------|---------|---------------|------------|----------|
| 1   | Dist. Bandung       |                             | 25.76   | 24.81         | 25.76      | 25.76    |
| 2   | Dist. Bandung Barat |                             | 25.09   | 25.17         | 24.38      | 25.37    |
| 3   | Dist. Bekasi        |                             | 25.24   | 25.24         | 24.34      | 25.18    |
| 4   | Dist. Bogor         |                             | 25.24   | 25.21         | 24.43      | 25.12    |
| 5   | Dist. Ciamis        |                             | 25.37   | 25.13         | 25.18      | 24.32    |
| 6   | Dist. Cianjur       |                             | 25.18   | 25.09         | 24.43      | 25.3     |
| 7   | Dist. Cirebon       |                             | 24.43   | 25.06         | 25.29      | 25.21    |
| 8   | Dist. Garut         |                             | 25.14   | 25.18         | 24.44      | 25.24    |
| 9   | Dist. Indramayu     |                             | 24.47   | 25.21         | 25.29      | 25.03    |
| 10  | Dist. Karawang      |                             | 25.27   | 25.21         | 24.35      | 25.17    |
| 11  | Dist. Kuningan      |                             | 25.23   | 24.3          | 25.23      | 25.23    |
| 12  | Dist. Majalengka    |                             | 25.38   | 25.09         | 24.22      | 25.31    |
| 13  | Dist. Pangandaran   |                             | 25.28   | 25.29         | 25.2       | 24.24    |
| 14  | Dist. Purwakarta    |                             | 25.3    | 25.11         | 25.31      | 24.28    |
| 15  | Dist. Subang        |                             | 25.41   | 25.02         | 24.31      | 25.26    |
| 16  | Dist. Sukabumi      |                             | 25.35   | 25.24         | 25.25      | 24.17    |
| 17  | Dist. Sumedang      |                             | 25.23   | 24.3          | 25.23      | 25.23    |
| 18  | Dist. Tasikmalaya   |                             | 25.41   | 25.25         | 25.05      | 24.29    |
| 19  | Bandung City        |                             | 24.3    | 25.23         | 25.23      | 25.23    |
| 20  | Banjar City         |                             | 25.23   | 25.23         | 24.3       | 25.23    |
| 21  | Bekasi City         |                             | 25.11   | 25.47         | 24.31      | 25.11    |
| 22  | Bogor City          |                             | 25.23   | 25.23         | 24.3       | 25.23    |
| 23  | Cimahi City         |                             | 25.02   | 25.29         | 24.21      | 25.48    |
| 24  | Cirebon City        |                             | 25.23   | 24.3          | 25.23      | 25.23    |
| 25  | Depok City          |                             | 25.23   | 25.23         | 24.3       | 25.23    |
| 26  | Sukabumi City       |                             | 25.21   | 25.12         | 24.4       | 25.27    |
| 27  | Tasikmalaya City    |                             | 25.43   | 25.12         | 25.15      | 24.3     |
| 28  | West Java Average   |                             | 22.62   | 24.86         | 25.46      | 27.05    |

- Description and prediction.

Based on the Markov chain stationary distribution, the probability of changes in rainfall from low phenomena is 22.62%, medium is 24.86%, high is 25.46%, and very high is 27.05%. This result indicates that in the long term, the phenomena of monthly rainfall data in West Java area has a tendency to be change is still high with probability 0.27. It is still quite large compared to other state spaces low, medium, and high. The result can be given as a recommendation for a
related institutions such as LAPAN and BMKG. Furthermore, the institutions can announce a special attention to inform the situation of rainfall phenomena to society.

Acknowledgment
The authors thank Rector Universitas Padjadjaran for funding this research through Academic Leadership Grant and RKDU 2018. Also thanks the researchers from LAPAN Bandung for the data and discussion.

References
[1] Abdullah A S, Matoha S, Lubis D A, Falah A N, Jaya I G N M, Hermawan E, Ruchjana B N, 2018 Implementation of Generalized Space Time Autoregressive (GSTAR)-Kriging Model for Predicting Rainfall Data at Unobserved Locations in West Java. Applied Mathematics and Information Science, 12 (3), 607-615.
[2] http://bidinagtuns.blogspot.co.id/2010/11/curah-hujan.html, accessed at 13 March 2018
[3] Rustiana S, Ruchjana B N, Abdullah A S, Hermawan E, Sipayung S B, Jaya I G N M and Krismianto. Rainfall prediction of Cimanuk watershed regions with canonical correlation analysis (CCA), Journal of Physics: Conf. Series, 89312021, doi:10.1088/1742-6596/893/1/012021, 2017.
[4] Sandi R 2015 Simulasi Curah Hujan Harian Menggunakan Stokastik Rantai Markov Orde 3 × 3 (Studi Kasus: Daerah Aliran Sungai Kampar). Jom FTEKNIK Volume 2 No. 2 October 2015.
[5] Sujatmoko and Bambang 2012 Analisa Kehandalan Stokastik Rantai Markov untuk Simulasi Data Curah Hujan Harian pada Das Kampar. Jurnal Sains dan Teknologi 11 (1), ISSN 1412-6625.
[6] Tovler A 2016 An Introduction to Markov Chains. Department of Mathematical Sciences, University of Copenhagen, Denmark.
[7] Ross S M 1996. Stochastic Processes Second Edition. University of California, Berkeley, United States of America.
[8] Hermawan E, Ruchjana B N, Abdullah A S, Jaya I G N M, Sipayung S B, Rustiana S Development of the statistical ARIMA model: an application for predicting the upcoming of MJO index, Journal of Physics.: Conf. Ser. 893 012019, doi:10.1088/1742-5969/893/1/012019, 2017.
[9] Doubleday K J and Julius N E 2011. Application of Markov Chains to Stock Trends. Journal of Mathematics and Statistics, 7 (2): 103-106
[10] Osaki S 1992 Applied Stochastic System Modelling. Berlin: Springer-Verlag Berlin Heidelberg.
[11] https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa, accessed at 20 January 2019
[12] Abubakar U Y, Lawal A, Muhammed A 2014 Markov Chain Model to Annual Rainfall Distribution for Crop Production, American Journal of Theoretical and Applied Statistics, Vol. 3, No.2, 2014, pp. 39-43.
[13] Firdaniza 2016. Distribusi Stasioner Rantai Markov untuk Prediksi Curah Hujan di Wilayah Jawa Barat. Proceeding of Seminar Matematika dan Pendidikan Matematika. ISBN: 978-602-6122-20-9
[14] Munkhammar J and Widen J 2018 Statistical ARIMA model: an application for predicting the upcoming of MJO index, Journal of Physics.: Conf. Ser. 893 012019, doi:10.1088/1742-6596/893/1/012019, 2017.
[15] Chung C H 2007 Vegetation response to climate change on Jeju Island, South Korea, during the last deglaciation based on pollen record. Geosciences Journal, vol. 11, no. 2, pp. 147–155.
[16] Yoo C, Lee J, Ro Y 2016 Markov Chain Decomposition of Monthly Rainfall into Daily Rainfall: Evaluation of Climate Change Impact. Research Article, Advances in Meteorology. DOI: 10.1155/2016/7957490
[17] Zohadi B M and Salam A C A 1981. A Stochastic Model of Daily Rainfall for Universiti Pertanian Malaysia, Serdang, Pertanika, 4 (1), 1-9
[18] Han J and Kamber M 2006 Data Mining: Concept and Techniques Second Edition, Morgan Kaufmann Publishers.
[19] Tamiselvi R, Sivakshiti B and Kavitha R 2015 An Efficient Preprocessing and Postprocessing Techniques in Data Mining. International Journal of Reseach in Computer Applications and Robotics. ISSN: 2320-7345
[20] Diaz J L, Herrera M, Izquierdo J, Garcia R P 2010 The tasks of pre and post-processing in Data Mining applied to a real world problem. International Congress on Environmental Modelling and Software. Brigham Young University
[21] Wang Y, Steele T and Zhang E 2016 QQ Plot. https://math.illinois.edu/system/files/inline-files/Proj9AY1516-report2.pdf