Designing Classification Models of Patron Visits to an Academic Library using Decision Tree

Aisyah Larasati[1,4*], Muhammad Farhan[1], Puji Rahmawati[1], Nabila Azzahra[1], Apif Miftahul Hajji[2,4], Anik Nur Handayani [3]

[1] Department of Industrial Engineering, Universitas Negeri Malang, Jl. Semarang No.5 Malang - Indonesia
[2] Department of Civil Engineering, Universitas Negeri Malang, Jl. Semarang No.5 Malang - Indonesia
[3] Department of Electrical Engineering, Universitas Negeri Malang, Jl. Semarang No.5 Malang - Indonesia
[4] PUI-PT Disruptive Learning Innovation (DLI) Universitas Negeri Malang, Jl. Semarang No.5 Malang - Indonesia

[*]Corresponding author: aisyah.larasati.ft@um.ac.id

Keywords: decision tree, academic library, classification models, patron visits

Abstract. Classification models of patron visits in library may help the library to reveal factors that affect patron visit and to predict how frequent a patron visit the library. This research aims to design classification models of patron visit in Library of Universitas Negeri Malang using decision tree model. Data is collected using online and offline surveys. The total number of usable responses are 883, in which 402 of the responses collected through on-line survey and 481 of the responses collected through a direct survey at the library area. The sampling method is a convenience random sampling. The classification model is built using Decision tree model. The model accuracy of the classification model is 87.5%. The result shows that a library customer tends to visit library more often when they have an assignment or need references for their thesis/final project. In contrast, a self-motivated patron tends to rarely visit library. This study finds nine attributes that highly affect the frequency of customer visit to the academic library are semesters, faculty, department, internet service, bag storage, reading rooms, OPAC services, staff services and the last is book collection.

Introduction

The existence of a library cannot be separated from human culture. Moreover, in ancient times, people recorded their knowledge to be remembered and conveyed to others. They used body language or images to express thoughts and feelings. Also, they used body language or images as communication with others. Hence, the existence of the library began to be needed. The invention of printing machines, the development of recording techniques, and the development of digital technology Information and communication-based technology cause the library to grow faster. The complexity of library management is increasing. Library as a management system of ideas, thoughts, experiences, and knowledge of mankind, has a main function; to preserve the result of human culture, particularly in the form of printed documents and other record works, also to transform ideas, thoughts, experiences, and knowledge from one generation to the next. The target of implementation of this function is the formation of a society that has a culture of reading and learning at all times.

Faced with an increasingly dynamic environment and changes, libraries are beginning to realize that there is a vast largely untapped assets diffused in the organization – knowledge [1]. The

conventional image of the library known as a quiet place to read, most of the collection in the library are printed books, no internet access, all of those concept is changing to a modern image of library. This change caused by new challenges that library has to face. This changing has a huge effect of the library service, with the existence of new technologies in computers, internet facilities around the library, e-book collection, etc.

University library user consists of various behavior, depending on the purpose of visiting the library, how often the patron visits the library in a week or a month, etc. In order to reach the effective system of library, that is to help users/patrons to find information more easily, it is necessary to understand the patron's behavior. By knowing more about behavior of patrons, we could understand the needs of patrons, and how to respond to those needs. The behavior of patrons also have a role as suggestions to reach the improvement of the library. We use data mining to process the patron's behavior data.

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can consist of data warehouse, databases, the web and other information repositories or data that streamed into the system dynamically [2]. Data mining is used to classify which useful or useless information from large datasets and change it to an input that has a visualization that is easy to interpret. One of the most common-used methods for data mining is decision tree. This method is easy to use, no ambiguity, and it is robust even in the presence of missing values. Target variables or independent variables can use both discrete and continuous variables. Assessing the relative importance of variables is one of the reasons for using the decision tree. Sometimes when relevant variables are identified, researchers need to know which variables have the major role. Generally, the importance of variable is known by computing based on the reduction of model accuracy (or in the decision tree, in the prioritized of nodes in the tree) when the variable is removed. In this case, we use data mining to classify the patron visit using decision tree model. In most circumstances the more records a variable have an effect on, the greater the importance of the variable [3].

It is important to explore the patron's behavior (how often the patron visit the library in a week, how many books does the patron read in one time visit, etc.) in order to understand the patron needs, so that library understand which factors should be prioritized in order to increase the library's service performance. In addition, understanding the patron behaviour help the library improving its service quality effectively, as a result the patron revisit intention gets higher. In the academic library, the patron experiences during their visit to the library are closely related to their learning outcome in the university [4].

The aim of this study is to classify the library visitors according to the dependent variable that exists, in this case, the dependent variable is how often the person comes to visit the library within the specified time as the parameter.

Literature Review

Library are a collection of materials that exist as physical system which for many library users, especially in academic institutions, size, scope and Collection currencies are more important than other aspects service [5]. Library as a physical space to enable and facilitate learning is really important to higher education learning environment [6]. In this age of digital era, finding references for academic research or completing assignment is making direct visits to the library decreased [7]. So, this study was conducted to determine the frequency of patron visits whether they included frequent, rare or erratic visits and how it was influenced by several other factors.

The first factor is motivation. From research on motivation that has been carried out in the library field, most of it focuses on patron behavior or habits such as patron behavior when looking for information [8]. Usually, motivation comes from needs. There are five categories of needs: 1) psychology needs; 2) Safety needs; 3) Belongingness needs; 4) Esteem needs (build a positive-

image and appreciation), and 5) Self-actualization needs [8,9]. Besides of that, there is a need for achievements. Students who have a high need for achievement tend to search more information for their learning tasks (e.g. by searching literature that can answer or relevant with their needs in library or online searching) [8].

On the other hand, reading habits have a correlation with the motivation of students to come to the library such as students who have poor reading habits are not most likely to be library users but, students who have good reading habits are very likely to be library users [10]. But, reading habits will have no effect if the collection of books provided to patrons is inadequate. Because Customer satisfaction is achieved by not only providing the required facilities, services but also the collections of a library [11]. Patrons also have their own criteria for choosing the books that are needed and desired. However, in the presence of these criteria, from the customer's perspective, This criterion has caused most collections of library to be far greater than the average customer's needs [12]. Customers who are satisfied it's depends on their perception of perceived service and mostly they are would like to re-visit the library [13].

The second factor is the quality of service (e.g. Type of service and Collection of books). There are four dimensions to measure library user's perceptions of service quality [11,14] such as : 1) Access to information; 2) Individual control which relates to the attitude of library staff to the customer; 3) Library as place, like a symbol of intellect, that related to equipment, facilities and infrastructure (library environment); and 4) Affective service (e.g. empathy, responsiveness, and assurance), which related to services offered by the library. The frequency of patron visits may reflect patrons' perception and expectation on the library service quality.

First of all, the expectations of patrons are described from patrons' perception about how hard it to get/find the library products, and how likely it is that the patrons can get access to the product that they need. Second, when patron needs for some product, they believe that they can find the product through access to the product library. Third, various product libraries deliver different satisfaction for patrons to meet their information needs [12]. After using a service, patrons evaluate and provide assessment of the service based on their experiences [11]. Patrons are motivated to use existing products more often in the library if they can use all services in the library either through online or direct visits. However, patrons' motivation to use library products and service tend to be very low, if their perceptions to get an access to products and service are difficult. This happens because patrons have not had any experience with all of library product and services[12]. Thus, it becomes important to reveal the behavior of patrons who tend to frequently or rarely visit the academic library. One approach that enable to classify patrons such that is a classification method using a decision tree model. This model is one of the most machine learning or method that effective, easy to use, has no ambiguity, and robust even there is a missing value for classification problems [15]. Data that can be analyzed using decision tree are categorical or numerical, in which each categorical classes has to have a record [16]. The structure of a decision tree is shown in Figure 1.

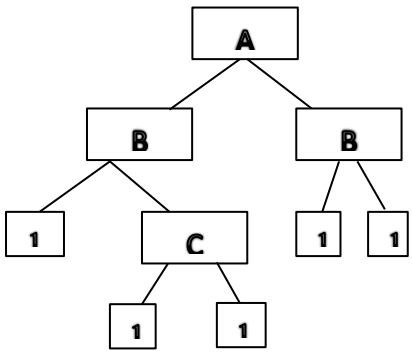

Figure 1. Decision Tree Classifier

Research Methodology

This research aims to classify patrons that come to the academic library at Universitas Negeri Malang based on their frequency of visits. This research designs the classification models of patron visit in the academic library of Universitas Negeri Malang using decision tree model. Data is collected using both online and offline surveys. The total number of samples used in the study are 921 respondents. The usable responses are 883, in which 402 of the responses collected through on-line survey and 481 of the responses collected through a direct survey at the academic library area. The sampling method is a convenience random sampling.

The flowchart of the research process in this study is shown in Figure 2.

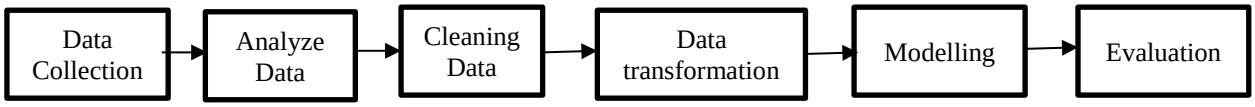

Figure 2. Flow Chart of the Research

This study collects data using questionnaire and google form which contain five indicators and have 27 attributes. The direct survey is performed using accidental sampling technique. The indicator variable used in this study is shown in Table 1. Respondent profile attribute includes name, faculty, semester, and department.

Table 1. Indicator in questionnaire

Variable	Indicator	Questionnaire item
Independent variable	Respondent Profile	4 profile question
	Motivation	Item 3
	Type of services	Item 7
	Books collection	Item 16
	Service Quality	Item 20, 23, 24, 25, 26
Dependent variable	Frequency of visits to the library	Item 10

The respondents used were students from Universitas Negeri Malang. The distribution of respondents based on their faculty is shown in Table 2.

Table 2. Distribution of respondent faculty

No	Faculty	Total	Percentage
1	Faculty of Economics (FEc)	146	16.53%
2	Faculty of Sports Sciences (FSpS)	19	2.15%
3	Faculty of Education (Fed)	67	7.58%
4	Faculty of Social Sciences (FSoS)	66	7.47%
5	Faculty of Mathematics and Natural Sciences (FMNS)	122	13.81%
6	Faculty of Psychology Education (FPE)	38	4.30%
7	Faculty of Letter (FL)	81	9.17%
8	Faculty of Engineering (FEng)	344	38.95%

According to Table 2, the data distribution is quite good because data has similar proportion to the total number of students in each faculty at Universitas Negeri Malang.

Data is cleaned to eliminate missing value in the data. The usable rate is 95.87%. From total 921 samples, only 883 data can be processed. Data is then transformed to simplify the model of the decision tree; the transformed data can be seen in Table 3.

Table 3. Data transformation

No	Attribute	Number of class	Transformation number of class
1	Motivation to the library	5	2
2	Frequently used services	9	3
3	Frequency of library visits every month	5	2
4	Library collection	5	2
5	Internet services	5	2
6	Benefits of OPAC	5	2
7	Comfort of library space	5	2
8	Library staff services	5	2
9	Security storage bags	5	2

In order to build the decision tree model, data is split into two data partitions: training datasets (80%) and testing datasets (20%) to avoid over fitting models. The parameters of decision tree are: the selection criterion= gain ratio, the maximal depth = 10, pruning is applied with confidence = 0.1, pre-pruning is applied with minimum gain = 0.01, minimum leaf size = 2, minimum size for split = 4, and total number of pre-pruning = 3.

Results and Discussion

Decision Tree is formed from independent variables and dependent variables. Independent variables predictor attributes to determine the classification of library visitors. While the dependent variable is the target attribute, in the form of frequency of visits to libraries which are classified as frequent (F) and rare (R). The Decision Tree model is shown in Figure 3.

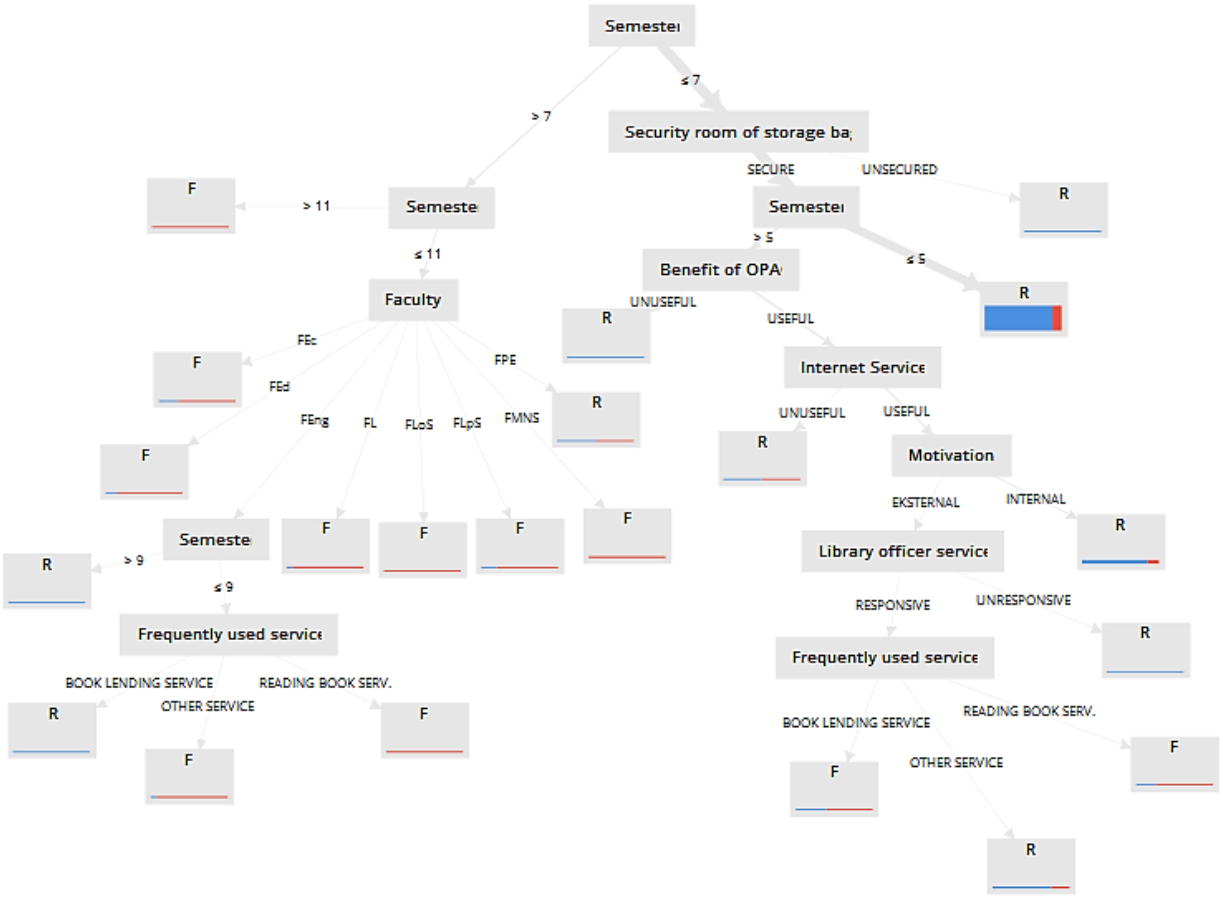

Figure 3. The Resulted Decision Tree

The results of the decision tree indicate that students in the 7th to 11th semester are classified as students who often visit the library. While the Faculty of Engineering students who are under the 9th semester, with services that are often used are books borrowing tend to rarely visit the library, while those who use reading services and other services tend to often visit the library. Students who are under the 7th semester tend to rarely visit the library, with the perception that the storage bags room is safe, the benefits of OPAC services, The helping from Internet services, motivation to the library mostly comes from external factors, library staff services that has high responsiveness and specifically for services that are often used reading books services and books lending services are classified as students who often visit the library.

Accuracy of Decision Tree is shown in Table 4. Accuracy of training data is 87.27% and accuracy of testing data is 87.5%. Accuracy of training data and testing data tends to be consistent so there is no indication of over fitting. From the results of testing data that is quite high it can be concluded that the data can be used to predict students who often (F) visit the library and who rarely (R) visit the library.

Table 4. The Model Accuracy

Data partition	Training Data Set (80%)		Testing Data Set (20%)	
	Count	%	Count	%
Correct Classification	617	87.27%	154	87.5%
Misclassification	90	12.73%	22	12.5%
Total	707		176	

Based on the results of the analysis using the decision tree, the attributes that influenced the frequency of customer visits to the library are shown in table 5. Table 5 shows the 10 most influential variables on the frequency of customer visits to the library. These results indicate that the main factors that influenced the frequency of customer visits to the library are semesters.

Table 5. The attributes that most influence the frequency of customer visits to the library

No	Attributes	Importance (%)
1	Semesters	44.45%
2	Faculty	15.59%
3	Department	13.69%
4	Benefits of Internet Service	4.63%
5	Security room of storage bag	4.44%
6	Convenience of reading rooms	4.39%
7	Benefits of OPAC services	4.38%
8	Library officer service	2.41%
9	Book Collection	1.62%

Conclusion

According to what has been classified before, students who have a high frequent of visit to the library is students from semester 7^{th} until 11^{th}, also students from Faculty Engineering which are currently in semester 9^{th} and under. Moreover, students with low frequent of visit to the library is students who are under the 7^{th}. Factors of these results are; the storage room's safety, the OPAC service's convenience, the support of the internet, and the hospitality of the library's staff.

Also from the result there is nine attribute that has high influenced to the frequency of customer visit to the library. Attribute semesters has the high percentage of importance that is 44,46%. Followed by faculty with 15,59%, department (13,69%), internet service (4,63%), security storage

bag (4,44%), comfort of reading rooms (4,39%), benefits of OPAC services (4,38%), library staff services (2,41%), and the last is book collection (1,62%).

This result implies that motivation is not usually become a attribute that can give an impact to the frequency of patron visit to the academic library. In fact, faculty, department and semesters also give an impact for frequency of customer visit to the library,

Acknowledgement

Acknowledgments to Universitas Negeri Malang (UM) which has funded this research through Islamic Development Bank (IsDB) - UM Research Grant.

References

[1] Abdulsalami LT, Okezie QI, Agbo AD. The role of the library in the promotion of knowledge societies in Nigeria. Pelagia Res Libr Adv Appl Sci Res 2013;4:58–70.
[2] Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.
[3] Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 2015;27:130–5. doi:10.11919/j.issn.1002-0829.215044.
[4] Montenegro M, Clasing P, Kelly N, Gonzalez C, Jara M, Alarcón R, et al. Library resources and students' learning outcomes: Do all the resources have the same impact on learning? J Acad Librariansh 2016;42:551–6.
[5] Abbott C, Webb SP. Performance measurement in library and information services. Aslib London; 1994.
[6] Fields A. Collaboration in Libraries and Learning Environments. Libr Rev 2013;62:626–7. doi:10.1108/LR-06-2013-0070.
[7] Li LH, Wu F, Su B. Impacts of Library Space on Learning Satisfaction – An Empirical Study of University Library Design in Guangzhou, China. J Acad Librariansh 2018;44:724–37. doi:10.1016/j.acalib.2018.10.003.
[8] Small R. Designing motivation into library and information skills instruction. Sch Libr Media Q 1998;1:1–15.
[9] Maslow A. Classics in the history of psychology. Choice Rev Online 2013;35:35SUP-462-35SUP – 462. doi:10.5860/choice.35sup-462.
[10] D'Elia G. The development and testing of a conceptual model of public library user behaviour. Libr Q 1980;50:410–30.
[11] Cristobal AS. Expectations on library services, library quality (LibQual) dimension and library customer satisfaction: Relationship to customer loyalty. Libr Philos Pract 2018;1:1–23.
[12] Lee S. Vroom's expectancy theory and the public library customer motivation model. Libr Rev 2007;56:788–96. doi:10.1108/00242530710831239.
[13] Banwet DK, Datta B. Effect of service quality on post-visit intentions over time: The case of a library. Total Qual Manag 2002;13:537–46. doi:10.1080/09544120220149322.
[14] Ahmed SMZ, Shoeb MZH. Measuring service quality of a public university library in Bangladesh using SERVQUAL. Perform Meas Metrics 2009;10:17–32. doi:10.1108/14678040910949666.
[15] Song Y, Lu Y. Decision tree methods : applications for classification and prediction 2015;27:130–5.
[16] Alsabti K, Ranka S, Singh V. CLOUDS : Tree Classifier for Large Datasets. Proc Fourth Knowl Discov Data Min Conf 1998:2–8.