Analysis of the Markov Chain Approach to Detect Blood Sugar Level

Arif Ridho Lubis, Santi Prayudani, Muharman Lubis, Al-Khowarizmi
Politeknik Negeri Medan, Jalan Almamater No. 1, Medan, 20155
Telkom University, Jalan Telekomunikasi No. 1, Bandung, 40257
Universitas Muhammadiyah Sumatera Utara, Jalan Kapten Mucthar Basri, Medan
Email: arifridho@polmed.ac.id

Abstract. Diabetes mellitus is a health disorder in which the level of sugar in a person's blood is getting higher due to the lack of insulin or insulin receptors in the body do not properly function. Indonesia ranks the 4th largest in the number of Diabetes Mellitus sufferers in the world. Many people initially did not know that they had diabetes mellitus and only found out that they had had diabetes after experiencing some complications in some organs of the body. To predict blood sugar levels, this study tries to use one method of probability approach, i.e. Markov chain. It is a calculation technique that is commonly applied in modeling many conditions so that this technique is applied to help predict changes that might occur in the future. The results of the analysis applying this method could predict blood sugar levels, which then the results can be used as material or basis for creating a system that is able to help people affected by diabetes mellitus in predicting the high and the low of blood sugar levels.

1. Introduction

The development of science and information technology that is getting faster and faster has changed the people's way of life in the world in running their daily activities. The existence and role of information technology in all sectors of life entering the era of globalization faster than imagined. The impact is not only affecting the economic and political sectors of the entire country, even entering the social and cultural aspects of humans. It is not strange that the development of computers has brought the world to a new era of information age, not only for the fulfillment of information but also for solving problems in everyday life. One of the problems that can be addressed is in the health sector. Where this information technology plays a role in the advancement of the health sector [1]. Diabetes mellitus is a health disorder in which the level of sugar in a person's blood becomes high due to lack of insulin or insulin receptors do not function properly. Diabetes caused by insulin deficiency is called Type 1 DM or Insulin Dependent Diabetes Mellitus (IDDM). Diabetes caused by improperly functioning insulin is called Type 2 DM or Non-Insulin Dependent Diabetes Mellitus (NIDDM). Diabetes mellitus has become the fourth largest cause of death in the world and the number is increasing year by year [3].

According to WHO data [2], Indonesia ranks the 4th largest in the number of people with Diabetes Mellitus in the world. Many people initially did not know that they suffer from diabetes mellitus, in Asian countries more than 50 percent (some even reach 85 percent) new diabetics know themselves have diabetes after experiencing complications in some organs of the body [3]. This ignorance is caused by the lack of information about diabetes, its symptoms and the lack of diabetes specialist doctors.
In insufficient knowledge about the symptoms and how to deal with diabetes mellitus as well as the limited number of specialist doctors of diabetes mellitus is one reason why the number of people affected by the disease is increasing. To predict blood sugar levels, this study tries to apply one method of probability approach i.e. Markov chain. It is a calculation technique that is generally applied in modeling many conditions so that it is applied to assist in estimating changes that may occur in the future [4]. Conducted [5] a study applying the Markov chains method in predicting the results of round wood production in kabupaten Malang, in which the study had a 100% compatibility in producing round wood. Conducted [6] research with markov chains to find out the opportunity to move the brand of GSM prepaid cellular cards. In a different method [7] identified diabetes by using the k-Nearest Neighbor method. Based on the testing that had been done, it obtained the best accuracy results of 93.33% with an error rate of 6.67%. Based on these results, a system applying the Modified K-Nearest Neighbor (MKNN) method can be implemented in daily life in cases of diabetes mellitus.

Conducted [8] a study to determine diabetes mellitus by applying the certainty factor method. Based on the results of the implementation and testing it was concluded that SPPK diabetes mellitus could diagnose patients with 80% accuracy. Based on the problems described above, this research will apply Markov chain to predict blood sugar levels, which then can be used as material or basis for making a system that is capable of making people affected by diabetes mellitus in predicting the high and low levels of blood sugar with the Markov chain approach.

2. Research Methodology

This study analyzes the Markov chain approach to predict the blood sugar levels of diabetic (DM) people. In DM there are 4 attributes as references for calculating blood sugar levels in patients. These attributes will later form a pattern with the existing food menu class. From these data, doctors can detect patterns by applying several methods, one of which applies data mining. Where in this study will be carried out to predict blood sugar levels based on previous behavior by calculating the probability of the Markov Chain. The applied data in this study are the dataset of diabetic patients which were obtained from UCI Machine Learning from the study of Michael Kahn, MD, PhD, Washington University, St. Louis, MO. The parameters to determine the food menu for diabetics were four parameters, they are height, weight, age, and activity. Diabetic patient data obtained for this study are 100 data. The applied data are the incidence data of patients in 4 periods. Markov chain is a probability technique that can analyze the movement from one condition to another. It can also be applied to analyze the events in the future mathematically. Where the basic concept of Marcov Chains was first introduced by Russian mathematician Andrei A. Marcov in 1907. The state of transition is a change from a state to another in the next period, where in this transition state is a random process and expressed in the form of probability. Then it is called the transition probability applied to determine the probability of the next state or period.

\[ P = [p_{ij}]_{N\times N} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix} \]

Given \( P = [p_{ij}] \) becomes probability of the transition matrix \( N \) at DTMC \( \{X_n,n\ge 0\} \) with state space \( S = \{1,2,\ldots,N\} \) therefore:

\[ P_{ij} \ge 0, 1 \le i,j \le N \]
\[ \sum_{i=1}^{N} p_{i} = 1, 1 \le i \le N \]

(1)

(2)

Probability of transition \( n \)-step fill the following equation, called Chapman-Kolmogorov equation:

\[ p_{ij}^{(n+m)} = \sum_{k=1}^{N} p_{ik}^{(n)} p_{kj}^{(m)} \]

(3)

From above equation, it can be stated that, \( p_{ij}^{(n+m)} = p_{ij}^{(n)} \cdot p_{ij}^{(m)} \cdot p_{ij}^{(n+m)} = p_{ij}^{(n)} \cdot p \). with \( n = 0 \) it obtains such equation
From the recursive formula above so that the equation can be obtained:

\[
p^{(1)} = p^{(0)} * P
\]

\[
p^{(2)} = p^{(1)} * P = p^{(0)} * P * P = p^{(0)} * p^2
\]

\[\vdots \]

\[
p^{(n)} = p^{(n-1)} * P = \cdots = p^{(0)} * p^n
\]

From the recursive formula above so that the equation can be obtained:

\[
p^{(n+1)} = p^{(0)} * p^{n+1}
\]

The recursive formula, prediction can be made based on the interpretation of dynamic system. If P is the probability and \( \pi_0 \) is the reference point therefore the value of \( \pi_n \) can be obtained by:

\[
\pi_n = \pi_{(n-1)}^P
\]

\[
\pi_{(n-1)} = \pi_{(n-2)}^P; \pi_{(n-2)} = \pi_{(n-3)}^P
\]

Due to the repetition, it can be obtained by

\[
\pi_n = \pi_0 P^n\]

The probability of steady state is a transition probability in the future that does not depend on the initial state. Where this probability is a transitional opportunity that has reached a balance so that it will not be affected by the changes of time. This principle is applied to observe how many states towards a balanced point. So that n-steps are obtained which will become steady state. The steady state conditions are stationary distribution, ergodic, and interlinked property of state.

Based on the research methodology, the steps in the Markov chain application include:

1. Can display and use initial data (existing data).
2. Forming intervals and states, determining state probabilities and transition probabilities, and writing the transition state matrix probability with transfer state.
3. Perform simulations.
4. Analyze the results of the simulation.

3. Results and Discussion

Results of Markov Chain Probability on Blood Sugar Levels

A. Division of 3 State

To find the probability of state \( p_i = \frac{s_i}{n} \) is the amount of data. With \( n = 100 \), then look for the difference from the maximum and minimum values to get the state interval. Data included in S1 as many as 84 data, S2 as many as 16 data, and for S3 3 data. Then divided into 3 intervals and obtained the following results:

| State | Interval | Probability | Score |
|-------|----------|-------------|-------|
| S1    | \( X < 671.76 \) | \( P_1 \) | 0.8909 |
| S2    | \( 671.76,1343.52 \) | \( P_2 \) | 0.092 |
| S3    | \( X \geq 134.52 \) | \( P_3 \) | 0.0172 |

The transition probability of each state to the number of each state, i.e, \( p_{ij} = S_{(i-j)} / S_i \) So that a transition probability matrix can be formed. In Period 1 blood sugar levels enter at intervals S1 and there is no last state transition, so many data including S1 are 81 data, 162 16 data, and for S3 3 data. Here is the probability of transitioning blood sugar levels:

| Transition State | Transition | Probability | Score |
|------------------|------------|-------------|-------|
| S1-1             | 147        | \( p_{1,1} \) | 0.9545 |
| S1-2             | 7          | \( p_{1,2} \) | 0.0455 |
| S1-3             | 0          | \( p_{1,3} \) | 0     |
| S2-1             | 6          | \( p_{2,1} \) | 0.375 |
| S2-2             | 8          | \( p_{2,2} \) | 0.5   |
| S2-3             | 2          | \( p_{2,3} \) | 0.125 |
From table 2, it was obtained transition probability matrix state of 3 x 3 for 3 state:

\[
p = \begin{pmatrix}
0.9545 & 0.0455 & 0 \\
0.375 & 0.5000 & 0.125 \\
0.3333 & 0.3333 & 0.3333
\end{pmatrix}
\]

\[
\pi = \pi_0 * P = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} * \begin{pmatrix}
0.9545 & 0.0455 & 0 \\
0.375 & 0.5000 & 0.125 \\
0.3333 & 0.3333 & 0.3333
\end{pmatrix} = \begin{bmatrix} 0.9545 & 0.0455 & 0 \end{bmatrix}
\]

Next it is obtained the transition probability matrix of n-step 3x3 in steady state:

\[
\pi = \pi P = \begin{pmatrix}
\pi_1 & \pi_2 & \pi_3
\end{pmatrix} = \begin{pmatrix}
\pi_1 & \pi_2 & \pi_3
\end{pmatrix} = \begin{pmatrix}
0.9545 & 0.0455 & 0 \\
0.375 & 0.5000 & 0.125 \\
0.3333 & 0.3333 & 0.3333
\end{pmatrix}
\]

\[
\pi_1 = 0.9545\pi_1 + 0.0375\pi_2 + 0.3333\pi_3
\]

\[
\pi_2 = 0.0455\pi_1 + 0.5\pi_2 + 0.3333\pi_3
\]

\[
\pi_3 = 0.125\pi_2 + 0.3333\pi_3
\]

\[
\pi_1 + \pi_2 + \pi_3 = 1
\]

From the above equation it is obtained \(0.6667\pi_3 = 0.125\pi_2\) therefore \(\pi_3 = 0.1875\pi_2\) from the eliminated equation to \(\pi_1 = 9.6154\pi_2\)

Substitute to the following equation

\[
9.6154\pi_2 + \pi_2 + 0.1875\pi_2 = 1
\]

\[
\pi_2 = 0.092568
\]

\[
\pi_2 = 0.09257
\]

Limited numbers where \(n = 25\) so that each State can be reached from the state that can be achieved from other states with \(n = 25\)

\[
\pi_1 = \pi_2 + \pi_3 = 1
\]

\[
P_{ij}^{25} = \begin{bmatrix}
0.8901 & 0.0925 & 0.0173 \\
0.8901 & 0.0925 & 0.0173 \\
0.8901 & 0.0925 & 0.0173
\end{bmatrix}
\]

**Table 3. The Prediction of Blood Sugar Level 3 State Division**

| No | Month | Prediction | Level |
|----|-------|------------|-------|
| 1  | Period I | X < 671,76 | 209.72 |
| 2  | Period II | X < 671,76 | 309.72 |
| 3  | Period III | X < 671,76 | 241.64 |
| 4  | Period IV | X < 671,76 | 53.87 |

In Table 4.3 it can be seen that the prediction of blood sugar levels in the next 4 periods has a match with 100% data when using a 3-state divider

**B. 4 State Division**

In division 4 state the completion step is the similar to the 3 state divider but the difference is divided by 4 to get the interval of each state

**Table 4. The Prediction of Blood Sugar Level 4 State Division**

| State | Interval | Probability | Score |
|-------|----------|-------------|-------|
| S1    | X < 503.82 | P1 | 0.8506 |
| S2    | [503.82, 1007.64] | P2 | 0.1092 |
| S3    | [1007.64, 1511.64] | P3 | 0.023 |
Then it is obtained the transition probability matrix $n$-steps $4 \times 4$ in steady state by applying similar method it is obtained in $n=21$:

$$P = (1 \ 0 \ 0 \ 0) \cdot \begin{pmatrix} 0.9320 & 0.0544 & 0.0136 & 0 \\ 0.4211 & 0.4211 & 0.0526 & 0.1053 \\ 0.5000 & 0.5000 & 0 & 0 \\ 0 & 0.3333 & 0.3333 & 0.3333 \end{pmatrix} \cdot \begin{pmatrix} 0.0497 & 0.1098 & 0.0231 & 0.0173 \\ 0.0497 & 0.1098 & 0.0231 & 0.0173 \\ 0.0497 & 0.1098 & 0.0231 & 0.0173 \end{pmatrix}$$

From Table 4.5 it is seen that the prediction of Blood Sugar Levels for the next 4 periods for 4 periods is at the first interval $X < 503.8$, so that the prediction results with the existing data are 100% matching at the 4-state division.

### C. 5 State Division

At the division of 5 states the completion step is similar to the division of 3 states but the difference is divided by 5 to get the interval of each state. The following is the interval distribution on Blood Sugar Levels:

#### Table 6. State Division and the Probability of Blood Sugar Level

| State | Interval | Probability | Score |
|-------|----------|-------------|-------|
| $S_1$ | $X < 403.056$ | $P_1$ | 0.7816 |
| $S_2$ | [403.056,806.112] | $P_2$ | 0.1437 |
| $S_3$ | [806.112,1209.168] | $P_3$ | 0.0517 |
| $S_4$ | [1209.168,1612.224] | $P_4$ | 0.0057 |
| $S_5$ | $X \geq 1612.24$ | $P_5$ | 0.0172 |

$$P = \begin{pmatrix} 0.9185 & 0.0593 & 0.0222 & 0 & 0 \\ 0.32 & 0.56 & 0.08 & 0 & 0.04 \\ 0.3333 & 0.2222 & 0.3333 & 0 & 0.1111 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.3333 & 0 & 0.3333 & 0.3333 \end{pmatrix}$$

$$\pi_1 = \pi_0 \cdot P$$

$$P = (0.9185 \ 0.0593 \ 0.0222 \ 0 \ 0)$$

5
By applying $\pi_{i+j} = \pi_i * P$, Next it is obtained transition Probability Matrix $n$- steps $5 \times 5$ at steady state with $n = 28$

$$p_{ij}^{28} = \begin{pmatrix}
0.7803 & 0.1445 & 0.0520 & 0.0058 & 0.0173 \\
0.7803 & 0.1445 & 0.0520 & 0.0058 & 0.0173 \\
0.7803 & 0.1445 & 0.0520 & 0.0058 & 0.0173 \\
0.7803 & 0.1445 & 0.0520 & 0.0058 & 0.0173 \\
0.7803 & 0.1445 & 0.0520 & 0.0058 & 0.0173 \\
\end{pmatrix}$$

Table 7. The Prediction Score of Sugar Blood 5 State Division

| No. | Month   | Prediction  | Production |
|-----|---------|-------------|------------|
| 1   | Periode I | $X < 403.056$ | 209.72     |
| 2   | Periode II | $X < 403.056$ | 309.41     |
| 3   | Periode III | $X < 403.056$ | 241.64     |
| 4   | Periode IV  | $X < 403.056$ | 53.87      |

From Table 7 it is seen that the prediction of blood sugar levels in the next 4 periods is at the first interval $X < 403.056$, so that the prediction results with the existing data have a 100% matching in the 5 state division.

4. Discussion

The dataset that will be applied are is 100 data. Testing of a combination of training and test data will be conducted to determine the greatest level of accuracy of the system. Some combinations are applied to test the dataset shown in the following table.

Table 8. Result of Dataset Combination

| Test  | Combination of The Training Data : Test Data | Accuracy |
|-------|---------------------------------------------|----------|
| 1     | 20:80                                       | 18.75 %  |
| 2     | 40:60                                       | 25 %     |
| 3     | 50:50                                       | 30 %     |
| 4     | 60:40                                       | 37 %     |
| 5     | 80:20                                       | 75 %     |

Based on the results of testing the combination of training data sets and test data 80 and 20 provide the highest average accuracy value, i.e. 75%. Based on the results of testing the dataset in Figure 4.4, the following is a test using 20 test data shown in Table 9.

Table 9. The Result of Test with 20 Test Data

| Parameter | Actual | Result of The System |
|-----------|--------|----------------------|
| Age       |        |                      |
| R         | R      | C                    | 3         | 3         |
| R         | ST     | C                    | 3         | 3         |
| R         | T      | R                    | 2         | 2         |
| SR        | C      | R                    | 3         | 3         |
| R         | C      | R                    | 3         | 3         |
| SR        | C      | C                    | 5         | 5         |
| SR        | C      | C                    | 3         | 3         |
| SR        | R      | ST                   | 3         | 3         |
| SR        | C      | R                    | 3         | 4         |
| SR        | C      | R                    | 3         | 3         |
| SR        | C      | R                    | 5         | 4         |
| SR        | C      | C                    | 3         | 3         |
| SR        | C      | R                    | 2         | 2         |
| SR        | R      | R                    | 4         | 0         |
| C         | T      | R                    | 5         | 0         |
Based on the results of testing using 20 datasets showed an accuracy of 75%. The resulting accuracy is still quite good, this is because the Markov chain probability process will lead to a steady state condition which means that after the process has been running for several periods, the probability of status will be fixed and this is called the steady state probability. Repeatedly shown that the number of columns of the probability of transition in each row of the transition matrix is one. If all the number of matrix columns are equal to one, the transition matrix is called Double Stochastic. For each double stochastic transition matrix where the number of states is m, then each steady state probability is 1 / m.

5. Conclusion

The decision method of Probability Markov Chain is formulated to determine the menus for diabetics from states formed from processed data training. Based on the results of testing the menu determination system for diabetic people by applying the Markov Chain Probability Method. While the best accuracy when testing a combination of data training and testing on the number of data training 80 data and the number of testing data 20 data were obtained. Dataset of training dataset are not good so the former state is not good enough so that when testing with new data it will get inappropriate results. The Probability of Markov Chain method carries out a decision based on the state formed from the results of state formation. When there are several data that have similarities and different class results, only one class will be formed so that when tested the results of the system cannot classify the data that will be due to no state.

References

[1] Sulaksono, J, Jauhari. M H & Hariri. F R, 2014, “Sistem Pendukung Keputusan Penentuan Penyakit Diabetes Millitus dengan Menggunakan Metode LVQ”, Seminar Nasional Teknologi Informasi dan Multimedia.

[2] BJK/AIS, “www.liputan6.com,” sctv, 22 Dec 2011. [Online]. Available: https://www.liputan6.com/health/read/368590/diabetes-melitus-indonesia-duduki-peringkat-ke-4-dunia. [Accessed 21 6 2018].

[3] Tandra H, (2008), Segala sesuatu yang harus Anda ketahui tentang Diabetes, Penerbit Gramedia, Jakarta

[4] Ching, Wai-Ki., K Ng, Michael. 2006. Markov Chains model, Algorithms and Application, Hongkong.

[5] Rizanti I N dan Soehardjoepri, 2017, “Prediksi Produksi Kayu Bundar di Kabupaten Malang dengan Menggunakan Metode Markov Chains”, Jurnal Sains dan Seni ITS, Vol 6, No 2

[6] Allo G D, Hadtija D & Paendong. M, 2000, “Analisis Rantai Markov untuk mengestahui peluang perpindahan merek kartu seluler prabayar GSM”, Jurnal MIPA UNSRAT.

[7] Fernanda. S. I, Ratnawati. D. K & Adikarya. P. P, 2017, “Identifikasi Penyakit Diabetes Millitus menggunakan Metode MKNN”, Jurnal Pengembangan Teknologi Informasi dan Komputer, Vol 1, No 6, pp. 507-513.

[8] Santoso. H F, Farwiyanto. E & Wesety U N , 2016, “Analisis dan Implementasi Metode Certainty Factor pada Sistem Pendukung Keputusan untuk penyakit Diabetes Millitus”, e-Prosiding of Engineering, Vol 3, No 3, pp. 5189.