Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 4 (2023) 895 - 903 ISSN Media Electronic: 2580-0760

Tree Algorithm Model on Size Classification Data Mining

Agis Abhi Rafdhi[1], Eddy Soeryanto Soegoto[2], Senny Luckyardi[3], Chepi Nur Albar[4]
[1]Department of Information System, Faculty of Engineering and Computer Science, Universitas Komputer Indonesia
[2,3]Department of Management, Faculty of Economic and Business, Universitas Komputer Indonesia
[4]Department of English, Faculty of Humanities, Universitas Komputer Indonesia
[1]agis@email.unikom.ac.id, [2]eddy.soeryantos@email.unikom.ac.id, [3]chepi.nuralbar@email.unikom.ac.id

Abstract

The goal of this research is to use a tree algorithm to categorize student clothing in order to acquire an accurate size. This research is qualitative approach through descriptive analysis, while the analysis employed C.45 Tree algorithm classification. Manual calculations utilizing the tree algorithm formula revealed that the majority of students require XL-sized clothing. On the X5 (Shoulder length) characteristic, the maximum entropy and information gain values were obtained at 0.212642462. According to the forecast, the shoulder length attribute is the first calculation in developing a decision tree scheme since it has the largest entropy and information gain value. Lastly, the findings of this study analysis can be used as a mapping prediction to make decisions on the size of the student group's clothing.

Keywords: *classification; data mining; decision tree*

1. Introduction

Data mining is a technique for generating patterns from a set of processed data in order to perform further analysis and produce reliable prediction results [1]. Data mining techniques can be used for a variety of purposes, including getting decision predictions and data distribution patterns, as well as sorting data into specific groups [2]. Classification is one of the data mining approach strategies used in data processing used to arrange datasets that have been separated into specified groupings [3]. This approach, for example, can categorize car types as Sedan, MPV, and SUV. The same rules can be used to categorize clothing sizes for undergraduate students. The datasets for this investigation were created based on the clothing sizes of 32 undergraduate students. To generate prediction patterns from the analyzed data, the classification stage employs a decision tree method.

Several studies use data mining classification techniques, one of which is in health sector carried out by Llias Tougui, Song, and Elhoseny [4]–[6] in forming decision patterns. The Naïve Bayes algorithm was utilized to classify the dataset and build data patterns in order to make decisions in health instances during the investigation. In the field of education, the algorithm can be used to determine the relationship and impact between a student's academic achievements and their social abilities. This study assesses any qualities that may influence student's potential to excel, both individually and as a group [7]. In another research, the Naïve Bayes classification technique was utilized to identify the typical pattern of consumer activity over a specific time period in order to assess the company's inventory control. In the research, entropy and information gain calculations are employed to detect general customer behavior in order to generate output for the company's policy-makers [8], [9], the research used the Naïve Bayes classification algorithm to quickly and simply determine the dataset pattern. However, in this study, data processing is carried out using a decision tree algorithm with the goal of conveniently visualizing the results of data processing so that decision making is relatively correct. The goal of this project is to use data mining classification algorithms to assess a dataset of students' clothing sizes. A decision tree technique with entropy and information gain calculations was employed to process the dataset.

2. Research Methods

This study used descriptive analysis in conjunction with a qualitative method. Furthermore, Classification techniques were applied in the data processing, while the Decision tree algorithm was used in the algorithm approach. The algorithm is typically used to solve a problem by representing the criteria as interconnected

Accepted: 07-10-2022 | Received in revised: 01-06-2023 | Published: 12-08-2023

nodes in the form of a tree [10], [11]. The tree, also known as the Decision tree, is a predictive decision-making approach model that is represented in a tree-shaped structural hierarchy [12], [13]. The phases of data processing using a decision tree begin with defining the selected qualities and go to the decision. The gainration calculation was utilized in this study to determine the selected qualities. Formula 1 is the gainration computation formula.

$$Gainratio(S,A) = \frac{Gain(S,A)}{SplitInformation(S,A)} \quad (1)$$

S is the Sample and A is the Attribute.

3. Results and Discussions

3.1 Data Collection

The dataset used in this study used 32 undergraduate students data as a test sample. Table 1 shows the dataset used in this study.

Table 1. Dataset of Student Clothing Measurement

Respondent	X1	X2	X3	X4	X5	TARGET
1	VERY WIDE	VERY WIDE	SHORT	SHORT	MEDIUM	XL
2	VERY WIDE	WIDE	LONG	SHORT	MEDIUM	L
3	WIDE	WIDE	LONG	TALL	WIDE	XL
4	WIDE	WIDE	VERY LONG	TALL	WIDE	XL
5	WIDE	WIDE	LONG	TALL	WIDE	XL
6	WIDE	WIDE	LONG	TALL	WIDE	XL
7	WIDE	WIDE	SHORT	TALL	WIDE	XL
8	WIDE	WIDE	VERY LONG	VERY TALL	WIDE	XL
9	WIDE	WIDE	SHORT	TALL	WIDE	XL
10	VERY WIDE	WIDE	SHORT	TALL	MEDIUM	XL
11	WIDE	WIDE	SHORT	TALL	MEDIUM	L
12	WIDE	WIDE	SHORT	TALL	MEDIUM	XL
13	WIDE	NARROW	SHORT	SHORT	MEDIUM	L
14	VERY WIDE	WIDE	SHORT	TALL	MEDIUM	XL
15	VERY WIDE	WIDE	SHORT	SHORT	MEDIUM	XL
16	VERY WIDE	WIDE	SHORT	VERY TALL	WIDE	XL
17	VERY WIDE	WIDE	SHORT	VERY TALL	MEDIUM	XL
18	VERY WIDE	VERY WIDE	VERY LONG	VERY TALL	WIDE	XL
19	WIDE	WIDE	SHORT	TALL	MEDIUM	L
20	SMALL	NARROW	SHORT	SHORT	MEDIUM	XL
21	VERY WIDE	VERY WIDE	LONG	TALL	MEDIUM	XL
22	VERY WIDE	WIDE	LONG	VERY TALL	MEDIUM	XL
23	WIDE	WIDE	SHORT	TALL	MEDIUM	L
24	VERY WIDE	WIDE	SHORT	TALL	MEDIUM	XL
25	WIDE	NARROW	SHORT	LOW	MEDIUM	L
26	WIDE	WIDE	SHORT	LOW	MEDIUM	L
27	WIDE	WIDE	SHORT	VERY TALL	MEDIUM	XL
28	WIDE	WIDE	LONG	TALL	MEDIUM	L
29	NARROW	WIDE	SHORT	TALL	MEDIUM	XL
30	WIDE	WIDE	SHORT	TALL	MEDIUM	L
31	WIDE	WIDE	SHORT	TALL	MEDIUM	L
32	WIDE	WIDE	VERY LONG	TALL	WIDE	XL

DOI: https://doi.org/10.29207/resti.v7i4.4572
Creative Commons Attribution 4.0 International License (CC BY 4.0)

In Table 1, there are 5 main attributes from X1 to X5. X1 means Bust, X2 is Stomach Circumference, X3 is Arm Length, X4 is Body Length, and X5 means Shoulder Length. On the other hand, the Target Attribute is a label attribute that will be a decision prediction.

3.2 Data Analysis

The first step in analyzing the obtained dataset is to determine the value of entropy and information gain. Entropy is an information theory measure used to determine the characteristics of an existing dataset [14], [15]. The entropy value is a critical metric in determining the Information Gain (IG) value for each characteristic. The largest IG value of the current attributes is used to pick characteristics as nodes, both as roots and as branches [16]. The formula for computing the Entropy value is shown in Formula 2.

$$\mathrm{H(X)} = -\sum_{i=1}^{n} P(X = i) log_2\ P(X = i) \qquad (2)$$

H is the Set entropy, Xis the Case set, n is the Number of X Partition, and Pi is the Number of samples for *i* and proportion of undergraduate students on *X*.

To calculate the entropy value in this dataset, there are 2 classes that will form the entropy value, namely L and XL. From this formula, the value of L class is 0.52439747 and the value of XL class is 0.371640762, resulting in entropy value of H(X) is 0.896038233. The next stage is the calculation of the IG value. Information gain is a measurement of attribute values used to test each node in the tree [17]. Formula 3 is a formula to get the IG value for each tested attribute:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * entropy\ S_i \qquad (3)$$

S is the Case set, A is the Attributes, n is the Number of partitions on Attribute A, |Si| is the Number of cases on i-th partition, and |S| is the Number of cases on S.

The largest information gain is observed in the X5 attribute with a value of 0.212642462, hence the tree algorithm test in Figure 1 is performed on the X5 attribute with Medium and Wide instruments. The results of the tree algorithm test on the X5 Medium and X5 Wide attributes are shown below (see Figures 2 and 3).

In Figure 2, the X5 Medium characteristic was evaluated against other qualities to acquire the IG value so that further nodes for this decision tree technique could be obtained. According to the test findings, the X1 characteristic had the highest IG value, which is 0.519358614. Based on the value, the following node from the X5 Medium property, namely the X4 attribute, will be evaluated (see Figure 4-6). While testing the X5 Wide attribute in Figure 3, the results produced are homogenous in XL size on all other qualities. As a result, the X5 Wide attribute indicates that the decision obtained is the XL size.

INFORMATION GAIN							
Attribut (X1)	L	XL	L	XL	Entropy(X1;c)		IG(S,X1)
VERY WIDE	0	10	0,314493784	0,125003203	0,439496987	0,151077089	0,166211507
WIDE	10	10	0,487368115	0,487368115	0,97473623	0,578749637	
SMALL	0	2	0	0	0	0	
	10	22	32			0,729826726	
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	3	0	0	0	0	0,08642354
WIDE	8	18	0,523212221	0,367279419	0,89049164	0,723524458	
NARROW	2	1	0,389975	0,528320834	0,918295834	0,086090234	
	10	22	32			0,809614692	
Attribute (X3)							IG(S,X3)
VERY LONG	0	4	0	0	0	0	0,078075935
LONG	2	5	0,516387121	0,346733448	0,863120569	0,188807624	
SHORT	8	13	0,530406637	0,428305246	0,958711883	0,629154673	
	10	22	32			0,817962298	
Attribute (X4)							IG(S,X4)
SHORT	4	3	0	0	0	0	0,212232976
TALL	6	13	0,525146846	0,374596913	0,899743759	0,683805257	
VERY TALL	0	6	0	0	0	0	
	10	22	32			0,683805257	
Attribute (X5)							IG(S,X5)
MEDIUM	10	12	0,517047056	0,476983155	0,994030211	0,68339577	0,212642462
WIDE	0	10	0	0	0	0	
	10	22	32			0,68339577	

Figure 1. Tree Algorithm Examination on All Attributes

DOI: https://doi.org/10.29207/resti.v7i4.4572
Creative Commons Attribution 4.0 International License (CC BY 4.0)

INFORMATION GAIN (X5 = "MEDIUM")							
Attribut (X1)	L	XL	L	XL	Entropy(X1;c)		IG(S,X1)
VERY WIDE	0	8	0,352213889	0,151044446	0,503258335	0,141541407	0,519358614
WIDE	10	2	0,23686905	0,447169385	0,684038436	0,235138212	
SMALL	0	2	0	0	0	0	
	10	12	22			0,376679619	
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	2	0	0	0	0	0,298304848
WIDE	8	9	0,511747219	0,485755327	0,997502546	0,529923228	
NARROW	2	1	0,389975	0,333333333	0,723308334	0,067810156	
	10	12	22			0,597733384	
Attribute (X3)							IG(S,X3)
VERY LONG	0	0	0	0	0	0	0,213557949
LONG	2	2	0,5	0,5	1	0,125	
SHORT	8	10	0,519966667	0,471109393	0,99107606	0,557480284	
	10	12	22			0,682480284	
Attribute (X4)							IG(S,X4)
SHORT	4	3	0,5	0,530639062	1,030639062	0,225452295	0,398869581
TALL	6	6	0,311278124	0,311278124	0,622556249	0,155639062	
VERY TALL	0	3	0	0,530639062	0,530639062	0,116077295	
	10	12	22			0,497168652	

Figure 2. Information Gain of Attribute X5 Medium

INFORMATION GAIN (X5 = "WIDE")							
Attribut (X1)	L	XL	L	XL	Entropy(X1;c)		IG(S,X1)
VERY WIDE	0	2		0	0	0	0,896038233
WIDE	0	8	0	0	0	0	
SMALL	0	0	0	0	0	0	
	0	10	10			0	
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0	0	0	0,896038233
WIDE	0	9	0	0	0	0	
NARROW	0	0	0	0	0	0	
	0	10	10			0	
Attribute (X3)							IG(S,X3)
VERY LONG	0	4	0	0	0	0	0,896038233
LONG	0	3	0	0	0	0	
SHORT	0	3	0	0	0	0	
	0	10	10			0	
Attribute (X4)							IG(S,X4)
SHORT	0	0	0	0	0	0	0,896038233
TALL	0	7	0	0	0	0	
VERY TALL	0	3	0	0	0	0	
	0	10	10			0	

Figure 3. Information Gain of Attribute X5 Wide

INFORMATION GAIN (X1 = "WIDE")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	0	0	0	0	0	0,035246117
WIDE	8	10	0,511747219	0,450314557	0,962061776	0,860792115	
NARROW	2	0	0	0	0	0	
	10	10	20			0,860792115	
Attribute (X3)							IG(S,X3)
VERY LONG	0	3	0	0	0	0	0,145266522
LONG	2	3	0,5	0,311278124	0,811278124	0,170795395	
SHORT	8	4	0,389975	0,528320834	0,918295834	0,579976316	
	10	10	20			0,750771711	
Attribute (X4)							IG(S,X4)
SHORT	4	0	0	0	0	0	0,170080659
TALL	6	8	0,523882466	0,46134567	0,985228136	0,725957574	
VERY TALL	0	2	0	0	0	0	
	10	10	20			0,725957574	

Figure 4. Information Gain of Attribute X1 Wide

In Figure 4, the X1 Wide attribute was evaluated on the X2, X3, and X4 attributes, with the X4 attribute yielding the highest IG of 0.170080659. As a result, the X4 attribute is a continuation node of the X1 Wide attribute (see Figures 7-9). While the X1 Very Wide attribute was evaluated on the X2, X3, and X4 attributes, the greatest value was recorded at the X3 attribute with an IG value of 0.789681414. However, the attribute value

Creative Commons Attribution 4.0 International License (CC BY 4.0)

was considered to be homogeneous at the XL size, hence the decision of attribute X1 Very Wide was XL size. Furthermore, the maximum value on the IG of X1 Small attribute is on the X3 attribute of 0.896038233, which is the same as the entropy value, implying that there is no option for the X1 Small attribute.

INFORMATION GAIN (X1 = "VERY WIDE")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	3	0	0	0	0	0,77344591
WIDE	0	7	0	0,168564443	0,168564443	0,122592322	
NARROW	0	0	0	0	0	0	
	0	10	10			0,122592322	
Attribute (X3)							IG(S,X3)
VERY LONG	0	1	0	0	0	0	0,789681414
LONG	0	2	0	0,389975	0,389975	0,106356818	
SHORT	0	7	0	0	0	0	
	0	10	10			0,106356818	
Attribute (X4)							IG(S,X4)
SHORT	0	2	0	0,389975	0,389975	0,106356818	0,789681414
TALL	0	4	0	0	0	0	
VERY TALL	0	4	0	0	0	0	
	0	10	10			0,106356818	

Figure 5. Information Gain of Attribute X1 Very Wide

INFORMATION GAIN (X1 = "SMALL")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	0	0	0	0	0	0,680908609
WIDE	0	1	0	0,240438991	0,240438991	0,215129623	
NARROW	0	0	0	0	0	0	
	0	1	1			0,215129623	
Attribute (X3)							IG(S,X3)
VERY LONG	0	0	0	0	0	0	0,896038233
LONG	0	0	0	0	0	0	
SHORT	0	0	0	0	0	0	
	0	0	0			0	
Attribute (X4)							IG(S,X4)
SHORT	0	1	0	0,240438991	0	0	0,666276631
TALL	0	1	0	0,240438991	0,240438991	0,177165572	
VERY TALL	0	1	0	0,240438991	0,240438991	0,052596029	
	0	3	3			0,229761601	

Figure 6. Information Gain of Attribute X1 Small

INFORMATION GAIN (X4 = "SHORT")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0	0	0	0,306898649
WIDE	2	1	0,389975	0	0,389975	0,129991667	
NARROW	2	1	0,389975	0,528320834	0,918295834	0,459147917	
	4	3	7			0,589139584	
Attribute (X3)							IG(S,X3)
VERY LONG	0	0	0	0	0	0	0,748645114
LONG	1	0	0	0	0	0	
SHORT	3	3	0,442179356	0,442179356	0,884358713	0,147393119	
	4	3	7			0,147393119	

Figure 7. Information Gain of Attribute X4 Short

In Figure 7, the X4 Short attribute was combined with the X2 and X3 attributes to find the next node; the X3 attribute had the maximum information gain with a value of 0.734213133. With these findings, the X4 Short attribute node is being tested on the X3 attribute. Furthermore, in Figure 8, the X4 Tall attribute test was performed, and the largest information gain on the X2 attribute was reached with a value of 0.470565953, indicating that the next node for the X4 Tall attribute is tested for the X2 attribute. Finally, a test of the X4 Very Tall attribute in Figure 9 reveals that the IG is the same for both traits and is homogeneous at the XL size. It can be concluded that the X4 Very Tall decision tree is XL size. The next test is the node of the X4 short attribute is shown in Figure 10 to 12.

DOI: https://doi.org/10.29207/resti.v7i4.4572
Creative Commons Attribution 4.0 International License (CC BY 4.0)

INFORMATION GAIN (X4 = "TALL")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0,223575132	0,223575132	0,020960169	0,470565953
WIDE	6	12	0,525146846	0,418714745	0,943861591	0,40451211	
NARROW	0	0	0	0	0	0	
	6	13	19			0,425472279	
Attribute (X3)							IG(S,X3)
VERY LONG	0	2	0	0,341887107	0,341887107	0	0,094025454
LONG	1	4	0,223575132	0,473247898	0,69682303	0,146699585	
SHORT	5	7	0,506841952	0,530737271	1,037579223	0,655313193	
	6	13	19			0,802012779	

Figure 8. Information Gain of Attribute X4 Tall

INFORMATION GAIN (X4 = "VERY TALL")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0	0	0	0,896038233
WIDE	0	5	0	0	0	0	
NARROW	0	0	0	0	0	0	
	0	6	6			0	
Attribute (X3)							IG(S,X3)
VERY LONG	0	2	0	0	0	0	0,896038233
LONG	0	1	0	0	0	0	
SHORT	0	3	0	0	0	0	
	0	6	6			0	

Figure 9. Information Gain of Attribute X4 Very Tall

INFORMATION GAIN (X3 = "VERY LONG")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0	0	0	0,896038233
WIDE	0	3	0	0	0	0	
NARROW	0	0	0	0	0	0	
	0	4	4			0	

Figure 10. Information Gain of Attribute X3 Very Long

In Figure 10, testing the X3 Very Long attribute is carried out on the X2 attribute because the remaining attributes that have not been tested are only the X2 attribute. The results of the X3 Very Long IG test resulted in all homogeneous values in the XL parameter and the information gain value was the same as the entropy. Therefore, the decision tree of the X3 Very Long attribute is the size of XL. Furthermore, based on Figure 11 test of the X3 Long attribute, it is found that the X2 attribute is still diverse, so it is necessary to carry out further testing on the X2 attribute. Figure 12 shows the same case as X2 attribute, so further testing needs to be done (see Figure 13-15).

INFORMATION GAIN (X3 = "LONG")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0	0	0	0,108927518
WIDE	2	4	0,528320834	0,389975	0,918295834	0,787110715	
NARROW	0	0	0	0	0	0	
	2	5	7			0,787110715	

Figure 11. Information Gain of Attribute X3 Long

INFORMATION GAIN (X3 = "SHORT")							
Attribute (X2)	L	XL	L	XL	Entropy(X2;c)		IG(S,X2)
VERY WIDE	0	1	0	0	0	0	0,193537696
WIDE	6	11	0,530294238	0,406373144	0,936667382	0,702500536	
NARROW	2	1	0,363230923	0,240438991	0,603669913	0	
	8	13	21			0,702500536	

Figure 12. Information Gain of Attribute X3 Short

In Figure 12, the X2 Very Fat attribute test was carried out on the X3 attribute and obtained homogeneous results in the XL size. While in Figure 13, all homogeneous attribute of X2 Fat tested on the XL size. Likewise, the homogeneous X2 Flat attribute was tested at size L. It can be concluded that for testing the X2 attribute, the decision tree obtained for the X2 Very fat attribute is XL, the X2 Fat and X2 Flat attributes are L size. Figure 15 is the result of the decision tree, based from all the tests that have been carried out on all attributes and obtained the following decision tree.

Creative Commons Attribution 4.0 International License (CC BY 4.0)

INFORMATION GAIN (X2="VERY WIDE")							
Attribute (X3)							IG(S,X3)
VERY LONG	0	4	0	0	0	0	0,896038233
LONG	0	1	0	0	0	0	
SHORT	0	1	0	0	0	0	
	0	6	6			0	

Figure 13. Information Gain of Attribute X2 Very Fat

INFORMATION GAIN (X2="WIDE")							
Attribute (X3)							IG(S,X3)
VERY LONG	4	0	0	0	0	0	0,896038233
LONG	2	0	0	0	0	0	
SHORT	6	0	0	0	0	0	
	12	0	12			0	

Figure 14. Information Gain of Attribute X2 Fat

INFORMATION GAIN X2="NARROW"							
Attribute (X3)							IG(S,X3)
VERY LONG	4	0	0	0	0	0	0,896038233
LONG	6	0	0	0	0	0	
SHORT	2	0	0	0	0	0	
	12	0	12			0	

Figure 15. Information Gain of Attribute X2 Flat

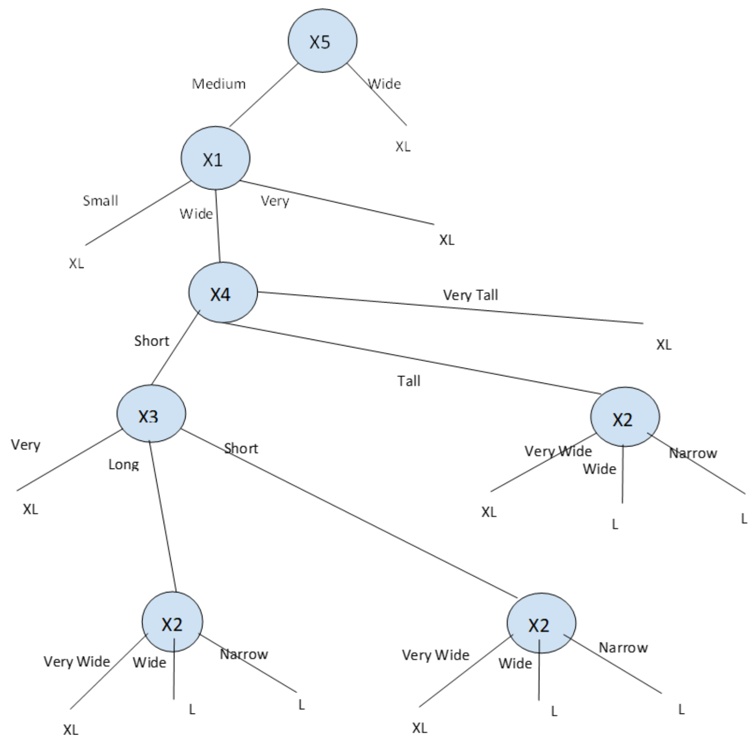

Figure 16. Results of the Decision Tree

From the results of the decision tree obtained in Figure 16, it can be described. The results of the decision tree algorithm were processed manually using theoretical formulas.

1. IF X5="MEDIUM" AND X1="SMALL" THEN Ukuran ="XL"
2. IF X5="WIDE" THEN Ukuran = "XL"
3. IF X5="MEDIUM" AND X1="WIDE" AND X4="VERY TALL" THEN Ukuran ="XL"
4. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="VERY LONG" THEN Ukuran ="XL"
5. IF X5="MEDIUM" AND X1="WIDE" AND X4="TALL" AND X2="VERY FAT" THEN Ukuran ="XL"
6. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="LONG" AND X2="VERY FAT" THEN Ukuran ="XL"

DOI: https://doi.org/10.29207/resti.v7i4.4572
Creative Commons Attribution 4.0 International License (CC BY 4.0)

7. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="LONG" AND X2="FAT" THEN Ukuran ="L"
8. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="LONG" AND X2="FLAT" THEN Ukuran ="L"
9. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="SHORT" AND X2="VERY FAT" THEN Ukuran ="XL"
10. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="SHORT" AND X2="FAT" THEN Ukuran ="L"
11. IF X5="MEDIUM" AND X1="WIDE" AND X4="SHORT" AND X3="SHORT" AND X2="FLAT" THEN Ukuran ="L"
12. IF X5="MEDIUM" AND X1="VERY WIDE" THEN Ukuran="XL"

In future research, it can be done using an application that can generate and display more accurate results. This manual test aims to explain the flow of calculations in the prediction of decision tree decision making [18]–[20]. Data mining techniques, such as the tree algorithm formula, can be used to evaluate business activities, such as mapping the most popular products and examining the relationships between those products and other products. Another example is that data mining can be used to evaluate the market in order to discover the link between products purchased by customers in the transaction database. The output of data mining assists the owner in developing the marketing strategy and procurement system [21]. Furthermore, data mining can be coupled with other technologies such as 3D Virtual Reality to provide accurate data and lifelike visuals. For example, it was discovered in the examination of clothing comfort from the numerical garment viewpoint that different regions of the human body affect the amount of comfort differently. The study aided in the development of a pattern for measuring the ease of quantitative calculations for each data sample [22].

The tree algorithm model works by recursively partitioning the feature space into subsets based on the values of the features. This is done by constructing a tree-like structure where each node represents a subset of the feature space and each branch represents a decision based on a feature value. At each node, the algorithm selects the feature that provides the most significant information gain and splits the data based on its values. The process continues until a stopping criterion is met, such as reaching a maximum depth or minimum number of samples in a leaf node.

To apply the tree algorithm model for size classification, the features used can be physical characteristics of the object, such as length, width, and height. These features can be used to train the model using a dataset of objects with known sizes. Once the model is trained, it can be used to predict the size of new objects based on their features.

Overall, the tree algorithm model is a powerful and flexible tool for size classification, and it has been successfully applied in many fields, such as agriculture, industry, and medical imaging. However, like all machine learning models, it requires careful data preprocessing and validation to ensure accurate and reliable results

4. Conclusion

Based on the results of the Decision tree test from the undergraduate student data, it was observed that the majority of predictions resulted on the XL size with 8 out of 12 decisions. he entropy value of 0.896038233. On the Information Gain, testing begins with the X5 attribute after attaining the highest Information Gain when testing all attributes, which is 0.212642462. Meanwhile, the next node is tested on the X1 attribute with a value of 0.519358614 for Information Gain. After testing on the X1 Wide instrument, the third node is tested on the X4 attribute with an Information gain value of 0.170080659. Furthermore, after receiving an Information Gain of 0.734213133 on the X4 Short attribute, the X3 attribute became the second last test. Finally, the X2 attribute is the final test that leads to the decision.

References

[1] V. Plotnikova, M. Dumas, and F. Milani, “Adaptations of data mining methodologies: A systematic literature review,” *PeerJ Comput. Sci.*, vol. 6, pp. 1–43, 2020, doi: 10.7717/PEERJ-CS.267.
[2] S. Asha Kiranmai and A. Jaya Laxmi, “Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy,” *Prot. Control Mod. Power Syst.*, vol. 3, no. 1, 2018, doi: 10.1186/s41601-018-0103-3.
[3] S. Sriramoju, G. Ramesh, and B. Srinivas, “An Overview of Classification Rule and Association Rule Mining,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3, no. April, pp. 1692–1697, 2018, [Online]. Available: https://www.researchgate.net/profile/Gadde-Ramesh/publication/323550246_An_Overview_of_Classification_Rule_and_Association_Rule_Mining/links/5ad98d64a6fdcc293586aa12/An-Overview-of-Classification-Rule-and-Association-Rule-Mining.pdf.
[4] I. Tougui, A. Jilbab, and J. El Mhamdi, “Heart disease classification using data mining tools and machine learning techniques,” *Health Technol. (Berl).*, vol. 10, no. 5, pp. 1137–1144, 2020, doi: 10.1007/s12553-020-00438-1.
[5] C. W. Song, H. Jung, and K. Chung, “Development of a medical big-data mining process using topic modeling,” *Cluster Comput.*, vol. 22, pp. 1949–1958, 2019, doi: 10.1007/s10586-017-0942-0.
[6] M. Elhoseny, K. Shankar, and J. Uthayakumar, “Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, 2019, doi: 10.1038/s41598-019-46074-2.
[7] M. M. Arcinas, G. S. Sajja, S. Asif, S. Gour, E. Okoronkwo, and M. Naved, “Role of Data Mining in Education for

DOI: https://doi.org/10.29207/resti.v7i4.4572
Creative Commons Attribution 4.0 International License (CC BY 4.0)

Improving Students Performance for Social Change," *Turkish J. Physiother. Rehabil.*, vol. 32, no. 3, pp. 6519–6526, 2021.

[8] R. Rahim *et al*., "C4.5 classification data mining for inventory control," *Int. J. Eng. Technol.*, vol. 7, no. July 2019, pp. 68–72, 2018, doi: 10.14419/ijet.v7i2.3.12618.

[9] M. Sharma, S. Sharma, and G. Singh, "Performance analysis of statistical and supervised learning techniques in stock data mining," *Data*, vol. 3, no. 4, pp. 1–16, 2018, doi: 10.3390/data3040054.

[10] M. Z. Arif, R. Ahmed, U. H. Sadia, M. S. I. Tultul, and R. Chakma, "Decision Tree Method Using for Fetal State Classification from Cardiotography Data," *J. Adv. Eng. Comput.*, vol. 4, no. 1, p. 64, 2020, doi: 10.25073/jaec.202041.273.

[11] M. Li, H. Xu, and Y. Deng, "Evidential decision tree based on belief entropy," *Entropy*, vol. 21, no. 9, 2019, doi: 10.3390/e21090897.

[12] T. Ç. AKINCI and H. S. Noğay, "Application of Decision Tree Methods for Wind Speed Estimation," *Eur. J. Tech.*, vol. 9, no. 1, pp. 74–83, 2019, doi: 10.36222/ejt.558914.

[13] K. N. Dey, S. Saha, A. Ghosh, and S. Bandopadhyay, "Missing value imputation in DNA microarray gene expression data: a comparative study of an improved collaborative filtering method with decision tree based approach," *Int. J. Comput. Sci. Eng.*, vol. 18, no. 2, p. 130, 2019, doi: 10.1504/ijcse.2019.10019160.

[14] T. Behdadnia, Y. Yaslan, and I. Genc, "A new method of decision tree based transient stability assessment using hybrid simulation for real-time PMU measurements," *IET Gener. Transm. Distrib.*, vol. 15, no. 4, pp. 678–693, 2021, doi: 10.1049/gtd2.12051.

[15] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74–78, 2018, doi: 10.26438/ijcse/v6i10.7478.

[16] S. Mauluddin, I. Ikbal, and A. Nursikuwagus, "Complexity and performance comparison of genetic algorithm and ant colony for best solution timetable class," *J. Eng. Sci. Technol.*, vol. 15, no. 1, pp. 276–290, 2020.

[17] E. Odhiambo Omuya, G. Onyango Okeyo, and M. Waema Kimwele, "Feature Selection for Classification using Principal Component Analysis and Information Gain," *Expert Syst. Appl.*, vol. 174, no. February, p. 114765, 2021, doi: 10.1016/j.eswa.2021.114765.

[18] K. Mathan, P. M. Kumar, P. Panchatcharam, G. Manogaran, and R. Varadharajan, "A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease," *Des. Autom. Embed. Syst.*, vol. 22, no. 3, pp. 225–242, 2018, doi: 10.1007/s10617-018-9205-4.

[19] I. Budiarti, R. Andrian, and A. W N Falah, "Application of Web Communication Relationship Management in Small and Medium Enterprises," *Int. J. Res. Appl. Technol.*, vol. 1, no. 1, pp. 49–54, 2021, doi: 10.34010/injuratech.v1i1.5611.

[20] W. Novianti and E. Erdiana, "Information Technology to Support E-Advertisement," *Int. J. Res. Appl. Technol.*, vol. 1, no. 1, pp. 134–139, 2021, doi: 10.34010/injuratech.v1i1.5656.

[21] E. Yakut and M. Yüksel Avcilar, "Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store," *Can. Soc. Sci.*, vol. 10, no. 3, p. 75, 2014, doi: 10.3968/4541.

[22] K. Liu, J. Wang, and Y. Hong, "Wearing comfort analysis from aspect of numerical garment pressure using 3D virtual-reality and data mining technology," *Int. J. Cloth. Sci. Technol.*, vol. 29, no. 2, pp. 166–179, 2017, doi: 10.1108/IJCST-03-2016-0017.

DOI: https://doi.org/10.29207/resti.v7i4.4572
Creative Commons Attribution 4.0 International License (CC BY 4.0)