CHAPTER 2

LITERATURE REVIEW

2.1 Student's Satisfaction

Student satisfaction is the extent to which students' hopes and desires for teaching and learning activities are expected with the reality received by students. From the statement above, it can be concluded that student satisfaction is a student's response to learning achievement that matches or even exceeds student expectations. Student satisfaction can be interpreted as a feeling of pleasure, satisfaction and relief for students both for the physical and non-physical services provided during the lecture proces, (Harmen et al., 2019).

In this study, the author refers to previous research that is relevant to this research. The following are some relevant research results that are used as study material for researchers, can be seen in table 2.1 :

Table 2.1 Previous Research Results

NO.	Author/Year	Problem	Method	Finding of Evaluation
1.	Sari et al., 2019	To determine students satisfaction by lecturers teaching method in STIKOM Tunas Bangsa	Naïve Bayes	Students satisfied with all aspects of lecturers teaching, such as communication, build a learning atmosphere, students evaluation, and delivery of learning materials
2.	Parlambang and Fauziah, 2020	Clustering students satisfaction by academic services	K-Means	From 248 students, 142 students dissatisfied, 23 students satisfied, and 83 students very satisfied to the academic services in Universitas Pandadaran

2.1 Previous Research Results(continue)

NO.	Author/Year	Problem	Method	Finding of Evaluation
3.	Wanda Rizki Fadillah, et al., 2020	Measuring students satisfaction level with the performance of computer laboratory assistants	C4.5 Algorithm	The most influenced aspect of the students satisfaction by performance of computer laboratory is the reliability aspect.

2.1.1 Factors That Affecting Student Satisfaction

Student's satisfaction is influenced by the difference between what students expect and the situation provided by the university in an attempt to fillfull the students expectations (Rahmawati, 2013). The main factors that affect service quality are expected service and perceived service. If the service is deemed to be in accordance to what students expected, then the service quality will be perceived well by students. Otherwise, if the service is perceived to be worse than expected, the service quality will be perceived as bad by students. Therefore, service quality will depend on the ability of service providers to consistently meet student's expectations (Sulastri, 2016).

Student's satisfaction is determined by several factors that lead to a sense of satisfaction and dissatisfaction that felt by students (Rahmawati, 2013). The factors that influence student satisfaction with learning performance are as follows:

1. Quality of service.

   Related to services, such as guarantees, responses, and problem solving. Students will feel satisfied if they get a good service and as students expected. Good and satisfactory service will produce a good response from students towards related educational institutions.

2. Quality of product.

   Related to the product, such as product quality, product form, and reliability. Students will be satisfied if the university using the high quality products to fulfill the student's expectation.

2.1.2 Dimensions of Student's Satisfaction

Student satisfaction is a personal perception that meets their needs and expectations of the quality of service or learning provided by the higher education institution where they study (Sukmanasa et al., 2017). Measuring the level of student satisfaction is difficult to do because services are not tangible. There are 5 aspects to calculate student satisfaction:

1. Reliability is a dimension that measures the reliability of higher education institutions in providing quality learning in accordance with the needs and expectations of students.
2. Responsiveness is a dimension that measures the response or accuracy of the actions of higher education institutions in responding to and overcoming student problems and complaints.
3. Assurance is a quality assurance dimension that assesses teaching staff in higher education institutions in instilling trust and confidence in students through interactions made by lecturers and students.
4. Empathy is a dimension that measures the services provided by higher education institutions in understanding the individual or personal needs of their students.
5. Tangibles is physical evidence that during learning the higher education institution provides adequate physical facilities and lecture equipment to its students.

2.2 Validity Test and Reliability Test

Revealing Data validity is the degree to which the collected data and the actual data in the data source are consistent. Substantial information will be acquired in the event that the information assortment instrument is additionally legitimate. The method commonly used to test the validity of the instrument is correlation analysis Sinulingga (2013). Correlation analysis is carried out using the Moment of Product correlation formula developed by Pearson, while the formula is 2.1

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}} \quad (2.1)$$

Where r_{xy} is the correlation coefficient X and Y, N is the number of subjects,∑XY is X and Y multiplication score, ∑X is the amount of X, $\sum X^2$ is sum of X squares, ∑Y is the amount of Y, and $\sum Y^2$ is the sum of Y squares.

Disclose the reliability of a measuring instrument with respect to the degree of consistency and stability of the data resulting from the data collection process using the instrument. Testing the consistency of the instrument using the Spearman-Brown formula is based on the split-half method, Sinulingga (2013). The correlation between the first and second hemispheres is calculated using the formula 2.2

$$r_{11} = \left(\frac{K}{k-1}\right)\left(1 - \frac{\sum \sigma_b^2}{\sigma_1^2}\right) \quad (2.2)$$

Where r_{11} is instrument reliability, K is many questions, $\sum \sigma_b^2$ is variansts items summary, and σ_1^2 is total variants.

2.3 Data Mining

Data mining is the process of extracting data into information that has not previously been conveyed, with the right techniques the data mining process will provide optimal results (Abdillah G., Putra F. A. and Renaldi F., 2016)". Using statistical and mathematical pattern recognition techniques, The process of filtering extremely large amounts of stored data to uncover meaningful new relationships, patterns, and trends is also referred to as data mining.

Data mining is not a brand-new field. One of the difficulties in defining it is that many aspects and methods of data mining are derived from established scientific fields. By utilizing expertise from a wide range of fields, data mining aims to enhance conventional methods in the following areas:

1. Huge amount of data
2. High data dimension
3. Heterogeneous data and different properties

Information mining is an interaction that utilizes factual procedures, math, man-made consciousness, and AI to remove and distinguish helpful data and related information from different huge data sets (Hendrian S, 2018). Broadly speaking, There are two primary categories of data mining:

a) Descriptive mining, or the process of finding the most important aspects of a database's data. Clustering, association, and sequential mining are examples of data mining methods included in descriptive mining.
b) Predictive, or the process of using a future variable to identify patterns in data. Classification is one of the methods used in predictive mining.

The most significant aspects of data mining are, in accordance with the definitions that have been provided, as follows:

1. Data mining is the automated process of using data that already exists.
2. The data that needs to be processed is very big.
3. Finding relationships or patterns that could provide useful clues is the goal of data mining.

2.3.1 Characteristics of Data Mining

Data mining also has characteristics as described in the study (Santoso H., Haryadi I. P. and Prayitno 2016), namely:

1. Data mining is related to the discovery of something hidden and certain data patterns that were not previously known.
2. Data mining usually uses very large data.
3. Data mining is useful for making critical decisions, especially in strategy.

2.3.2 Data Mining Tasks and Functions

Using tools, data mining looks for rules and patterns in data sets. By identifying data rules and features, the software is tasked with locating patterns in the data. It is anticipated that data mining tools will be able to recognize this pattern in the data with minimal user input. (Tampubolon K. et al, 2013).

2.4 Classification

Classification is the process of finding a set of models that describe and differentiate data classes. The purpose of classification is that the resulting model can be used to predict the class of data that does not have a class label. If given a set of data consisting of several features and classes, then classification is to find a model of that class as a function of the other features (Sukardi, Syukur A. and Supriyanto C., 2014).

There are also those who explain that, Finding the same properties on a set of objects in a database and classifying them into various classes using a defined classification model is the process of data classification. Finding a model from the training set that can classify attributes into the appropriate category or class is the goal of classification. The model is then used to classify attributes whose class is unknown. The classification technique is divided into several techniques, one of which is the Decision Tree (Lorena S., Zarman W. and Hamidah I., 2014).

Another study states that the classification technique is a systematic approach to building a classification model from a set of input data. For example, decision tree techniques, Bayesian (Naive Bayesian and Bayesian Belief Networks), Artificial Neural Networks (Backpropagation), concept-based techniques from mining association rules, and other techniques (K-Nearest Neighbor, genetic algorithm, technique with set approach rough and fuzzy). Classification is a technique of classifying data. The difference with the clustering method lies in the data, where in clustering there is no dependent variable, while in classification there is a dependent variable (Jamhur A. I., 2016).

2.5 Decision Tree

A decision tree is a decision tree where each branch shows a choice among a number of available alternative choices, and each leaf will show the chosen decision. Decision trees are commonly used to obtain information for the purpose of making a decision. There are those who define that a predictive model with a tree or hierarchical structure is called a decision tree.

The most widely used classification technique is the Decision Tree because it is not only quick to construct but also easy to comprehend in terms of the model's outcomes. The Decision Tree is a flowchart that resembles a tree structure. Each internal node displays a test on an attribute, each branch displays the results, and the leaf node displays the classes or class distribution. (Lorena S. et al, 2014). According to this definition, a decision tree is a flowchart or flowchart in which each node displays a tested attribute. each branch represents the results of the division that has been tested and each leaf node (internal node) represents a certain class group. The benefits of a decision tree can be seen as follows:

1. Reducing the complexity of the decision-making process into simpler ones so that decision makers can more effectively interpret problems' solutions.
2. Data exploration and the discovery of hidden relationships between a number of potential input variables and a target variable can also benefit from decision trees.
3. Choice trees join information investigation and demonstrating, so they are perfect as an initial phase in the displaying system in any event, when utilized as the last model of another strategy.

2.6 C4.5 Algorithm

A decision tree can be constructed from the data using the C4.5 algorithm. The ID3 algorithm, which is also a method for creating a decision tree, was developed into the C4.5 algorithm. The C4.5 algorithm makes a recursive visit to each decision node and selects the best branch until there are no more possible branches. (Rahmayuni I, 2014).

A common approach for classifying data mining is the decision tree or C4.5 algorithm. As previously explained, classification is a technique of finding a collection of patterns or functions that describe and separate data classes from one another to declare the object to be in a certain category by looking at the behavior and attributes of the defined group. It is able to categorize and demonstrate the connection between attributes, This approach is well-liked. A choice tree can be created using a variety of calculations, one of which is the C4.5 calculation. The C4.5 algorithm can handle both numerical and discrete data. The

(gain ratio) is utilized by the C4.5 algorithm. Prior to working out the procurement proportion, it is important to compute the data esteem in bits from an assortment of items, utilizing the idea of entropy (Supriyanti W., Kusrini and Amborowati A., 2016).

To create a decision tree, this algorithm starts by inserting training samples into the root node of the decision tree. Training samples are samples used to build a classifier model in this case a decision tree. Then an attribute is selected to partition this sample. For each value that this attribute has, a branch is created. After the branch is formed, a subset of the data set whose attributes have a value corresponding to that branch is inserted into the new node.

Additionally, it is mentioned that the C4.5 algorithm creates a decision tree from the database's training data in the form of cases or records (tuples). (Putra D. W. T., 2016).

1. The Concept of Entropy

   The estimated number of bits required to extract a class (positive or negative) from a number of random data in the sample space S is called entropy (S). Entropy is also known as the number of bits needed to represent a class. The amount of entropy used to extract a class is proportional to the entropy value. Entropy is used to measure S's originality. The following factors determine the amount of entropy in the sample space S:

$$Enthropy(S) = \sum - pi * \log_2 pi \quad (2.3)$$

   S denotes the case set, A denotes the feature, N denotes the number of partitions, and Pi denotes the ratio of Si to S.

2. Gain Concept

   Information is obtained from the output data or dependent variable S, which is grouped based on attribute A, denoted by gain (S,A). Gain (S,A) is the gain of information from attribute A relative to the output data S.

$$Gain(S,A) = Enthropy(S) - \sum_{i=I}^{n} \frac{|Si|}{|S|} * Enthopy(Si)(2) \quad (2.4)$$

   Where S is Case set, a is attribute, n is the number of attribute partitions a, and |Si| is umber of cases on the partition I, |S| is the umber of cases in S