Patient queue systems in hospital using patient treatment time prediction algorithm

Dwi Putri Handayani1, Mustafid2, Bayu Surarso3
Universitas Diponegoro, Indonesia1,2,3

Abstract
Patient Treatment Time Prediction Algorithm was very important to build an outpatient queue system at the hospital. This study aims to build a system of outpatient queues to predict the waiting time of outpatients in the eye clinic at one of Cirebon hospitals. Patient Treatment Time Prediction algorithm was calculated based on historical data or medical records of patients in the hospital with 120 patient data. The Patient Treatment Time Prediction algorithm was trained by improved Random Forest algorithm for each service and a waiting time for each service. Prediction of waiting time for each patient service was obtained by calculating the consumption of patient care time based on patient characteristics. The waiting time for each service predicted by the trained Patient Treatment Time Prediction algorithm is the total waiting time of patients in the queue for each service. This research resulted in a system that can show the time taken by patients in every service available in the eye clinic. Patient time consumption in each service produced varies according to the patient's condition, in this case based on the patient's gender and age. This research provides benefits in terms of improving performance in each department involved, optimizing human resources, and increasing patient satisfaction. This research can be developed for each department in the hospital.

1. Introduction
Outpatient queue system becomes very important in the success of hospital management which in general the hospital has not implemented a patient queue system so that it increases patient waiting time at the hospital [1]. The patient's desire to be outside the waiting room must be broken by the uncertainty of the waiting time so that the patient cannot leave the place so that it does not pass his turn [2]. Prolonged waiting times reduce the quality of care and increase side effects and sometimes cannot be changed [3]. Long waiting times for patients are usually seen outside patient facilities, and these difficulties contribute to various public health problems, including disruption of access to care, disruption of hospital work patterns, and patient dissatisfaction [4]. The uncertain arrival of outpatients significantly influences the daily operation of health facilities [5]. According to [6] based on research that has been done, cancer patients in Canada experience a long waiting time so will have a negative clinical impact. Most hospitals, patient queues that encourage and determine each tempo of the workflow, so the management of patient waiting time is an important thing [7]. One of the six principles is that it can reduce waiting times which sometimes delays can have a negative impact on those who receive and provide care [8]. A management in a hospital requires an increase in service, especially reducing waiting time to increase patient satisfaction. However, little is well documented, and its effects are rarely measured by robust methods [9].

Patient satisfaction will increase if it is able to predict waiting times accurately and delay scheduled appointments and allow nurses to serve patient anthens more accurately [10]. Linear regression models have been developed to predict waiting times with an average yield of 64.6 minutes. The average absolute difference between the actual and estimated waiting times is 29.0 minutes. Thus, this difference is the reason for completing research on this subject with other algorithms that can increase the inaccuracy can be accepted by patients and increase satisfaction more than providing information about waiting times [11]. Long waiting time can reduce the quality and satisfaction of patients. Among the US Emergency Department in 2010, only 31% achieved the appropriate triage target, while only 48% admitted that it took 6 hours to service patients [12]. Research using the Dartmouth Microsystem Improvement Curriculum (DMIC) framework and the Plan-Do-Study-Act (PDSA) get the hypothesis that it will reduce waiting time in the targeted waiting room by 20 minutes, whereas in the doctor's examination service is 10 minutes. However, the results of the research that has been done, the average waiting time in the waiting room for patients is 5.33 minutes shorter than the pre-implementation period. The waiting time at the doctor's examination is 1.81 minutes [13]. Male patients tend to be more satisfied despite the delay, while female patients also tolerate the delay but demand clearer time information [14].
One of the innovative categories of queue management is to use a queue system to eliminate the waiting process for patients. The patient queue system provides scheduling information to patients before arriving at the service point, so patients can wait at home rather than in the hospital lobby [15]. The patient queue system must be able to manage the patient's time and recommend the arrangement of services to be performed by the patient and still consider the time of arrival of the patient [16].

The algorithm that can be used to solve the problem described above is the Patient Treatment Time Prediction algorithm [17]. Patient queue system using Patient Treatment Time Prediction will facilitate patients in registering, knowing the order of services that must be passed, and knowing how much time he will take to perform a series of services at the hospital, including waiting time and service time so that patients can take advantage of waiting time well. This patient queue system applies to non-emergency and outpatient patients who usually had to get services such as blood pressure checks, treatment at the clinic, checking blood in the lab, undergoing x-rays, taking drugs, payment, and so on [18]. Patient Treatment Time Prediction algorithm will calculate based on historical data or medical records of patients in the hospital. Patient Treatment Time Prediction algorithms are trained based on the Random Forest algorithm which is improved for each service, and the waiting time for each service [19]. However, with the limitations of the original Random Forest algorithm and hospital characteristics, the Random Forest algorithm needs to be improved in order to get effective results [20].

2. Research Method

Patient Treatment Time Prediction algorithm is generated based on historical hospital data. The waiting time of each service is predicted by the Patient Treatment Time Prediction taken from the number of patient waiting times in line. Later, Chen et al. Built a Queuing-Recommendation Hospital system that featured the system’s recommendation on efficient and comfortable service plans with the least amount of waiting time for patients [17].

Patient waiting time can be calculated when the duration of service time of each patient is known. The duration of each patient's service time will vary according to the patient's condition. Therefore, the Random Forest algorithm is used to train time duration data based on the patient's characteristics to build the Patient Treatment Time Prediction. The Random Forest algorithm is enhanced in 4 aspects to build Patient Treatment Time Prediction.

First, all the selected data variables are sorted from unnecessary features, such as the patient's name, address, and phone number. Secondly, the duration of service time is the variable that is targeted in this process which is a continuous variable, so the CART model will be used. However, there are also independent variables that are of nominal data, such as time span (0-23) and days (Monday-Sunday). In this case, the two-fork tree model of the traditional CART could not fully produce a good analysis. Therefore, to build a regression tree that appropriately required a multi-branch model in the construction process. Furthermore, the process of creating a regression tree must go through identifying the defective data and then delete it to reduce its influence on accuracy. Lastly, the Random Forest algorithm uses a traditional voting in selecting a sample data that allows it to retrieve the incorrect data. Patient Treatment Time Prediction improves how to select the data to be used. Therefore, classifications increase the accuracy of the classification of the Random Forest algorithm and reduce errors. So, the Patient Treatment Time Prediction is superior in terms of accuracy and performance than the Random Forest algorithm. Details of the Patient Treatment Time Prediction process to calculate the total waiting time of the patient are presented as a flowchart in Figure 1.

At this stage as Equation 1, the data needed from the hospital is collected, such as patient data and service data.

\[ S = \{s_1, s_2, ..., s_N\} \]  
(1)

The hospital provides several interconnected and unrelated services during one day visit and each patient can get several services, can be seen in Equation 2.

\[ X|s_i = \{x_1, x_2, ..., x_K\} \]  
(2)

That Equation 3 states the set of patient service data. X is the data of each service, such as registration, examination, taking drugs, and others.

\[ Y|x_i = \{y_1, y_2, ..., y_N\} \]  
(3)

The patient data set with the treatment plan consists of \( y_i \), that is data for each patient doing some service. The data that will be used for the Patient Treatment Time Prediction process must be tested. The duration of service time will differ depending on the condition of the patient, so the duration of service time of the patient cannot be measured with absolute standards. Improvised Random Forest algorithm to improve accuracy of Patient Treatment Time Prediction. Some patient data sets from the original data are taken at random to be sampled, can be seen in Equation 4.
Equation 5 showing, data not selected as samples will also be organized into Out-of-Bag (OOB) data.

\[ S_{\text{OOB}} = \{s_{\text{OOB}1}, s_{\text{OOB}2}, \ldots, s_{\text{OOB}k}\} \]  

**Figure 1. Flowchart Patient Treatment Time Prediction**

Service duration is the target variable in this process which is a continuous variable, so that the single decision tree type in the Random Forest model used is a regression tree. Thus, the CART regression tree model is created for each \( S_{\text{train}} \) test data set. The first optimization aspect of the Random Forest algorithm is the process of making each CART tree. All M features of each \( S_{\text{train}} \) test data are used in the testing process. For each tree node separation process, each feature variable \( y_j \) is required and each potential split point value \( v_p \) of \( y_j \) is chosen to calculate the loss function \( (y_j, v_p) \), which is defined as follows Equation 6.

\[
(y_j, v_p) = \arg \min \left[ \sum_{x \in R_l(y_j, v_p)} (y_i - c_L)^2 + \sum_{x \in R_R(y_j, v_p)} (y_i - c_R)^2 \right]
\]

\( y_j \) is every feature of the test data, \( 1 \leq j \leq M \). \( v_p \) is any potential separator point value of \( y_j \). \( R_L(y_j, v_p) \) is the first data set (left), whereas \( R_R(y_j, v_p) \) is the second data set (right). \( c_L \) as the average value of the set \( R_L(y_j, v_p) \), whereas \( c_R \) the average value of the set \( R_R(y_j, v_p) \).

Noisy data may still exist even if it has been deleted at the preprocessing stage. Therefore, the third optimization aspect of the Random Forest algorithm is to reduce the effect of noisy data on the accuracy of the algorithm. This data removal method is done in calculating the value of each CART node. Data on leaf nodes is currently sorted in ascending order. Then, the values of the three data points \( Q1, Q2, \) dan \( Q3 \) are calculated. The inner limit of noisy data is determined as follows Equation 7.
The outer limits of noisy data are determined as follows Equation 8.

\[\text{OL} = Q_3 + 1.5(Q_3 - Q_1)\]  
(8)

Data that is outside of the range of \(\{IL; OL\}\) will be deleted as bad data. After construction a tree and get a node, the average value \(c_j\) from data \(y_j\) counted at each leaf node of the regression tree. The calculation equation is defined as follows Equation 9.

\[c_j = \frac{1}{k} \sum y_j, (IL \leq y_j \leq OL)\]  
(9)

The following is the pseudocode of the Patient Treatment Time Prediction algorithm:

**Algorithm:** Patient Treatment Time Prediction Algorithm Process based on Random Forest

**Input:** the test data set 
**Output:** decision tree

1: make a data sample
2: create a test data set that is not used
3: make an empty tree
4: for each independent variable in the do test data sample
5: calculate split point candidates
6: for each candidate do split point
7: calculate the best split point
8: end for
9: add a node
10: separate data for the left branch
11: separate data for the right branch
12: for each data in the right and left branches do
13: calculate the best split value of the left branch
14: if the best split value is left or right branch> best split point then
15: add node sections
16: separate data into two left and right forks / branches
17: else
18: collect data for leaf nodes
19: calculate the main value of a leaf node
20: end if
21: end for
22: end for

3. Results and Discussion

The author has analyzed the time consumption of patient services in the eye clinic at one of the hospitals in Cirebon based on the time factor and patient characteristics. Based on the contents of the activity and the circumstances, the consumption of patient care time on care assignments in each service can vary. In fact, time consumption in the same service may be different, different officers, different periods, and different patient conditions.

This study uses historical data of patients who perform services at the eye clinic at one of Cirebon hospitals. The data used are patient data in the period from October 1, 2019 to October 7, 2019 with a total of 120 patients. Patient data consisting of age and sex are complemented by the day of service, the time when the patient starts receiving service, and the time when the patient has finished receiving service. There are three services to be analyzed, namely registration, check-up, and service by a doctor. So, the amount of data used in this study is 360 data.

The first step to the node separation process is required for each feature variable \(y_j\) and any potential split point values \(v_p\) from \(y_j\) chosen. From the available feature variables, the potential split point with gender. The next feature variable that can be used for the split point is the patient's age. The split point based on age can be determined by Equation 6. The split point based on the age of the patient will be obtained based on the results of Equation 6 which has the smallest value. \(c_L\) interpreted as the average value of the set \(R_L(y_j, v_p)\), while \(c_R\) interpreted as the average value of the set \(R_R(y_j, v_p)\). Then,
\[ c_L = \frac{\text{Total age of the patient in the set } R_L}{\text{Total data in the set } R_L} = \frac{2721}{50} = 54.4 \]

\[ c_R = \frac{\text{Total age of the patient in the set } R_R}{\text{Total data in the set } R_R} = \frac{3987}{70} = 57 \]

After got the average value of each set, the age of the patients in each set is processed with \( c_L \) and \( c_R \) for patient in set \( R_L \), then it will produce a value. If the value is the smallest of each set, then the age of the patient who has the smallest value will be used for the split point candidate in each set. Example calculation to get the smallest value on the set \( R_L \) with ID P089 aged 15 years, while on the set \( R_R \) with ID P020 aged 6 years.

ID P089 were 15 years old, subtracted by the average age of patients in the \( R_L \) set, which was 54.4. After that the results are squared.

\[ (y_i - c_L)^2 = (15 - 54.4)^2 = 1553.9 \]

ID P020 were 6 years old, reduced by the average age of patients on the \( R_R \) set, which is 57. After that the results are squared.

\[ (y_i - c_R)^2 = (6 - 57)^2 = 2596.6 \]

From the results of calculations the \( R_L \) set, the smallest value obtained is 0.3 in patients with ID P113 who have 55 years of age. Thus, the split point candidate in the age of patients in the \( R_L \) set is 55. Whereas, in the \( R_R \) set, the smallest value obtained is 0 in patients with ID P120 who have 57 years of age. Thus, the split point candidate in the age of patients in the \( R_R \) set is 57.

After getting a split point candidate at each set, make the candidate a separator between members of the set. Separate members of the \( R_L \) set into left and right sets, as well as the \( R_R \) set.

The left \( R_L \) set has 23 patient data, while the right \( R_L \) set has 27 patient data. The left \( R_R \) set has 30 patients and the right \( R_R \) set has 40 patients. After this grouping, count again in the same way to find the best split point. If you do not meet the \( \Phi(v_p|y_j) \geq \Phi(v_p|y_j) \) requirements then continue counting to find the best split point.

\[ c_L = \frac{\text{Total age of the patient in the left } R_L \text{ set}}{\text{Total data in the left } R_L \text{ set}} = \frac{1025}{23} = 44.6 \]

Patients with ID P089 are 15 years old. reduced by the average age of patients in the left \( R_L \) set, which is 44.6. After that the results are squared:

\[ (y_i - c_L)^2 = (15 - 44.6)^2 = 876.2 \]

After all members of the left \( R_L \) set are calculated, this calculation gets a split point with a value of 45 in patients with ID P076 which means this point does not meet the requirements greater than the candidate split point, then the calculation looks for the best split point followed by counting on the right \( R_L \).

\[ c_L = \frac{\text{Total age of the patient in the right } R_L \text{ set}}{\text{Total data in the right } R_L \text{ set}} = \frac{1696}{27} = 62.8 \]

Patients with ID P039 are 56 years old. reduced by the average age of patients in the right \( R_L \) set, which is 62.8. After that the results are squared.

\[ (y_i - c_L)^2 = (56 - 62.8)^2 = 46 \]

After all members of the right \( R_L \) set are calculated, this calculation gets a split point at age 63 which means it has met the greater requirements than the split point candidate of 55. Thus, 63 in patients with ID P097 becomes the best split point for the \( R_L \) set. An overview of the trees formed as a result of finding the best split point set of \( R_L \) is presented in Figure 2.

Figure 2. The Multi-Branch Regression Tree Male Gender of the PTTP Model

Next, calculate the best split point in the $R_R$ set.

$$c_L = \frac{\text{Total age of the patient in the left } R_R \text{ set}}{\text{Total data in the left } R_R \text{ set}} = \frac{1338}{30} = 44.6$$

Patients with ID P020 are 6 years old, reduced by the average age of patients in the left $R_R$ set, which is 44.6. After that the results are squared.

$$(y_i - c_R)^2 = (6 - 44.6)^2 = 1490$$

After all members of the left $R_R$ set are calculated, this calculation gets a split point with a value of 45 in patients with ID P106, which means that this point does not meet the requirements greater than the split point candidate of 57, then the calculation looks for the best split point followed by counting on the right $R_R$ set.

$$c_R = \frac{\text{Total age of the patient in the right } R_R \text{ set}}{\text{Total data in the right } R_R \text{ set}} = \frac{2649}{40} = 66.2$$

Patients with ID P083 are 58 years old, reduced by the average age of patients in the right $R_R$ set, which is 66.2. After that the results are squared.

$$(y_i - c_R)^2 = (58 - 66.2)^2 = 67.24$$

This calculation gets a split point at age 67 meaning that it has met the greater requirements than the candidate split point that is 57. So, at point 67 precisely in patients with ID P112 being the best split point for the $R_R$ set. An overview of the trees formed as a result of finding the best split point set of $R_R$ is presented in Figure 3. be the best split point.

Figure 3. The Multi-Branch Regression Tree Female Gender of the PTTP Model

The next step is to erase noisy data that might still be present in the set before getting the leaf node value using Equation 7 and Equation 8. For node T1.

© 2020 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)
Q1 is the first data on the node set T1, while Q3 is the last data on the node set T1. IL is the fence boundary in the set of vertices T1 that yields a value of -27. Whereas, OL is the outer fence boundary of node T1 which produces a value of 85. Every data that is between the inner fence -27 and the outer fence 85, then the data is normal data, which means it is not noisy data so it does not need to be deleted.

Noisy data that needs to be deleted also applies to the \( R_r \) set which is likely to still have noisy data on the set before getting the leaf node value using Equation 7 and Equation 8.

\[
Q1 = 6; Q3 = 44
\]

\[
IL = Q1 - 1.5(Q3 - Q1) = 6 - 1.5(44 - 6) = -51
\]

\[
OL = Q3 + 1.5(Q3 - Q1) = 44 + 1.5(44 - 6) = 101
\]

IL is the fence boundary in the set of vertices T4 which yields a value of -51. Whereas, OL is the outer fence boundary of the T4 node set which produces a value of 101. Every data between the inner fence -51 and the outer fence 101, then the data is normal data, which means that it is not noisy data so it does not need to be deleted.

Patient waiting times at each service are predicted after being grouped into a node. The time value of each patient in the node will be calculated by the Equation 9.

Calculation of prediction of patient waiting time in the examination service by doctors with Equation 9 will be presented from one node of the male set and one node of the female set.

\[
c_j = \frac{8 + 8 + 7 + 5 + 6 + 5 + 7 + 7}{8} = 6 \text{ minutes}
\]

At node T1 which is a node in the male gender regression tree for doctor's examination services. This calculation yields 6 minutes meaning that for male patients aged between 15-43 years it takes 6 minutes when doing a doctor's examination.

\[
c_j = \frac{9 + 8 + 7 + 9 + 7 + 6}{7} = 7 \text{ minutes}
\]

At node T4 which is a node in the female gender regression tree for the doctor's examination service. This calculation yields 7 minutes means that for female patients aged between 6-44 years it takes 7 minutes when doing a doctor's examination.

The results of calculations with Equation 9 will form a tree node waiting time for patients at the examination service by a doctor which will be presented in Figure 4.

![Figure 4. PTTP Model Regression Tree in the Examination Service by a Doctor](image)

4. Conclusion

The results and discussion of this study indicate that the outpatient queue system that was built with the Patient Care Time Prediction algorithm is able to show the prediction of the waiting time needed by the patient as an estimated...
time to get service at each service that the patient will pass. Outpatients will be able to know when it is their turn to get services in the patient queue. This study uses historical data from one hospital in Cirebon, especially in the eye clinic that can only be inputted by the admin. Patients undergo several services, namely registration, check-up by nurses and examinations by doctors. First, the Patient Treatment Time Prediction algorithm classifies patients based on sex and age. So, this algorithm produces a prediction of waiting time in each patient classification. Testing of this system has been carried out. Test results show the system can function properly according to design. The system can perform calculations using the Patient Care Time Prediction algorithm to get the service time that will be taken by each outpatient. The results of manual calculations with calculations using the system have the same results so the system can be said to be valid. The data used were 120 patients.

This outpatient queue system only applies to dependent services. Suggestions for further research development are wider use in other departments in the hospital, as well as independent services so that they can provide recommendations in terms of service to patients to reduce waiting times at the hospital.

**Notation**

The example of notation can be described with the following description:

- \( N \) : the number of data
- \( S \) : collection of patient data
- \( X \) : the data of each service
- \( K \) : the number of data \( X \)
- \( Y \) : the patient data set with the treatment plan
- \( M \) : the number of data \( Y \)
- \( S_{\text{train}} \) : collection of patient data chosen to be sampled
- \( S_{\text{OOB}} \) : Data not selected as samples will also be organized into data
- \( y_j \) : the number of \( S_{\text{train}} \) data and \( S_{\text{OOB}} \) data
- \( v_p \) : required and each potential split point value \( v_p \)
- \( v_p \) : potential separator point value of \( y_j \)
- \( R_L(y_p, v_p) \) : the first data set (left)
- \( R_R(y_p, v_p) \) : the second data set (right)
- \( c_L \) : the average value of the set \( R_L(y_p, v_p) \)
- \( c_R \) : the average value of the set \( R_R(y_p, v_p) \)
- \( Q_1 \) : the bottom data point
- \( Q_2 \) : the median data point
- \( Q_3 \) : the top data point

**References**


© 2020 The Authors. Published by Universitas Muhammadiyah Malang

This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)


