

Application of Markov Chain in Predicting Sugar Production at Candi Baru Sugar Factory, Sidoarjo

Agustina Pradjaningsih^{1, a)}, Ardelia Nani Vidatiyasa^{2, b)}, Kiswara Agung Santoso³⁾

^{1,2,3}*Department of Mathematics, Faculty of Mathematics and Natural Science, Jember University*

^{a)}Corresponding author: agustina.fmipa@unej.ac.id

^{b)}ardeliananitiyasa@gmail.com

Abstract. Granulated sugar is a sugar commonly used daily to manufacture food and beverages. The demand for granulated sugar continues to increase, but the number of sugar factories and the area of sugar cane in Indonesia is decreasing. This causes a gap between the demand for sugar which continues to increase, and the production of granulated sugar continues to decline, resulting in Indonesia being the largest country importer of sugar. The imbalance between the demand and production of granulated sugar At Candi Baru Sugar Factory, Sidoarjo, East Java, resulted in not achieving the target to meet these needs. Therefore, predictions are made to get an overview of production planning to optimize granulated sugar production so that sugar needs can be met. The prediction method used at the Candi Baru sugar factory, Sidoarjo, East Java, for the 2022 milling period is the Markov Chain method with a four-state divisor, namely drastically down, down, up, and up drastically. The application of Markov Chains produces predictions for each state. It is predicted that the production of sugar with the highest percentage for May – December upstate.

Keywords: *Markov Chain, Prediction, Granulated Sugar Production*
2020 Mathematical Subject Classification : 90C40

INTRODUCTION

Sugar is a natural sweetener and a simple carbohydrate that is easily soluble in water and turned into energy for the body [14]. Granulated sugar is a type of sugar used daily in the manufacture of food or drinks and comes from sugar cane juice which is crystallized into sugar granules. Demand for granulated sugar continues to increase in Indonesia, caused of several things: the habit of consuming granulated sugar, population growth, community welfare, and the development of industries that use granulated sugar as raw material [8]. Indonesia has the potential as the largest sugar-producing country in the world. Still, in reality, the number of sugar factories is decreasing, and the sugar cane area is decreasing, so the State of Indonesia is currently the largest sugar importer. The gap between the demand for sugar continues to increase, and production continues to decline. Production planning is needed to optimize the production strategy so that sugar needs and requests are met. Production planning is an activity related to the process of converting raw materials into an item [20]. Production planning determine based on the results of forecasting or predictions. Markov Chain can use as a prediction method. Markov Chain studies the properties of a variable in the present and past to estimate variables in the future. The result of Markov Chain analysis is probabilistic information. This analysis uses to assist decision-making. PG Candi Baru Sidoarjo is one of the companies that produce granulated sugar, which is experiencing an imbalance between demand and realization of sugar. From the description above, researchers are interested in conducting research, namely applying the Markov Chain in predicting sugar production with the help of QM software for Windows.

METHODOLOGY

Data is a collection of facts that can be trusted to be accurate; the available data is often raw data that has not been compiled and has no information, so the data needs to be processed in such a way to get information. Statistics can be

divided into two types: descriptive and inductive. An example of descriptive statistics is frequency distribution, where data are arranged according to size or category and presented in tables or graphs. The steps in forming the frequency distribution are as follows [22]:

- a. Determine the number of classes

The number of classes can be determined freely according to needs or using the Sturges formula below

$$K = 1 + 3.3 \log a \tag{1}$$

with K is the number of classes, and a is the number of data.

- b. Determine the length of the class interval

In determining the class length, the data's distance or range must be known. The range is the distance between the largest data and the smallest data

$$R = \text{largest data} - \text{smallest data}$$

The known range is divided by the number of classes (K), as follows [22]

$$I = \frac{R}{K} \tag{2}$$

with I is class length, R is the range, and K is the number of classes.

Inductive statistics are statistics that assess characteristics, predict and draw general conclusions. The probability of an event is a measure of the likelihood of an event occurring in the future. Probability has a conditional event rule that the occurrence of one event is used as a condition for the occurrence of another event which is stated in the following Equation: [22]

$$P(A|B) = \frac{P(A \cap B)}{P(B)}; P(B) \neq 0 \tag{3}$$

with $P(A|B)$ is conditional probability of event A if event B is known, $P(A \cap B)$ is the probability that both A and B occur at the same time, and $P(B)$ is probability of event B.

A Markov Chain is a series of event processes in which the conditional probability of future events depends on the current events. The results of Markov Chain analysis are probabilistic information that can be used to assist decision-making [1]. Markov analysis is a special form of probabilistic model, which is more commonly called a stochastic process. A stochastic process is a set of random variables. All possible values that can occur in the random variable of the stochastic process are called state space, so the stochastic process is expressed in Equation. [17]

$$X = \{X(t), t \in T\} \tag{4}$$

with X is set of random variables, $X(t)$ is random variables, T is time index set, and t is time.

A stochastic process X_t is said to have Markov properties if for $t = 1, 2, 3 \dots$ it has a transition probability or a probability of moving state i at time t to state j at time $t + 1$, which is called P_{ij} [17].

The Markov Chain is represented in a transition probability matrix called a one-step transition probability matrix. P_{ij} is the probability is in the state i , then the process will transition to the state j , as in the matrix P below [17]. The relationship between states can represent in Figure 1.

$$P = \begin{bmatrix} P_{00} & P_{01} & \dots & P_{0j} \\ P_{10} & P_{11} & \dots & P_{1j} \\ \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & \dots & P_{ij} \end{bmatrix}$$

with $P_{ij} \geq 0$; $i, j \geq 0$; $\sum_{j=0}^{\infty} P_{ij} = 1$; $i = 0, 1, \dots$, P is a one-step transition probability matrix, and P_{ij} is transition probability from state i to state j .

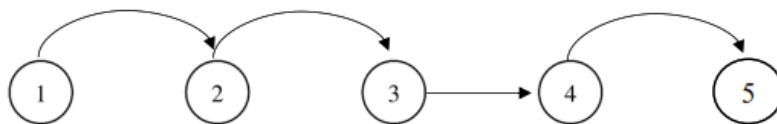


FIGURE 1. Relationship between states

The one-step transition probability matrix is the probability of the initial state of a system so that to find out the conditional probability that a system in the state i will be in a state j , an n -step transition probability (P_{ij}^n) is performed. So to calculate the probability of an n -step transition, you can use the Chapman-Kolmogorov equation as follows [17]

$$P^n = P^{n-1}P \tag{5}$$

with P^n is probability of the n^{th} state transition, $n = 1, 2, \dots$, P^{n-1} is a probability of state transition at $n-1$, and P is transition probability matrix P .

The state of the Markov Chain transition matrix for each state in each -step generally cannot be determined with certainty, so a probability for each state in the initial conditions is needed, called the initial state vector (v^0). The initial state vector indicates an equal chance of every state in a system starting a transition [11].

$$v^0 = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix}$$

where v_1 is the probability at state 1, v_2 is the probability at state 2, and v_k is the probability at state k . If P^n is the transition matrix of the Markov Chain and v^0 is the initial state vector, then

$$v^n = v^0 P^n \tag{6}$$

with v^n is state vector, $n \geq 1$, v^0 is initial state vector, and P^n is transition probability matrix, $n \geq 1$

This iterative process is used to predict the probability of the future state of the transition process [10]. After performing the n -step transition process, the system reaches a steady state where the transition probability has reached a fixed condition so that the transition probability value does not change again after makin n -step transitions. Repeated probability calculations will take time, so the calculation process is not efficient. In an ergodic (recurrent and aperiodic) Markov Chain, the steady-state probability can be solved using the following Equation [17]

$$\pi = P\pi \tag{7}$$

RESULTS AND DISCUSSION

The data was used during the sugarcane milling period from 2010 – 2021. According to the data obtained, the Candi Baru sugarcane milling period occurs from May to December, so the predictions obtained at the end of the calculation are predictions for the 2022 milling period. The data are in the form of data production of granulated sugar obtained from the Candi Baru Sugar Factory. The required data are in Table 1.

TABLE 1. Data variable

Variable	Description
X	Granulated sugar production (Tons)

The sugar production data is sorted, then arranged by category using frequency distribution and presented in tables. The first step is to find the range of granulated sugar production data. Range data is obtained by finding the difference between the largest and smallest data values. Based on the sugar production data, the largest data value is 7281.1 tons, and the smallest data is 262 tons, so the range of granulated sugar production data is

$$R = 7281.1 - 262 = 7019.1.$$

The range that has been obtained is used to find class intervals with Equation (2) as follows:

$$i = \frac{R}{4} = \frac{7019.1}{4} = 1754.8 \approx 1755.$$

The length of class intervals obtained can be used to form class intervals. The class interval will contain all the smallest to largest data that has been partitioned into four classes, with each class having a class interval length of I

[22]. So that state categories can be formed, as in Table 2, and Table 2 is used to classify each data according to the state variable category in Table 3.

TABLE 2. State category

State	Description	Range
1	Drastically decrease	262.0 – 2016.9
2	Decrease	2017.0 – 3771.9
3	Increase	3772.0 – 5526.9
4	Drastically increased	5527.0 – 7281.9

TABLE 3. Sugar production data classification

Data every month	Classification	Data every month	Classification	Data every month	Classification	Data every month	Classification	
May-2010	1630.1	1	Oct-2012	7222.1	4	Aug-2015	6736.5	4
Jun-2010	3642.9	2	Nov-2012	2035.8	2	Sep-2015	5995.4	4
Jul-2010	4044.7	3	May-2013	383.7	1	Oct-2015	6879.7	4
Aug-2010	4430.9	3	Jun-2013	2012.9	1	Nov-2015	2269.5	2
Sep-2010	2605.4	2	Jul-2013	4401.1	3	Jun-2016	2434.5	2
Oct-2010	4276.8	3	Aug-2013	3012.2	2	Jul-2016	2749.2	2
Nov-2010	3814.2	3	Sep-2013	5540.5	4	Aug-2016	5263.0	3
Des-2010	1112.3	1	Oct-2013	5621.6	4	Sep-2016	4811.9	3
May-2011	2859.7	2	Nov-2013	5866.1	4	Oct-2016	5420.0	3
Jun-2011	4531.1	3	Des-2013	3985.6	3	Nov-2016	5032.3	3
Jul-2011	4935.1	3	May-2014	946.4	1	Des-2016	4408.1	3
Aug-2011	4562.0	3	Jun-2014	4702.9	3	May-2017	1218.2	1
Sep-2011	3480.5	2	Jul-2014	4044.2	3	Jun-2017	3397.7	2
Oct-2011	7281.1	4	Aug-2014	4790.2	3	Jul-2017	6308.1	4
Nov-2011	786.1	1	Sep-2014	6246.7	4	Aug-2017	6373.8	4
May-2012	1589.2	1	Oct-2014	5936.9	4	Sep-2017	5746.6	4
Jun-2012	4829.5	3	Nov-2014	4657.8	3	May-2018	3201.2	2
Jul-2012	5978.3	4	May-2015	262.0	1	Jun-2018	3162.4	2
Aug-2012	3982.6	3	Jun-2015	5457.7	3	Jul-2018	4478.2	3
Sep-2012	6695.8	4	Jul-2015	3933.1	3	Aug-2018	5557.1	4

The following process is the preparation of the transition matrix obtained from the state transitions in Table 3. The transition matrix formed is used to form a one-step transition matrix to get an overview of the relationship between each state from the sugar production data. From Table 3, a transition matrix can be formed where each element contains the number of state transfers in each data sequence.

$$X = \begin{bmatrix} 2 & 3 & 5 & 0 \\ 2 & 5 & 6 & 4 \\ 4 & 5 & 14 & 6 \\ 1 & 5 & 4 & 12 \end{bmatrix}$$

Elements in the X matrix indicate a state transfer for all states (i) to all states (j) in the sugar production data. Each element in the transition matrix is solved by Equation (3) to produce the one-step transition probability matrix below.

$$X = \begin{bmatrix} \frac{2}{10} & \frac{3}{10} & \frac{5}{10} & \frac{0}{10} \\ \frac{2}{17} & \frac{5}{17} & \frac{6}{17} & \frac{4}{17} \\ \frac{4}{29} & \frac{5}{29} & \frac{14}{29} & \frac{6}{29} \\ \frac{1}{22} & \frac{5}{22} & \frac{4}{22} & \frac{12}{22} \end{bmatrix} = \begin{bmatrix} 0.20 & 0.30 & 0.50 & 0 \\ 0.12 & 0.29 & 0.35 & 0.24 \\ 0.14 & 0.17 & 0.48 & 0.21 \\ 0.05 & 0.22 & 0.18 & 0.55 \end{bmatrix}$$

From each one-step transition matrix that has been formed, it can be described in terms of the relationship between sugar production states as Figure 2 below.

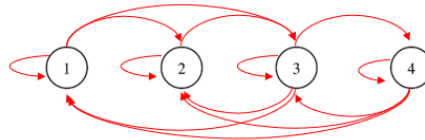


FIGURE 2. Relationship between sugar production state

Based on the illustration of the relationship between sugar production states shows a relationship that is not mutually exclusive, so other conditions influence the probability of each state.

The process is continued by looking for probabilities with n -step transitions; researchers use the help of QM for Windows software to get probabilities in each n -step until they reach steady state conditions. In this study, a dataset with the name of granulated sugar production will be created, and the number of states that will be used is four. Press "OK" after everything is in order. The initial state vector used in this study is $v^0 = [0.25 \ 0.25 \ 0.25 \ 0.25]$, so that the sugar production dataset template is obtained, which contains the one-step transition matrix elements of granulated sugar production, as shown in Figure 3 below.

Produksi gula pasir					
	Initial	State 1	State 2	State 3	State 4
State 1	,25	,2	,3	,5	0
State 2	,25	,12	,29	,35	,24
State 3	,25	,14	,17	,48	,21
State 4	,25	,05	,22	,18	,55

FIGURE 3 Sugar production dataset

Based on the data template in Figure 3, using the "number of transition" tool, then press "run" so that each iteration will produce an output of a transition matrix, ending probability (given initial)/state vector/ v^n and steady-state probability. The transition of the granulated sugar production dataset is shown in Figure 4 below.

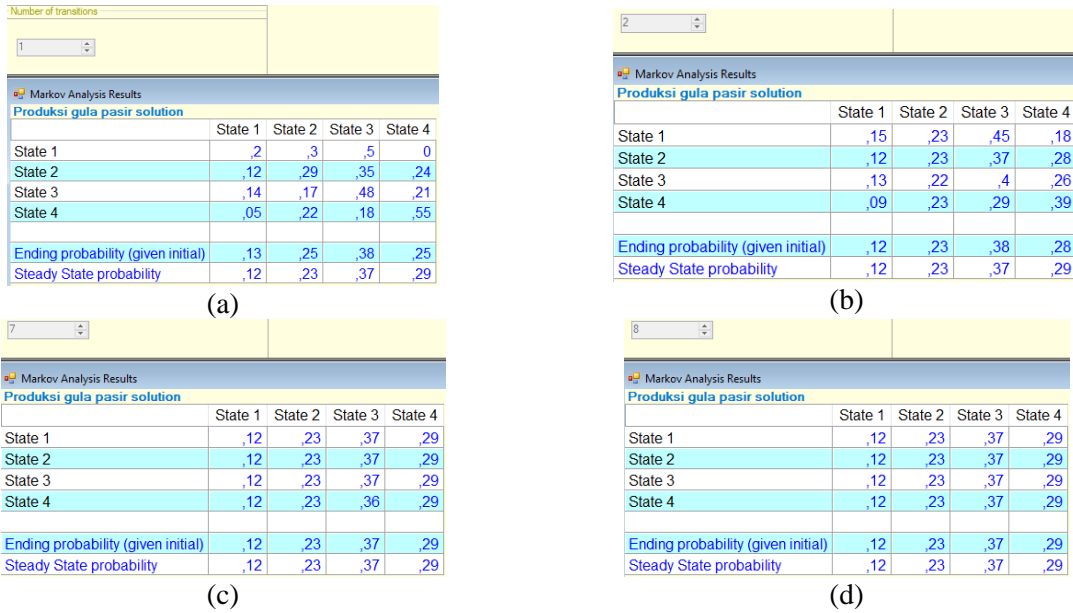


FIGURE 4. (a) first transition, (b) second transition, (c) seventh transition, (d) eighth transition.

Based on Figure 4 (d), sugar production reached a stable condition in the eighth iteration. Each dataset transition of each variable produces an ending probability/state vector, which is the probability of each transition which is presented in Table 5 below.

TABLE 5. Sugar production probability

Probability	State 1	State 2	State 3	State 4
$v^1 =$	0.13	0.25	0.38	0.25
$v^2 =$	0.12	0.23	0.38	0.28
$v^3 =$	0.12	0.23	0.39	0.29
$v^4 =$	0.12	0.23	0.38	0.28
$v^5 =$	0.12	0.23	0.37	0.29
$v^6 =$	0.12	0.23	0.37	0.29
$v^7 =$	0.12	0.23	0.37	0.29
$v^8 =$	0.12	0.23	0.37	0.29

Table 5 contains the production probabilities for granulated sugar during eight transitions, indicating that the largest probability is in state 3. Based on classified sugar production data, most state transitions ended in state 3. In QM for Windows, the state of granulated sugar production can be analyzed by the tool "State analysis." The state of the sugar production system is classified in Table 5.

State	Type	Class
State 1	Recurrent	1
State 2	Recurrent	1
State 3	Recurrent	1
State 4	Recurrent	1

FIGURE 5. Sugar production state analysis

Figure 5 shows that State 1 to state 4 in sugar production fulfills the steady-state ergodic properties so that through Equation (7), the steady-state probability of granulated sugar production can be seen in Table 6.

TABLE 6. Steady-state probability for sugar production

	π_1	π_2	π_3	π_4
Sugar production	0.12	0.23	0.37	0.29

Table 6 is a table that shows the steady-state probability of sugar production. It is found that the long-term probability of granulated sugar production in a drastically decreasing condition is $\pi_1 = 0.12$, in a down condition $\pi_2 = 0.23$, in an up condition $\pi_3 = 0.37$, and a drastically up condition $\pi_4 = 0.29$. The steady-state probability shows a fixed probability after performing the -step transition process; the steady-state probability can change if additional data is added and the Markov Chain process is carried out again. Based on Table 5, the prediction of granulated sugar production is obtained by multiplying each state vector by 100% so that the prediction of sugar production can be seen in Figure 6.

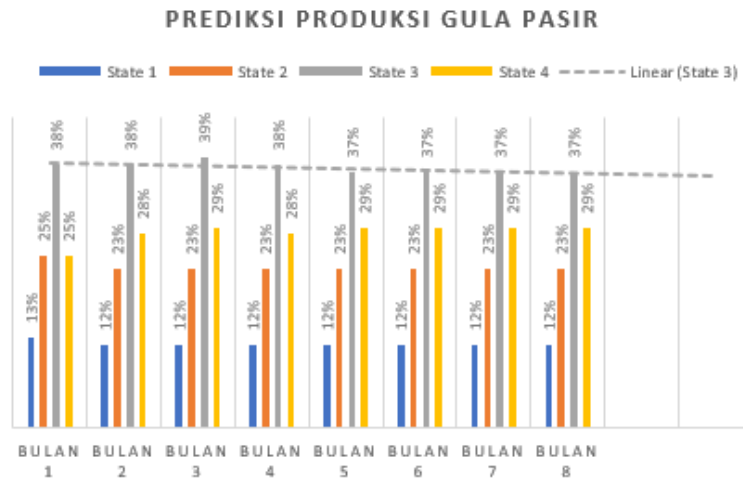


FIGURE 6. Sugar production prediction

Based on the picture above, it is found that production predictions from May to December will be in state 1 (down drastically) of 12% - 13%. The prediction of sugar production from May to December will be in state 2 (down) at 23% - 25%. Sugar production from May to December is predicted to be in state 3 (up) at 3772 – 5526.9 ton) with 37% - 38%. Sugar production from May to December is predicted to be in state 4 (a drastic increase) by 25 – 29%. The explanation above shows that the largest prediction is in state 3, so for the months of May to December 2022, sugar production is predicted to be in an upward condition. After December, sugar production will be stable at state three or an increasing condition.

CONCLUSION

Based on the results and discussion of Markov Chain predictions on the four data variables, it was successfully carried out with the help of QM for Windows. The most significant prediction is that sugar production for May to December will be in increasing condition (3772 – 5526.9 tons) with a percentage level of 37% - 39%

ACKNOWLEDGMENTS

The author would like to thank the University of Jember LP2M, which provided funding for the 2022 KeRis-DiMas Internal Grant under the MOC0 (Mathematical Optimization and Computation) Keris of Mathematics Department, Faculty of Mathematics and Natural Sciences, the University of Jember with research contract number 4615/UN25.3.1/LT/2022

REFERENCES

- [1] A.R, Putra, *Riset Operasional dengan POM-QM for Windows* (Desanta Multivisitama, Serang), pp. 59–60, 2018.
- [2] A, Ridhwan., D.E, Ratnawati and B, Rahayudi, “Peramalan Produksi Gula Pasir Menggunakan *Fuzzy Time Series* Dengan Optimasi Algoritma Genetika (Studi Kasus PG Candi Baru Sidoarjo),” *Journal of Information Technology and Computer Science*. (Brawijaya University, Malang), pp. 2543–2548, 2018.
- [3] A.S, Rachman., I, Cholissodin and M.A, Fauzi, “Peramalan Produksi Gula Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation Pada PG Candi Baru Sidoarjo,” *Journal of Information Technology and Computer Science*. (Brawijaya University, Malang), pp. 1683–1689, 2018
- [4] Aulia, “Aplikasi Rantai Markov Dalam Memprediksi Ekspor dan Impor Migas Di Indonesia,” B.Sc. Thesis, Sumatera Utara University, 2021.
- [5] Aulia and Iqbal, “Pendekatan Rantai Markov Waktu Diskrit Dalam Memprediksi Perencanaan Produksi Padi Terhadap Lahan Panen Di Sumatera Utara,” B.Sc. Thesis, Sumatera Utara University, 2018.
- [6] Basorudin and Dona, “Penerapan Metode Markov Chain Untuk Memprediksi Hasil Panen Kelapa Sawit dan Karet di Kabupaten Rokan Hulu,” *Journal of Informatics and Computer Science*. (Pasir Pengaraian University, Riau), pp. 116–123, 2020
- [7] B, Harstanto, *Naskah Tutorial QM for Windows* (Universitas Padjajaran, bandung), pp. 3, 2011
- [8] E, Satriana., E, Tety and A, Rifai, “Faktor-Faktor yang Mempengaruhi Konsumsi Gula Pasir di Indonesia,” *Journal of Agribusiness*. (Riau University, Riau), pp. 01–15, 2014.
- [9] E.Y, Darmayanti., B.D, Setiawan and F.A, Bachtiar, “Particle Swarm Optimization Untuk Optimasi Bobot Extreme *learning Machine* Dalam Memprediksi Produksi Gula Kristal Putih Pabrik Gula Candi Baru-Sidoarjo,” *Journal of Information Technology and Computer Science*. (Brawijaya University, Malang), pp. 5096–5104, 2018
- [10] H.A, Taha, *Operations Research An Introduction 10th Edition* (Pearson Education, England, 2017), pp. 632.
- [11] H.J, Wells, *Software for Decision Science: Quantitative Methods Production and Operations Management*. (England, Pearson Education), pp. 147, 2015.
- [12] I.N, Rizanti and Soehardjoepri, “Prediksi Produksi Kayu Bundar Kabupaten Malang Dengan Menggunakan Metode *Markov Chain*,” *Journal of Sciences and Arts*. (Teknologi Sepuluh November Institute of Technology, Surabaya), pp. 2337–3520, 2017
- [13] Karmini, *Ekonomi Produksi Pertanian* (Mulawarman University press, Samarinda), pp. 16, 2018.
- [14] L.J, Cseke., A, Kirakosyan., P.B, Kaufman., S.L, Warber., J.A, Duke., H.L, Brielmen, *Natural Product From Plant Second Edition* (CRC Press, Boca Raton), pp. 27, 2006.
- [15] Nurfitrianti, “Penerapan Data Mining Untuk Prediksi Harga Beras Di Indonesia Menggunakan Model Markov,” B.Sc (Tech). Thesis, Sultan Syarif Kasim Islamic State University, 2019.
- [16] S.D, Anitasari., D.N.R, Sari., I.A, Astarini., M.R, Defiani, *Teknologi Kultur Mikroskopa Tebu* (LPPM IKIP PGRI Jember Press, Jember), pp. 5–9, 2018.
- [17] S.M, Ross, *Introduction to Probability Models 10th Edition* (Elsevier Inc, California), pp. 192–215, 2010.
- [18] S, Maghfiroh dan F, Hilmiyah, “Prediksi Hasil Produksi Pajale di Kabupaten Jember Menggunakan Metode Markov Chain,” *Journal of Science, Technology dan Industry*. (Jember University, Jember), pp. 145–150, 2021.
- [19] S, Sasake., Y.A, Lesnussa and A.Z, Wattimena, “Peramalan Cuaca Menggunakan Metode Rantai Markov (Studi Kasus: Cuaca Harian Di Kota Ambon),” *Journal of Mathematics*. (Pattimura University, Ambon), pp. 01–09, 2021.
- [20] S, Sinulingga, *Perencanaan dan Pengendalian Produksi Edisi Pertama* (Graha Ilmu, Yogyakarta, 2009), pp. 23.
- [21] T.A, Nurman., I, Syata and C.D, Wulandari, “Prediksi Hasil Panen Kopi Di Sulawesi Menggunakan Analisis Rantai Markov,” *Journal of Applied Mathematics and Statistics*. (Alauddin Islamic State University, Makasar), pp. 120–127, 2021
- [22] T, Sariwulan, *Pengantar Statistika Ekonomi dan Bisnis* (Samudra Biru, Yogyakarta), pp. 15–16, 2018.