Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2022. Vol. 1. No. 2

APPLICATION OF DATA MINING IN DETERMINING SOCIAL ASSISTANCE RECIPIENTS WITH C4.5 ALGORITHM

Rika Nur Adiha[1]*, Sundari Retno Andani[2]*, Widodo Saputra[3]

[1] STIKOM Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia
[2,3] AMIK Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia
*rikaadiha9@gmail.com

Abstract

The Gunung Maligas District Office is a government agency tasked with running a government program, namely the Social Assistance Receipt program, to run the social assistance program, many residents complain that they do not receive assistance, while some residents who are considered capable actually get assistance, where each aid program is have different criteria in determining the recipient. Due to the large number of existing aid programs with different criteria in determining the acceptance of the aid program, of course, local government staff will have difficulty in conducting the selection process. So we need a system that is able to help local government staff to more easily determine the recipients of the social assistance. Based on the historical data of beneficiaries, recommendations for the classification of beneficiaries can be made that will assist government staff. Classification can be done using the C4.5 algorithm. In this study, it has parameters, namely, occupation, income, housing conditions and number of dependents. By applying the C4.5 data mining algorithm, it is hoped that it will make it easier and faster for government staff to determine the recipients of social assistance at the Gunung Maligas District Office.

Keywords: *Data Mining, Classification, C4.5 Algorithm, Social assistance*

1. Introduction

The social welfare of the community is always faced with various kinds of problems. Social problems such as poverty that will have an impact on people's lives. To overcome this problem, the government provides assistance [1],[2]. Social assistance is assistance in the form of money or goods that are not continuous and selective which aims to improve people's welfare [3],[4],[5]. To overcome this problem, the Gunung Maligas District Office always routinely implements the government program regarding the provision of social assistance to underprivileged residents [6],[7]. The assistance program is one of the cash social assistance programs (BST), where each aid program has a procedure, the Gunung Maligas Subdistrict Office has a procedure, namely by conducting a selection process for each citizen, each aid program has different criteria in determining the recipient. With so many existing aid programs with different criteria, of course, local government staff will find it difficult to carry out the selection process [8],[9]. Based on these problems, a method is needed to make it easier to determine beneficiaries so that the people who get assistance are recipients who really need or are right on target [10],[11]. One method that can be used in this research is the application of data mining algorithm C4.5 [12],[13]. In determining the classification of aid program recipients based on predetermined variables [14],[15]. By using the C4.5 algorithm, it is expected to produce a decision tree model that can predict families who are eligible to receive assistance programs and can assist local government staff in determining decision trees [16],[17]. Previous research by Hariati et all, on the application of the C4.5 Algorithm in determining the recipients of local government assistance programs in Kutai Kartanegara district obtained the results that the C4.5 algorithm can be applied to the application of determining the recipients of local government assistance programs by using a rule formed from the results. the highest accuracy on training data is 85% with a total of 105 data [18].

2. Research Method

The following will explain how to describe the scientific way to solve research problems. The method used in this study is the C4.5 Algorithm. The results of this study were carried out to produce a decision rule tree model for recipients of social assistance programs in accordance with the requirements. The research design carried out for determining the recipients of social assistance using the C4.5 algorithm begins by analyzing problems related to recipients of social assistance and determining the variables used. Furthermore, studying literature is based on references made to obtain information in research. Collecting data using a questionnaire by distributing it to people who receive social assistance, then Analysis is a process carried out for determining social assistance recipients by using the variables used in the study. Implementation using the Rapidminer version 5.3 application as a tool and the results are the results obtained from this research. It will be used as

new knowledge in determining social assistance recipients. The data collection carried out in this study uses library research by utilizing libraries and journals as references used in research and field research (Field Work Research) research is carried out directly in the field using questionnaires, literature studies, interviews and observations.

Work Activity Diagram The research carried out is described in the activity diagram in Figure 1 below:

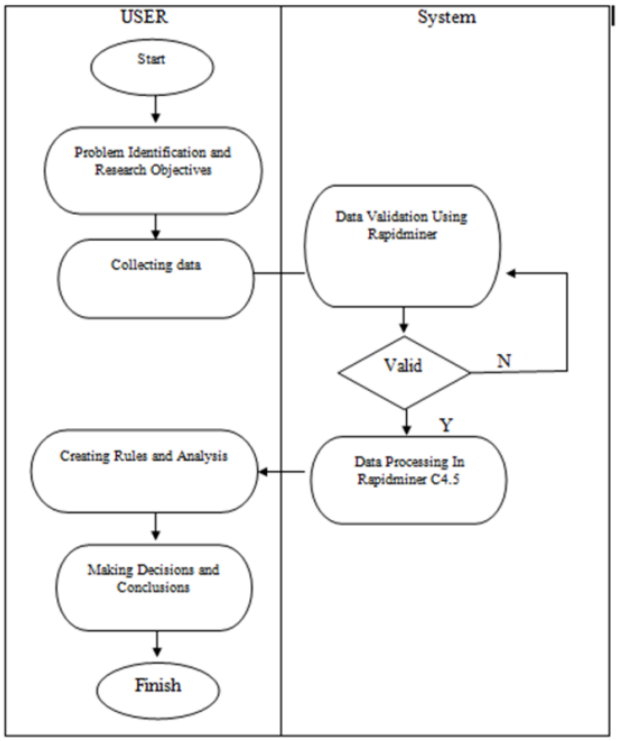

Figure 1. Research Activity Diagram

Figure 1 describes the author or user in identifying the problem and the purpose of the research carried out, collecting data in the form of a questionnaire given, on the Rapidminer application.

3. Results And Discussion

In this study, the authors obtained data from the questionnaire results from the Gunung Maligas sub-district office with a total of 50 data. To make it as a root, it was based on the highest gain value of the existing attributes. To calculate the gain, use a formula like the following [19],[20],[21] :

$$\boldsymbol{Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)} \quad (1)$$

Information:
A = Attribute
S = Set of cases
S1 = Number of samples

Before getting the gain value, the entropy value is sought first. To find the entropy value, use the following formula [22],[23],[24]:

$$\boldsymbol{E\ (s) = \sum_{i=1}^{n} - pi * log_2\ (Pi)} \quad (2)$$

Information :
S = Set of cases
S1 = Number of samples
Pi = Proportion\

3.1. Data Processing Using C4.5 . Algorithm

The following are the steps of data processing with the C4.5 Algorithm in order to obtain a model of decision tree rules for social assistance recipients in accordance with the data obtained from the questionnaire. For the calculation of the C4.5 algorithm, it can be described as follows [25]:
Step 1: Counting the number of cases, the number of cases for a feasible decision, and the number of cases for an invalid decision. Calculating the total Entropy:

$$Entropy\ [\text{Total}] = -\left(\frac{19}{50}\right) \times \log 2\left(\frac{19}{50}\right) - \left(\frac{31}{50}\right) \times \log 2\left(\frac{31}{50}\right) = 0{,}958042$$

Step 2: Then, calculate the entropy and gain for each data attribute. The following is the calculation of the Entropy and Gain values. Calculating Entropy and Job Gain:

Entropy [Work-Labor] =-(0/21)x log2⁡[(0/21)-(21/21)x log2⁡(21/21)]=0
Entropy [Job-PNS] = -(7/7)x log2⁡[(7/7)-(0/7)x log2⁡(0/7)]=0
Entropy [Job-Entrepreneur]= -(5/12)x log2⁡[(5/12)-(7/12)x log2⁡(7/12)]=0,979869
Entropy [Private-Employee Job]= -(7/10)x log2⁡[(7/10)-(3/10)x log2⁡(3/10)]=0,881291
Gain [Total, Employment] = 0,958042-((21/50)x0+(7/50)x0+(12/50)x0,979869+(10/50)x0,881291=0,546615)

Calculating Entropy and Gain Reliability:
Entropy [Excellent-Earning] = -(3/3)x log2⁡[(3/3)-(0/3)x log2⁡(0/3)]=0
Entropy [Good-Income] = -(4/4)x log2⁡[(4/4)-(0/4)x log2⁡(0/4)]=0
Entropy [Income-Sufficient] = -(11/15)x log2⁡[(11/15)-(4/15)x log2⁡(4/15)]=0,836641
Entropy [Income-Low] = -(0/0)x log2⁡[(0/0)-(19/19)x log2⁡(19/19)]=0
Entropy [Income-Very Less] = -(0/0)x log2⁡[(0/0)-(9/9)x log2⁡(9/9)]=0
Gain [Total, Earnings] = 0,958042-((3/50)x0+(4/50)x0+(15/50)x0,836641+(19/50)x0+(9/50)=0,707050)

Then do the same calculation for the next variable. After that, the gain is calculated for each attribute. The calculation results are shown in table 1 below.

Table 1. Calculation Results of Node 1

Node 1		Amount	Worthy	Not Worthy	Entrophy	Gain
Total		50	31	19	0,958042	
Work						0,546615
	Laborer	21	21	0	0,000000	
	PNS	7	0	7	0,000000	
	Entrepreneur	12	7	5	0,979869	
	Private employees	10	3	7	0,881291	
Income						**0,707050**
	Very good	3	0	3	0,000000	
	Good	4	0	4	0,000000	
	Enough	15	4	11	0,836641	
	Not enough	19	19	0	0,000000	
	Very less	9	9	0	0,000000	
Dependent Amount						0,13209
	A huge amount	3	2	1	0,000000	
	Lots	37	26	11	0,877962	
	A little	10	3	7	0,881291	
Home Condition						0,41613
	Permanent	15	2	13	0,566510	
	Semi Permanent	22	16	6	0,845351	
	Board	9	9	0	0,000000	
	Bamboo	4	4	0	0,000000	

From the results of the calculations in table 1, the highest value obtained is the Income attribute with a Gain value of 0.707050. So what is used as the root node is the Income attribute, which consists of five sub-attributes, namely Very good, Good, Enough, Not enough, and Very less. The decision tree obtained based on Node 1 calculations is as follows:

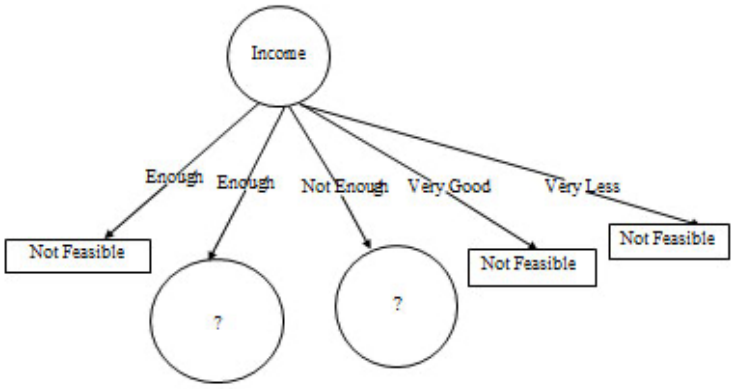

Figure 2. Decision Tree 1

From figure 2. from the calculation of Node 1, the Income attribute is used as the root node. As for the Enough and Not enough attribute classes based on the respective Entropy values, the results between the Worthy or Not Worthy decisions have not been obtained, so further calculations need to be carried out. The following is the result of the calculation of the Income = Enough attribute which can be seen in Table 2.

Tabel 1. Node Calculation Results 1.1

Node 1.1		**Amount**	**Worthy**	**Not Worthy**	**Entrophy**	**Gain**
Income = Enough		15	4	11	0,836641	
Work						0,241995
	Laborer	0	0	0	0	
	PNS	0	0	0	0	
	Entrepreneur	6	0	6	0,000000	
	Private employees	9	4	5	0,991076	
Dependent Amount						**0,327609**
	A huge amount	1	1	0	0,000000	
	Lots	8	3	5	0,954434	
	A little	6	0	6	0,000000	
Home Condition						0,303307
	Permanent	7	0	7	0,000000	
	Semi Permanent	8	4	4	1,000000	
	Board	0	0	0	0,000000	
	Bamboo	0	0	0	0,000000	

From Table 2. the calculation results that become branch nodes of Income-Enough are Dependent Amount with the highest Gain value of 0.327609. The attribute class values A huge amount, and A little, are zero, so no calculation is necessary. The value of the Enough attribute class has classified one decision . The decision tree obtained based on the calculation of Node 1.1 is as follows:

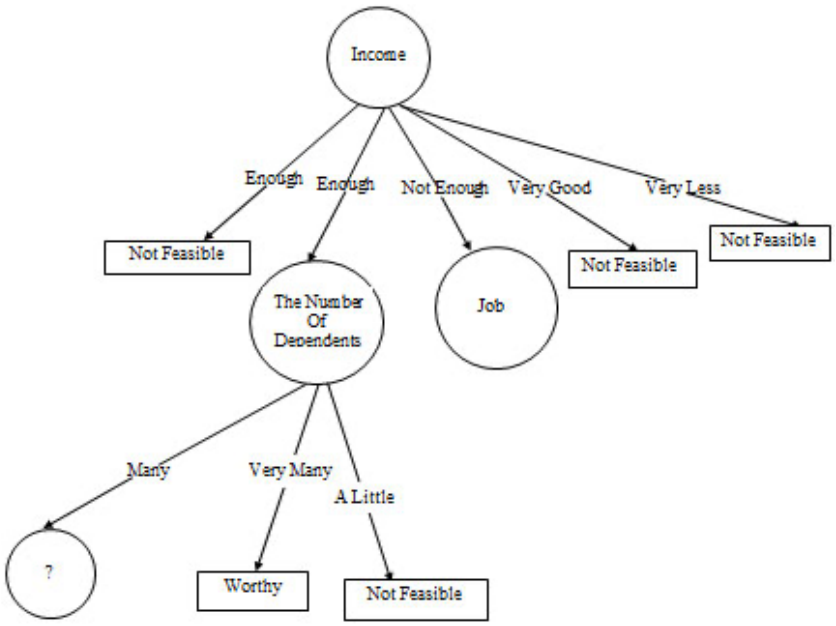

Figure 3. Decision Tree 2

In finding the results of the next calculation on the Income-Enough root node and the Dependent Amount-Lots branch node, it can be shown in the following table 3.

Table 3. Node Calculation Results

Node 1.2		**Amount**	**Satisfied**	**Not Satisfied**	**Entrophy**	**Gain**
Income = Enough & Dependent Amount = Lots		8	3	5	0,954434	**0,548795**
Work						
	Laborer	0	0	0	0,000000	
	PNS	0	0	0	0,000000	
	Entrepreneur	4	1	3	0,811278	
	Private employees	4	2	2	1,000000	
Home Condition						0,048795
	Permanent	4	1	3	0,811278	
	Semi Permanent	4	2	2	1,000000	
	Board	0	0	0	0,000000	
	Bamboo	0	0	0	0,000000	

From the calculation results in table 3, the attributes that become branch nodes of Income – Enough and Dependent Amount – Lots are Work.. Where the attributes of Work consist of Laborers, PNS, Private employees, and Entrepreneurs. Where Laborer has obtained a decision, civil servants have obtained a decision. As for the sub-attributes Private employees and Entrepreneurs have not received a decision, it will be recalculated. Then the decision tree can be described from the table above as follows:

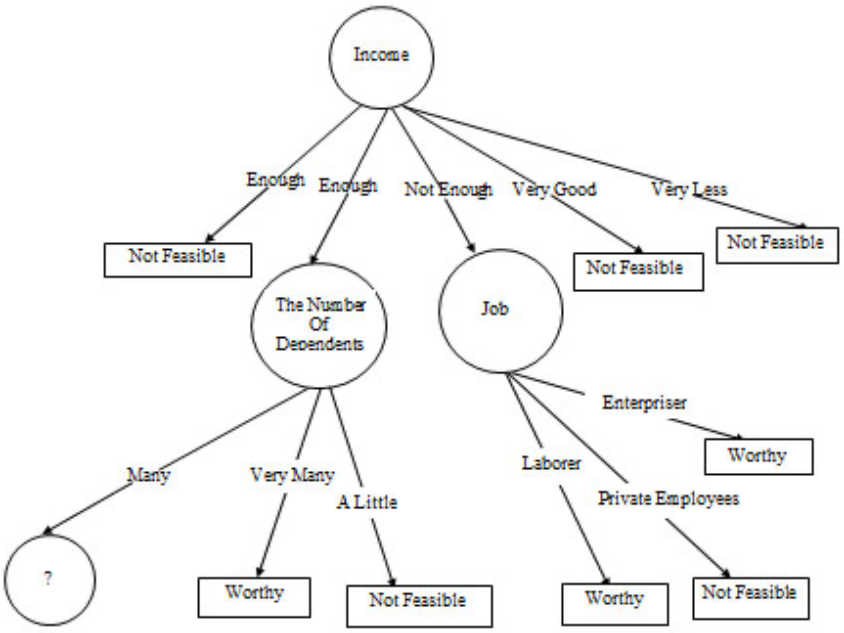

Figure 4. Decision Tree 3

Next, perform calculations on the branch nodes of the Income – Enough, Dependent Amount – Lots, Work – Private employees and Entrepreneur attributes as shown in table 4.

Table 4. Node Calculation Results 1.1.1

Node 1.1.1		**Amount**	**Satisfied**	**Not Satisfied**	**Entrophy**	**Gain**
Income = Enough & Dependent Amount = Lots & Work = Private employees & Entrepreneur		4	2	2	1,000000	0,000000
Home Condition						
	Permanent	2	1	1	1,000000	
	Semi Permanent	2	1	1	1,000000	
	Board	0	0	0	0,000000	
	Bamboo	0	0	0	0,000000	

From the calculation results in table 4, the attributes that become branch nodes of Income – Enough and Dependent Amount – Lots are Work.. Where the attributes of Work consist of Laborers, PNS, Private employees, and Entrepreneurs. Where Laborer has obtained a decision, civil servants have obtained a decision. As for the sub-attributes Private employees and Entrepreneurs have not received a decision, it will be recalculated. Then the decision tree can be described from the table above as follows:

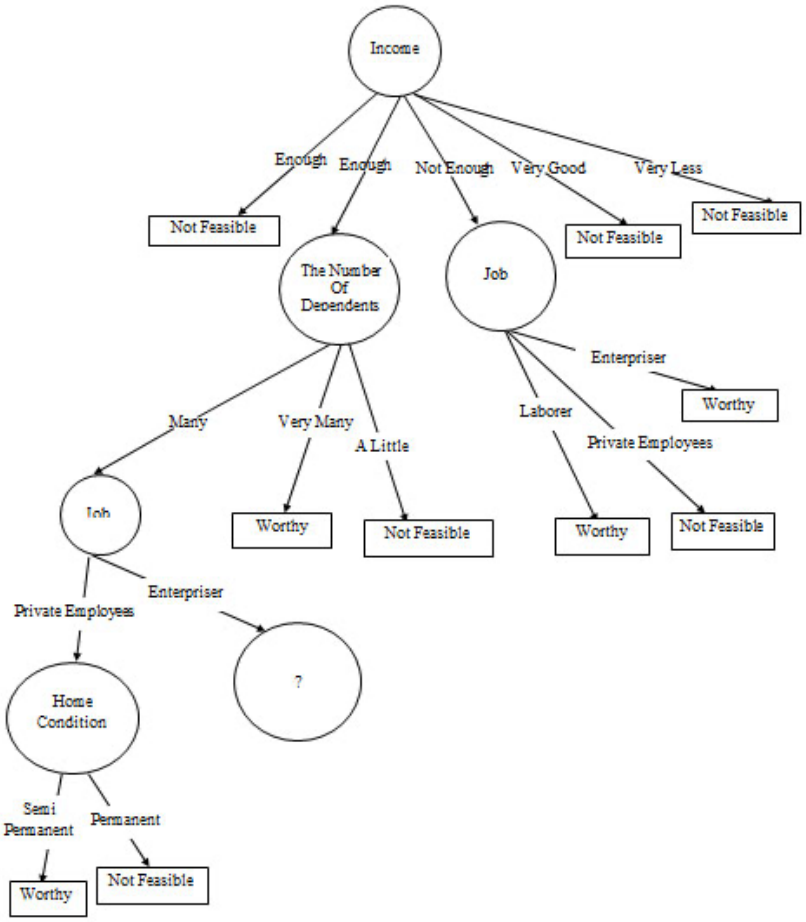

Figure 5. Decision Tree 4

Next, perform calculations on the branch nodes of the Income – Enough, Dependent Amount – Lots, Work – Private employees and Entrepreneur attributes as shown in table 5.

Table 5. Node Calculation Results 1.1.2

Node 1.1.2	Amount	Satisfied	Not Satisfied	Entrophy	Gain
Income = Enough & Dependent Amount = Lots & Work = Private employees & Entrepreneur	4	1	3	0,811278	0,000000
Home Condition					
Permanent	4	1	3	0,811278	
Semi Permanent	0	0	0	0,000000	
Board	0	0	0	0,000000	
Bamboo	0	0	0	0,000000	

From the above calculation results in table 5. The attributes that become branch nodes of Income – Enough, Dependent Amount – Lots, Work-Private employees and Entrepreneurs, are Home Condition – Permanent, and Semi Permanent. Attributes from Home Condition – Permanent, Semi Permanent, Board, and Bamboo. Where Semi Permanent, Board, Bamboo already have a decision. Meanwhile, the Permanent sub attribute has not yet received a decision, so it will be recalculated. Then the decision tree can be described from the table above as follows:

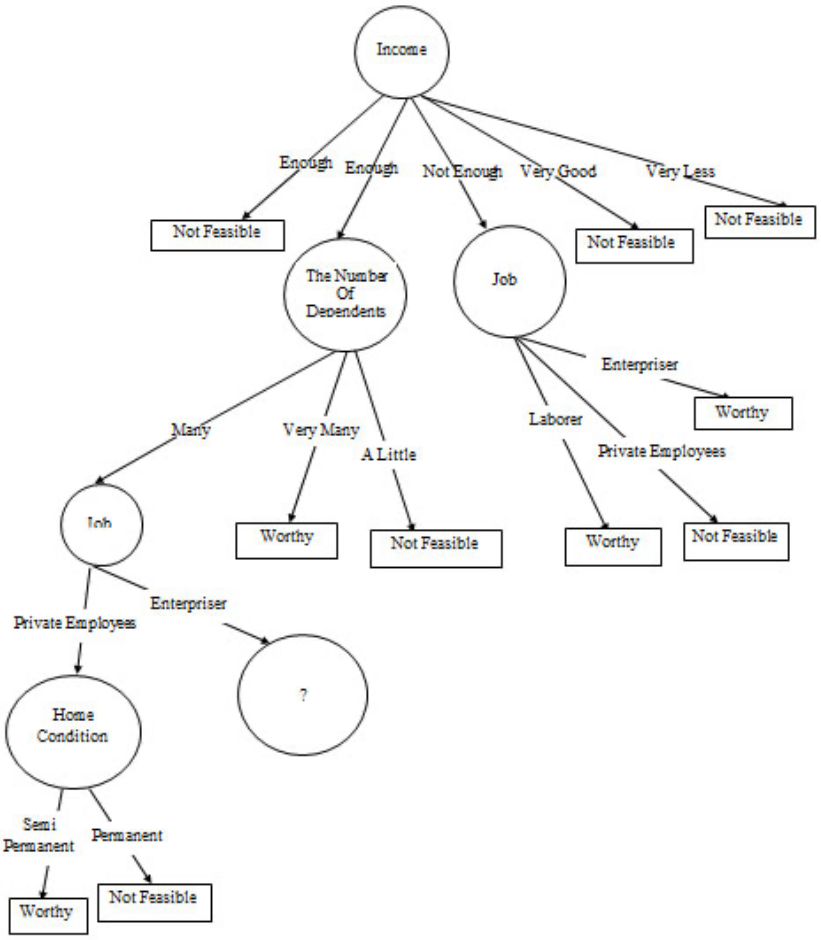

Figure 6. Decision Tree 5

Next, perform calculations on the branch nodes of the Income – Enough, Dependent Amount – Lots, Work – Private employees and Entrepreneur, Home Condition – Permanent attributes as shown in table 6.

Table 6. Node Calculation 1.1.3

Node 1.1.2	Amount	Satisfied	Not Satisfied	Entrophy	Gain
Income = Enough & Dependent Amount = Lots & Work = Private employees & Entrepreneur	4	1	3	0,811278	0,000000
Home Condition					
Permanent	4	1	3	0,811278	
Semi Permanent	0	0	0	0,000000	
Board	0	0	0	0,000000	
Bamboo	0	0	0	0,000000	

From the above calculation results in table 6, the attributes that become branch nodes of Income – Enough, Dependent Amount – Lots, Work-Private employees and Entrepreneurs, are Home Condition – Permanent, and Semi Permanent. Attributes of Home Condition – Permanent, Semi Permanent, Board, and Bamboo .. Where the sub attributes Permanent, Semi Permanent, Board, Bamboo already have a decision so the calculation is complete. Then the decision tree can be described from the table above as follows:

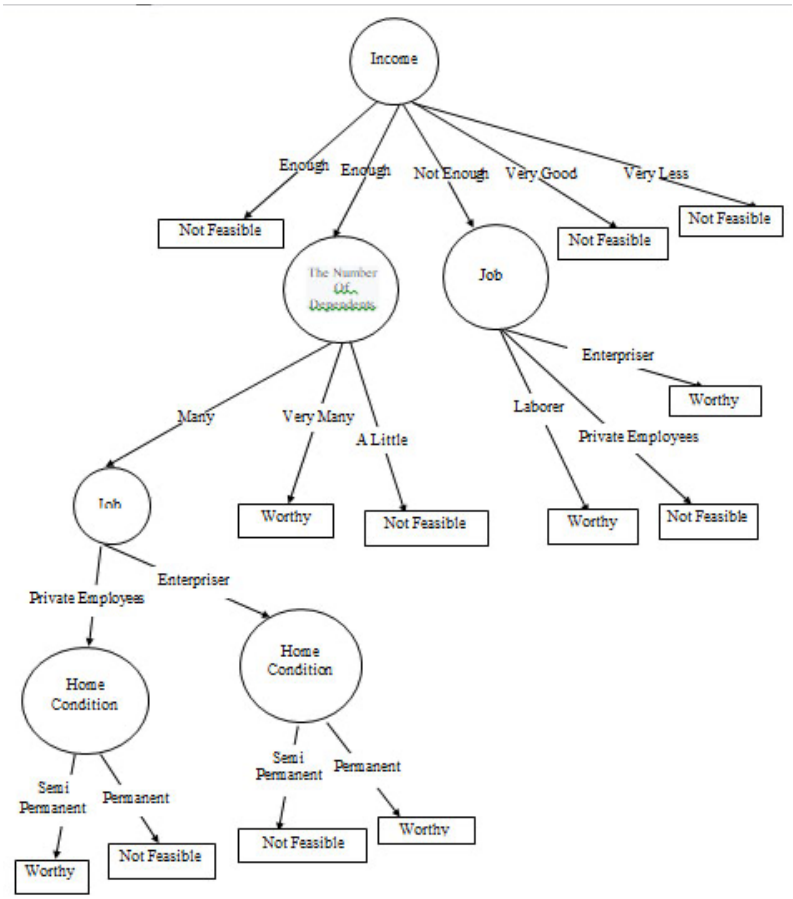

Figure 7. Decision Tree 6

In the calculation above, there are 12 rules that can be used as a reference in determining the main factors that influence the recipients of Social Assistance.

4. Conclusion

From the results of the research conducted, the conclusions that can be drawn based on the discussion of Social Assistance Recipients Using the C4.5 Algorithm are that the attribute that has the highest influence in determining the recipients of Social Assistance is Income. This is indicated by the Income attribute which occupies the root node in the decision tree diagram, and the Work attribute as the second factor in determining Social Assistance Recipients. It can be seen in the decision tree diagram that the Work attribute is located in a branch node under Income and based on the implementation of the test results with RapidMiner, a decision tree model is obtained that displays the rules that can be used for Social Assistance Recipients.

Acknowledgement

Acknowledgments to the supervisors and examiners who are lecturers at AMIK and STIKOM Tunas Bangsa who helped a lot in the process of compiling this research as one of the mandatory outcomes for undergraduate students to complete education at STIKOM Tunas Bangsa Pematangsiantar. I hope this research can be a reference for other researchers who use the C4.5 algorithm as a tool in solving problems in the research they are doing.

References

[1] H. Y. Dama *et al.*, "PENGARUH PRODUK DOMESTIK REGIONAL BRUTO (PDRB) TERHADAP TINGKAT KEMISKINAN DI KOTA MANADO," vol. 16, no. 03, pp. 549–561, 2016.
[2] P. Febrianti *et al.*, "TERHADAP ANAK TERLANTAR DI PANTI SOSIAL ASUHAN ANAK (PSAA) PUTRA UTAMA 03 TEBET," 2014.
[3] M. Luthfi, "Efektifitas Bantuan Sosial Program Keluarga Harapan Dalam Meningkatkan Kesejahteraan Keluarga," *J. Comm-Edu*, vol. 2, no. 1, pp. 81–89, 2019.
[4] K. Pintar, "' Virus Corona Jadi Pandemi Global ', Kelas Pintar, 1 April 2020," vol. 2019, no. April, 2020.
[5] D. I. Manggarai, "BANTUAN SOSIAL DAN PENDIDIKAN KESEHATAN BAGI MASYARAKAT PESISIR YANG TERDAMPAK SOSIAL-EKONOMI SELAMA PATOGENESIS COVID-19 Pendahuluan," vol. 16, no. 1, pp. 12–26, 2020.
[6] O. Nurdiawan, "Seleksi Penerima Bantuan Sosial Berdasarkan Sistem Pendukung Keputusan Dalam Upaya Mengurangi Siswa Rawan Putus Sekolah," vol. XIII, pp. 32–40, 2018.
[7] J. P. Gultom and A. Rikki, "Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan

Garoga untuk Menentukan Pola Penerima Beras Raskin," *Kumpul. Artik. Karya Ilm. Fak. Ilmu Komput.*, vol. 02, no. 01, pp. 11–19, 2020.

[8] M. M. M. ΘΕΟΔΩΡΟΥ, G. B. Paz, and A. A. B. Ruíz, "EFEKTIVITAS PROGRAM PEMERINTAH DAERAH TENTANG PEMBERIAN BEASISWA UNTUK MAHASISWA BERPRESTASI DAN NOT ENOUGH MAMPU DALAM KEBIJAKAN NOMOR 18 TAHUN 2017 DI KABUPATEN LUWU TIMUR i SKRIPSI," vol. 3, no. 2017, pp. 54–67, 2020, [Online]. Available: http://repositorio.unan.edu.ni/2986/1/5624.pdf.

[9] F. Harahap, "Penerapan data Mining dalam Pemilihan Mobil menggunakan Algoritma C4.5," *Ijccs*, no. x, pp. 11–20, 2018, [Online]. Available: https://voi.stmik-tasikmalaya.ac.id/index.php/voi/article/viewFile/99/42.

[10] A. S. Febri Wulandari, "Sistem Pengklasifikasian Pemilihan Penerima Beras Miskin (Raskin) Menggunakan Metode Naïve Bayes Classifier (Nbc)," *Progr. Stud. Tek. Inform. Fak. Teknol. Inf. Elektro*, 2019.

[11] M. Jannah, K. O. Putra, and L. Tambunan, "Penerapan Metode Analytic Network Process (ANP) Dalam Menentukan Penerima Bantuan Langsung Tunai (BLT)," *SATIN - Sains dan Teknol. Inf.*, vol. 07, no. 01, pp. 81–90, 2021, doi: 10.33372/stn.v7i1.708.

[12] S. Adi, "Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA Di Universitas Amikom Yogyakarta," *J. Mantik Penusa*, vol. 22, no. 1, pp. 11–16, 2018, [Online]. Available: http://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/342.

[13] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, "Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, pp. 87–96, 2018.

[14] A. A. Rahman and Y. I. Kurniawan, "Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan," *Progr. Stud. Inform. Univ. Muhammadiyah Surakarta*, 2016.

[15] Q. Iman and A. Wahyu, "Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/ Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine Classification of Poor Rice (Raskin)/ Prosperous Rice (Rastra) Recipient Hou," vol. 09, no. 2, pp. 178–184, 2021, doi: 10.26418/justin.v9i2.44137.

[16] P. Pascasarjana, "Program pascasarjana," p. 2012, 2012.

[17] T. Kenikan, K. Di, S. D. N. Citamiang, and K. Sukabumi, "Penerapan algoritma c.45 untuk klasifikasi tingkat kenikan kelas di sdn citamiang 2 kota sukabumi," 2020.

[18] M. Wati, B. Cahyono, U. Mulawarman, D. Mining, and B. K. Kartanegara, "Penerapan Algoritma C4 . 5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," vol. 2, no. 2, pp. 106–114, 2018.

[19] H. S. T. 3 Eka Satria Pribadi 1, Poningsih 2, "Analisa Tingkat KeSatisfiedan Masyarakat Terhadap Pelayanan Pengadilan Agama Pematangsiantar Menggunakan Algoritma," vol. 2, no. 1, pp. 33–40, 2020.

[20] T. Permana, A. M. Siregar, A. F. N. Masruriyah, and A. R. Juwita, "Perbandingan Hasil Prediksi Kredit Macet Pada Koperasi," *Conf. Innov. Appl. Sci. Technol.*, vol. 3, no. 1, pp. 737–746, 2020, [Online].Available:http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/1970.

[21] F. F. Harryanto, S. Hansun, U. M. Nusantara, G. Serpong, and C. Pegawai, "Penerapan Algoritma C4 . 5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," vol. 3, no. 2, pp. 95–103, 2017.

[22] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," vol. VI, no. 1, pp. 15–20, 2014.

[23] S. Takalapeta, "Penerapan Data Mining Untuk Menganalisis KeSatisfiedan Konsumen Menggunakan Metode Algoritma C4.5," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 3, pp. 34–38, 2018, doi: 10.37438/jimp.v3i3.186.

[24] S. Santoso and R. Nurmalina, "Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut)," *J. Integr.*, vol. 9, no. 1, pp. 84–91, 2017.

[25] W. D. Septiani, P. Studi, and M. Informatika, "DAN NAIVE BAYES UNTUK PREDIKSI PENYAKIT HEPATITIS," vol. 13, no. 1, pp. 76–84, 2017.