

MODEL BENEFICIARIES PROGRAM KELUARGA HARAPAN (PKH) OF LEBAK DISTRICT USING ALGORITHM C4.5

Iswadi¹, Dodi Rindiandiyah², Enting³, Muhammad Faisal⁴, Samsul Makin⁵.

¹Master of Computer Science Study Program, Budi Luhur University

²Master of Computer Science Study Program, Budi Luhur University

³Master of Computer Science Study Program, Budi Luhur University

⁴Master of Computer Science Study Program, Budi Luhur University

⁵Master of Computer Science Study Program, Budi Luhur University

Email : ¹iswadimista.pkh@gmail.com, ²dodirindiandiyah90@gmail.com,

³entingfirdaus80lebak@gmail.com,

⁴faisalmuhammadvin@gmail.com, ⁵samsulmakin25@gmail.com.

Abstract:

Poverty is one of the issues that were never completed, motherland Indonesia, since Indonesia's independence until now. To overcome these problems began in 2007 the Government of Indonesia through the Ministry of Social Affairs to implement the Program Keluarga Harapan (PKH) to ease the burden of life to poor families or family Beneficiaries (KPM), in Lebak Program Keluarga Harapan (PKH) starting in 2010, the Program Keluarga Harapan (PKH) is a program of the non-cash provision of social assistance conditional to a family Beneficiary (KPM) defined as poor. Program Keluarga Harapan (PKH) membrikan assistance to Beneficiaries Families (KPM), especially for poor families who have pregnant women, toddlers, elementary school children, middle school, high school, disabled and elderly,, with these criteria, the authors decided to use the algorithm C4.5 method. Model Beneficiaries Program Keluarga Harapan (PKH) will elaborate on the problem of multi factor or multi-criteria into a form of hierarchy to the recipient of the Program Keluarga Harapan (PKH) or KPM PKH, the C4.5 method test results in data mining will be conducted using 7 attributes , the result of the classification process is obtained that the C4.5 method has better accuracy rate, to provide assistance in the Program Keluarga Harapan (PKH) to KPM PKH, as for the kind of help a poor family was ranked bottom with the output from this method are eligible C4.5 or noneligible, beneficiaries Program Keluarga Harapan (PKH) that is provided to KPM PKH deemed unfit,

Keywords — **C4.5, Program Keluarga Harapan (PKH), Family Beneficiaries (KPM).**

I. INTRODUCTION

RI Ministry of Social Affairs Year (2015: 1) The Family of Hope Program is a social assistance program for Beneficiary Families (KPM) which are considered as poor or low-income families. PKH in Indonesia began in 2007. The poverty program known internationally as the Conditional Cash Transfer (CCT) which aims to increase poverty levels in developing countries.

Program Keluarga Harapan Harcommonly referred to as PKH, the program aims to improve

the socio-economic conditions to the families of the Program Keluarga Harapan Beneficiaries or KPM, in order to improve the education of children KPM and improve the health and nutrition status of pregnant women and children under five KPM. In the end, PKH is expected not only able to reduce poverty and improve human resources, especially in extreme poverty, but it can also break the chains of poverty itself.

Each state definitely has a problem each, are no exception Indonesia has some problems, such as in the socio-economic field in particular is the

problem of poverty. The poor are the primary concern of Indonesia command because considered shortcomings and inadequate in the face of the current globalization era growth. With poverty adany it will hinder the goals and ideals of the State of Indonesia.

Meanwhile, the government in carrying out its functions make efforts in tackling the problem of poverty. Poverty issues related to other aspects such as health, education, social, economic, cultural and other aspects. Basically, the poor have a weakness in the ability to make ends meet and the ability to strive and limited access to social and economic activities so far behind other communities that have a higher potential.

Program Keluarga Harapan (PKH) in Lebak district began in 2010 was conducted in 27 districts, at the present time throughout the districts in the Valley have received assistance PKH, as the number District in Lebak totaled 28 districts, in this study the authors take the data in 2017 as many as 35 017 KPM, research C4.5 algorithms have all the data that have received assistance Program Keluarga Harapan (PKH).

Table: 1 Number of Recipients PKH

DISTR ICTS	TOTAL KPM PKH 2017							DIS
	JML KPM	Pre gna nt wo men	CHI LD RE N	SD	SM P	Hig h Sch ool	EL DE RL Y	
LEBA K	35, 017	542	13. 268	31. 224	17. 399	6.2 80	1,4 65	282

Therefore there is need for efforts to tackle the problem of poverty in the form of policies such as development programs. As was explained earlier that one of the problems of the developing world is in the socio-economic field in particular is the problem of poverty. With this, the government needs to develop a strategy to address the problem of poverty. In order menanggulangan based household poverty, the government launched a special program called the Program Keluarga Harapan (PKH)

Inhibiting factors of CCT program in Lebak is a factor of poor families who want to become members PKH PKH and want to get help, because

there are many people in Lebak are not understand against the PKH aid program. Many of the recipients are already considered capable PKH then at Graduation / Non Eligiblekan (increased economic revenue), but sometimes do not want to get out of the membership.

From commentary The above will do an analysis of the data to obtain PKH PKH recipients information in Lebak using data mining techniques C4.5 Selection deemed appropriate as a key prediction accuracy of the results of a case being analyzed, namely Model Beneficiaries Program Keluarga Harapan (PKH) will elaborate on the problem of multi factor or multi-criteria into a form of hierarchy to the recipient of the Program Keluarga Harapan (PKH), the test results classification techniques of data mining will be conducted using 7 attributes, the results of the classification process is obtained that the C4.5 method has better accuracy rate, to provide assistance in the Program Keluarga Harapan (PKH), as for the kind of help a poor family was ranked bottom with output results or noneligible eligible, beneficiaries Program Keluarga Harapan (PKH) is not feasible, if already considered established in income / economy, or did not have the component as a key condition to get the help of the Program Keluarga Harapan (PKH).

II. LITERATURE REVIEW

A. Understanding Of PKH

According to the (Ministry of Social RI (2015) Program Keluarga Harapan (PKH) is a program of conditional cash assistance to Family Beneficiaries (KPM) qualified membership and established by the Ministry of Social Affairs. The general objective of CCT is to improve the quality of human resources, change behavior PKH participants were less supportive of efforts to improve the well-being, and to break the intergenerational transmission of poverty.

Hikmat (2017) Program Keluarga Harapan (PKH) is a conditional program providing social assistance to poor families (KM) which is defined as a family of CCT beneficiaries. In international terms is known as Conditional Cash Transfers (CCT).

So based on the above sources of exposure can be concluded that the Program Keluarga Harapan (PKH) is a program of non-cash money to the family of Beneficiaries (KPM) under the terms and conditions set by carrying out its obligations. This kind of program is internationally known as a program of conditional cash transfers (CCT) or the Conditional Cash Transfer program. Such requirements may include attendance at educational facilities (eg, school-age children), or attendance at health facilities (eg, for infants and children, or for pregnant women), which should be done by Families Beneficiaries (KPM).

B. Legal basis

1. Law Number 11 Year 2009 on Social Welfare;
2. Law No. 13 Year 2011 on Poverty Management;
3. Government Regulation No. 39 Year 2012 on the Implementation of Social Welfare;
4. Government Regulation No. 63 of 2017 on the distribution of Social Assistance in Non-Cash
5. Social Affairs Minister Regulation No. 20 / HUK / 2015 on Organization and Administration of the Ministry of Social;
6. Social Affairs Minister Regulation No. 10 / HUK / 2017 on the Program Keluarga Harapan;
7. Decree of the Director General of Protection and Social Security Number 12 / LJS / 09/2016 on general guidelines for PKH.

C. Interest of PKH

PKH general purpose is to improve the quality of human resources, change the behavior of participants of CCT is less support efforts to improve the well-being, and to break the intergenerational transmission of poverty. Specifically of interest is as follows PKH:

1. Improving the quality of health KPM;
2. Improve the education of children KPM;
3. Improving access to and quality of education and health services, especially for children KPM;

With specific goals are expected to improve the Human Development Index (HDI) for participants PKH.

D. Terms of membership PKH

Terms of membership PKH is Beneficiaries Families (KPM) which have components that have signed agreements PKH PKH participants as well as established by the Ministry of Social Affairs.

E. Components PKH

PKH components consisting of:

1. Pregnancy / puerperal mother;
2. Children aged under five years (Balita);
3. Elementary school children and equal;
4. Junior high school or equivalent;
5. High school or equivalent;
6. Children with disabilities;
7. Elderly.

III. THEORETICAL BASIS

A. Data Mining

According to Vercellis (2009) the term Data Mining is collected on all processes consisting of data analysis and analysis, models of inductive learning development and adoption of decision-making practices and ways based on acquired knowledge.

According to Han, Jiawei (2011), data mining is the process of finding interesting patterns and knowledge from large amounts of data. Meanwhile, according to Dunston and Yager (2008, p188), data mining is the process of searching through large amounts of data, in an attempt to discover patterns, trends, and relationships.

Based on some understanding of the above, it can be concluded that data mining is a process of analysis to explore the hidden information by using statistical and artificial intelligence in a database with a very large size, so it found a pattern of data that was previously unknown, and the pattern represented with computer graphics to be easily understood. The process of analyzing the data that many with the aim of finding an answer to be useful information in making decisions.

B. C4.5 Algorithm Method

According to Soewono (2014) C4.5 algorithm is one of the algorithms used in solving problems in the method of decision tree (decision tree) which is

a classification of data mining techniques. Method of decision tree (decision tree) is a method that changes the fact that a very large into a decision tree which represents the rules. Data in the decision tree is usually expressed in the form of tables with attributes and record.

According Soewono (2014) attribute declares a parameter known as the criteria in the formation of the tree. One attribute is an attribute that express the data solution per data item called a target attribute. Each attribute has a value that is called with the instance. Effective decision table is used when the conditions will be in the program is quite a lot and complicated.

According Kamagi and Hansun, (2014) The first step to building a decision tree is to choose an attribute as root, and then create a branch for each value and then divide the case into the branches. Repeat the process for each branch until all the cases to the branches have the same class

To select an attribute as roots, is based on the highest gain value from the existing attributes. To calculate the gain used formula as shown below:

$$Gain(S, A) = Entropy(S) - \sum_{[i=0]}^{|S_i|} \frac{|S_i|}{|S|} * Entropy(S_i)$$

Gain (S, A) = Entropy (S) -

Information:

S: The set of cases

A: Attributes

n: number of partitions attribute A

| Si |: The number of cases in the partition to-i

| S |: The number of cases in S

Before getting value Gain is to find the value of entropy. Entropy is used to determine how informative an input attribute to generate an attribute.

The basic formula of Entropy is as follows:

$$Entropy(S) = \sum_{i=0}^n - p_i * \log_2 (p_i)$$

Information:

S: The set Case

N: The number of partitions S

pi: The proportion of Si to S

IV. RESEARCH METHODS

A. Research Methods

Research method or methods of research (research method) In general, the definition of the research method is a process or way chosen specifically to resolve the issue raised in the research. While understanding the research methodology is a science that explains how it should be a study done.

According to Dawson (2009, 87) there are four commonly used research methods that action research, experiments, case studies and surveys. In the context of this study using the experiment, which is a method that is carried out with reference to problem solving that includes collecting data, formulating hypotheses, testing hypotheses, interpreting results, and conclusions Berndtsson (2008, 27).

This type of experiment is divided into two, namely the absolute experimental and comparative experiment. Absolute experiments lead to impacts resulting from the experiment. While the comparative experiment comparing two different objects, such as comparing two different algorithms to see each statistical results where better research Kothari (2004). The type of this research is an absolute experiment, the research that is testing, manipulate and influence the matters related to all variables or attributes.

This study if seen from the data and information managed, classified as quantitative research. Quantitative research is research that hypothesis can be tested by statistical techniques Moedjiono (2016).

B. Method of Collecting Data

This data collection techniques that data retrieved from Lebak District Social Service in the form of an excel format in 2017, with the following attributes:

1. pregnant women;
2. toddlers;
3. Elementary School student;
4. Junior High School student;
5. Senior High School student;
6. Elderly;
7. Disability.

Tabel 2 Example Beneficiaries PKH

NO	NO PKH	NAME MANAGEMEN	ADDRESS	EARTH	CHILD REN	SD	SMP	High Scho ol	ELDER LY	DIS
1	360117000400014	ANDSAH	KP. NORTH CITURJANG	0	0	0	0	0	0	0
2	360201000400001	DASIAH	KP. SIMPANG	0	0	1	2	0	0	0
3	360201000400003	UTI SUMIYATI	KP. SIMPANG	0	1	1	0	0	0	0
4	360201000400011	AREN	KP. SIMPANG	0	1	0	1	0	0	0
5	360201000400013	Majid	KP. SIMPANG	0	0	0	0	0	0	0

C. Mechanical Testing

Mechanical testing of the method that will be used *k-folds cross validation* with $k = 10$. This method of training data randomly split into 10 sections with nearly equal numbers in each group. At each iteration in the training process, the first part of the data used as test data and other data of 9 parts as training data. The training process performed 10 times. The test results will be obtained by calculating the average values of statistical evaluation and iteration.

V. MEASUREMENT OF RESEARCH

A more specific understanding of data mining, which is an application tool and uses statistical analysis on the data. Data mining is a process for the extraction or extracting data and information that is large, previously unknown, but it can be useful dipahamidan from large databases and used to make a very important keputusanbisnis.

Describe a collection of data mining techniques in order to find patterns that are not known at the data that has been collected. Data mining allows users to find knowledge in database data that may not be detected in the user.

Data mining can never be done without the availability of data to be calculated. That requires an understanding of what kind of data that will be used in the data mining process later. In this section we will discuss data processing.

A. Clustering and Data Analysis

In this study, the data used comes from the study entitled Using Data Mining for Predicting the recipient's models Assistance program family hopes (Data Mining to predict receiver models help family plan expectations), (Cortez & Silva, 2008) that has been uploaded to the repository UCI , The research data used in this study is a data set-hopes.csv-family program in the form of the report data model of CCT beneficiaries amounted to 35 017 records

and consists of 12 attributes with 7 attributes of type numeric, binary-type attribute 5. As for the attribute or independent variables used are (NoPKH, NamaUrus, Address, Pregnant women, toddlers, elementary school children, Children's junior high school children, Seniors, Disability, District, Village, eligible, noneligible).

Data will be grouped into two parts which is variable, data that meet the criteria CCT component, called eligible, if the data is not entered then enter noneligible PKH criteria. Data is built from two sources, a report accompanying the results in the field such as updating of KPM, and the operator proceeds used to supplement information from the data update companion must sign into the CCT as absenteeism / presence in the education and health of children KPM. The description of the attributes or variables to be researchers used are as follows:

Tabel 3 Variabel / Attributes Research

NO	VARIABLE / ATTRIBUTES	VALUE
1	NoPKH	numeric (1-16) Binary (M = male, F = female)
2	NamaUrus	binary (M = male, F = female)
3	Address	binary (U = Urban R = Rural)
4	Pregnant women	Numeric (0-9 months)
5	toddler	Numeric (0-6 years)
6	SD	Numeric (7-12 years)
7	SMP	Numeric (12-15 years)
8	High School	Numeric (16-18 years)
9	elderly	Numeric (60 years old)
10	disability	Numeric (0-9)
11	sub-district	binary (U = Urban R = Rural)
12	Ex / Village	binary (U = Urban R = Rural)

B. Validation and Evaluation

The main objective of this study was to determine the value of the accuracy of the C4.5 algorithm method. The results will be used to determine the accuracy of the two groups are eligible and Noneligible Which is best to improve the performance of the algorithm C4.5.

C. Model Design Method C4.5 Algorithm

By using two equations Method C4.5 algorithms it will get the value of entropy on the gain which is used as root in making the decision

tree. Counting the number of cases, amount to menentukan decision "Eligible", and the number of cases to determine the decision "Noneligible", as well as cases that are divided based on attributes that are used. After that, do the calculation of the gain for each attribute.

By using equation 4.1, we can calculate the value of entropy. Calculating the total entropy is done by counting the number of decisions "Eligible" and "Noneligible" of all cases.

$$\text{Entropy (total)} = ((-454 / 35,017) * \log_2 (454 / 35,017) + (-34,563 / 35,017) * \log_2 (34,563 / 35,017)) = 0.0998642$$

The total entropy is to calculate the total value of the decision "Noneligible" (454) and "Eligible" (34 563), while 35 017 is the total number of cases. Then calculate the entropy of each attribute value by using equation 4.2.

1. Entropy (Pregnant women) = $((-454 / 34,476) * \log_2 (454 / 34,476) + (- 34 022/34 476) * \log_2 (34,022 / 34,476)) = 0.10113352$
2. Entropy (Toddlers) = $((-454 / 23,026) * \log_2 (454 / 23..026) + (- 22 572/23 026) * \log_2 (22,572 / 2.3026)) = 0.139847727$
3. Entropy (SD) = $((-454 / 11766) * \log_2 (45.4 / 11 766) + (- 11 312/11 766) * \log_2 (11,312 / 11,766)) = 0.235769928$
4. Entropy (SMP) = $((-454 / 19,527) * \log_2 (454 / 19,527) + (- 19 073/19 527) * \log_2 (19,073 / 19,527)) = 0.159317951$
5. Entropy (SMA) = $((-454 / 29,134) * \log_2 (454 / 29,134) + (- 28 680/29 134) * \log_2 (28,680 / 29,134)) = 0.115864948$
6. Entropy (Elderly) = $((-454 / 33,642) * \log_2 (454 / 33,642) + (- 33 188/33 642) * \log_2 (33,188 / 33,642)) = 0.103160693$
7. Entropy (Disability) = $((- 454/34 751) * \log_2 (454 / 34,751) + (- 34 297/34 751) * \log_2 (34,297 / 34,751)) = 0.100483942$

Furthermore, by using equation (4.1) we can calculate the gain value of each attribute

1. Gain (Total, Bumil) = $(0.0998642) - ((34,476 / 35,017) * 0.10113352) - ((541 / 35,017) * 0) = 0.000293156$
2. Gain (Total, Toddlers) = $(0.0998642) - ((23,026 / 35,017) * 0.139847727) - ((10,844 / 35,017) * 0) - ((1,068 / 35,017) * 0) - ((79 / 35,017) * 0) = 0.007905044$
3. Gain (Total, SD) = $(0.0998642) - ((11,766 / 35,017) * 0.235769928) - ((16,755 / 35,017) * 0) - ((5,339 / 35,017) * 0) - ((1037 / 35,017) * 0) - ((120 / 35,017) * 0) = 0.020643565$
4. Gain (Total, junior) = $(0.0998642) - ((19,527 / 35,017) * 0.159317951) - ((13,758 / 35,017) * 0) - ((1,663 / 35,017) * 0) - ((69 / 35,017) * 0) = 0.011021591$
5. Gain (Total SMA) = $(0.0998642) - ((29,134 / 35,017) * 0.115864948) - ((5,524 / 35,017) * 0) - ((359 / 35,017) * 0) = 0.00346504$
6. Gain (Total, elderly) = $(0.0998642) - ((33,642 / 35,017) * 0.103160693) - ((1375 / 35,017) * 0) = 0.000754281$
7. Gain (Total, Disability) = $(0.0998642) - ((34,751 / 35,017) * 0.100483942) - ((2.66 / 35,017) * 0) = 0.000143565$

The calculation of entropy and gain for all the attributes to do, to get the highest gain value. The results of calculation of all the attributes shown in the table below.

Tabel 4 Entropy and Gain calculation determines

No	Attribute	value	Tot. Case	No. (S1)	Yes (S2)	Entropy	Gain
	TOTAL		35017	454	34563	0.0998642	
1	Pregnant Women						0.000293156
	35017	0	34476	454	34022	0.10113352	

2	Children						0.00 790 504 4
		0	23 026	454	22 572	0.13984 7727	
	35017	1	10 844	0	10 844	0	
		2	1,06 8	0	1,0 68	0	
		3	79	0	79		
3	SD						0.02 064 356 5
		0	11 766	454	11, 312	0.23576 9928	
		1	16 755	0	16 755	0	
	35017	2	5,33 9	0	5,3 39	0	
		3	1,03 7	0	1,0 37	0	
		4	120	0	120	0	
4	SMP						0.01 102 159 1
		0	19 527	454	19 073	0.15931 7951	
	35017	1	13 758	0	13 758	0	
		2	1,66 3	0	1,6 63	0	
		3	69	0	69	0	
5	High School						0.00 346 504
	35017	0	29 134	454	28 680	0.11586 4948	
		1	552 4	0	552 4	0	
		2	359	0	359		
6	Elderly						0.00 075 428 1
	35017	0	33 642	454	33 188	0.10316 0693	
		1	137 5	0	137 5	0	
7	DIS						0.00 014 356 5
	35017	0	34	454	34	0.10048	

			751		297	3942	
		1	266	0	266	0	

Based on the calculation results shown in Table 4.2 for data Entropy and Gain known that the attribute with the highest gain is elementary school children, so this attribute will be the root of the decision tree by C4.5 algorithm method. Attributes Children primary school had the highest scores compared with the attributes of the other, on the attribute value Son primary school have classified them into one, namely the decision Eligible so no need to do calculations again, while for other attributes, which are worth less or zero should be calculated at a later stage.

The calculation in the Attributes 1 can be described as a decision tree as shown in Figure 4.1 below.

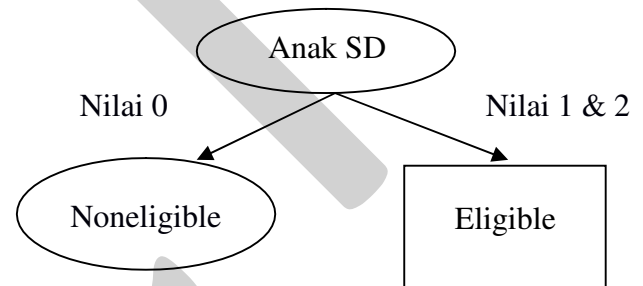


Figure 1 The decision tree Kids Gain calculation results SD

Pictured above is a decision tree be the root node is an attribute that has 2 branch facility that is in accordance with the value of the facility, the value Zero classify called Noneligible not receive assistance, while in grades 1 and 2 receive assistance called Eligible.

Design draft C4.5 algorithm in this study will be conducted using Weka application. The steps are as follows:

1. Open Tools Weka 3.8.3 or laptop computer that was installed weka,
2. Here is an early look weka.



Figure 2 Initial View Weka

3. To start using the Weka 3.8.3, select Explorer
4. Selected Open File-change to all files-select the data that is already stored in the folder -lalau will connect to the Open for the newly created database, or that have not been stored in the Weka 3.8.3.

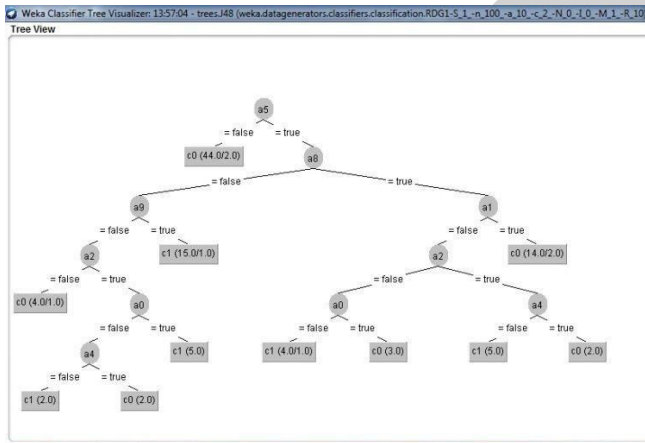


Figure 3 Display C4.5 Algorithm 3.8.3 Weka results

As a matter of proving the design of algorithm C4.5 method in this study will be carried out also proof by using a very simple application Microsoft Office Excel. The steps are as follows:

1. Open Applications Microsoft Office Excel, enter data as required.
2. Filter data to be calculated in the C4.5 Algorithm method.

Tabel 5 calculation method C4.5 Algorithm, in Microsoft Office Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	NO	NAMA	BUMI	BALI	SD	SMP	SMA	LANSIA	DIS	REKAMATAN	DESA	KET					
1	1	DASIAH	1	0	0	0	0	0	0	0	0	ELIGIBLE					
2	2	UTI SUMYATI	0	0	1	0	0	0	0	0	0	ELIGIBLE					
3	3	ARNI	0	1	0	0	1	0	0	0	0	ELIGIBLE					
4	4	ARNI	0	1	0	0	1	0	0	0	0	NONELIGIBLE					
5	5	MUJID	0	0	0	0	0	0	0	0	0	ELIGIBLE					
6	6	PIAH	1	0	0	0	0	0	0	0	0	ELIGIBLE					
7	7	YANI	0	1	4	1	0	0	0	0	0	ELIGIBLE					
8	8	ROSATI	0	2	2	1	0	0	0	0	0	ELIGIBLE					
9	9	ENCUK	0	1	0	0	0	0	0	0	0	ELIGIBLE					
10	10	ANAH	0	0	1	0	0	0	0	0	0	ELIGIBLE					
11	11	SONAH	0	1	1	1	0	0	0	0	0	ELIGIBLE					
12	12	ROHANAH	0	0	0	0	0	0	0	0	0	NONELIGIBLE					
13	13	JUBAEDAH	0	0	1	0	0	0	0	0	0	ELIGIBLE					
14	14	MISNAH	0	1	0	2	0	0	0	0	0	ELIGIBLE					
15	15	RUKMINAH	1	0	0	0	0	0	0	0	0	ELIGIBLE					
16	16	JUNI	0	0	0	0	0	0	0	0	0	NONELIGIBLE					
17	17	EPICH	0	0	1	0	0	0	0	0	0	ELIGIBLE					
18	18	SYAYAROH	0	0	0	0	0	0	0	0	0	NONELIGIBLE					
19	19	SAMINAH	0	0	1	0	0	0	0	0	0	ELIGIBLE					
20	20	ABID	0	0	1	0	0	0	0	0	0	ELIGIBLE					
21	21	IRIH	0	1	1	1	0	0	0	0	0	ELIGIBLE					
22	22	ACIH	0	1	1	1	0	0	0	0	0	ELIGIBLE					
23	23	DEDE	0	1	1	1	0	0	0	0	0	ELIGIBLE					
24	24	RIPAH	0	1	0	2	0	0	0	0	0	ELIGIBLE					
25	25	CICIH	1	0	0	0	0	0	0	0	0	ELIGIBLE					
26	26	HAMNAH	0	0	1	0	0	0	0	0	0	ELIGIBLE					

3. Next specify the attributes that will be the label, the validation method used to determine the accuracy of the C4.5 algorithm method in this study using a validation method tenfold cross-validation.

Tabel 6 Display C4.5 testing in Microsoft Office Excel

	A	B	C	D	E	F	G	H	I	J
	No	Nama	Bumi	Bali	SD	SMP	SMA	Lansia	Disabilit	Ket
1	1	DASIAH	1	0	0	0	0	0	0	ELIGIBLE
2	2	UTI SUMYATI	0	0	1	0	0	0	0	ELIGIBLE
3	3	ARNI	0	1	0	0	1	0	0	ELIGIBLE
4	4	ARNI	0	1	0	0	1	0	0	NONELIGIBLE
5	5	MUJID	0	0	0	0	0	0	0	ELIGIBLE
6	6	PIAH	1	0	0	0	0	0	0	ELIGIBLE
7	7	YANI	0	1	4	1	0	0	0	ELIGIBLE
8	8	ROSATI	0	2	2	1	0	0	0	ELIGIBLE
9	9	ENCUK	0	1	0	0	0	0	0	ELIGIBLE
10	10	ANAH	0	0	1	0	0	0	0	ELIGIBLE
11	11	SONAH	0	1	1	1	0	0	0	ELIGIBLE
12	12	ROHANAH	0	0	0	0	0	0	0	NONELIGIBLE
13	13	JUBAEDAH	0	0	1	0	0	0	0	ELIGIBLE
14	14	MISNAH	0	1	0	2	0	0	0	ELIGIBLE
15	15	RUKMINAH	1	0	0	0	0	0	0	ELIGIBLE
16	16	JUNI	0	0	0	0	0	0	0	NONELIGIBLE
17	17	EPICH	0	0	1	0	0	0	0	ELIGIBLE
18	18	SYAYAROH	0	0	0	0	0	0	0	NONELIGIBLE
19	19	SAMINAH	0	0	1	0	0	0	0	ELIGIBLE
20	20	ABID	0	0	1	0	0	0	0	ELIGIBLE
21	21	IRIH	0	1	1	1	0	0	0	ELIGIBLE
22	22	ACIH	0	1	1	1	0	0	0	ELIGIBLE
23	23	DEDE	0	1	1	1	0	0	0	ELIGIBLE
24	24	RIPAH	0	1	0	2	0	0	0	ELIGIBLE
25	25	CICIH	1	0	0	0	0	0	0	ELIGIBLE
26	26	HAMNAH	0	0	1	0	0	0	0	ELIGIBLE

D. C4.5 Algorithm Testing Results

The results of the testing method C4.5 algorithms in Microsoft Office Excel application are as follows:

Tabel 7 Calculation results of determining Eligible & Noneligible

ELIGIBLE	NON ELIGIBLE
34 563 (98.70%)	454 (1:30%)
35 017	

E. Application of C4.5 Algorithm

In this phase will be implemented method C4.5 algorithms which have higher accuracy in predicting aid Program Keluarga Harapan (PKH) to build a prototype prediction assistance Program Keluarga Harapan (PKH). By looking at the results of the evaluation and validation it can be seen that the C4.5 algorithm method has the best accuracy, so that the rule generated by the C4.5 algorithm method used as a rule for a prototype that can facilitate the prediction beneficiaries Program Keluarga Harapan (PKH). The rule used is the result of the processing methods C4.5 algorithms, with details aturanya rules are as follows:

1. Pregnant woman = IF (H2 = 1, "ELIGIBLE", IF (C2 = 0, "NONELIGIBLE", "-"))
2. CHILDREN = IF (OR (I2 = 1, I2 > 2), "ELIGIBLE", IF (D2 <= 0, "NONELIGIBLE", "-"))
3. SD = IF (OR (J2 = 1, J2 > 2), "ELIGIBLE", IF (E2 <= 0, "NONELIGIBLE", "-"))
4. SMP = IF (OR (K2 = 1, K2 > 2), "ELIGIBLE", IF (F2 <= 0, "NONELIGIBLE", "-"))
5. SMA = IF (OR (L2 = 1, L2 > 2), "ELIGIBLE", IF (G2 <= 0, "NONELIGIBLE", "-"))
6. ELDERLY = IF (OR (M2 = 1, M2 > 2), "ELIGIBLE", IF (H2 <= 0, "NONELIGIBLE", "-"))
7. DIS = IF (OR (N2 = 1, N2 > 2), "ELIGIBLE", IF (I2 <= 0, "NONELIGIBLE", "-"))

VI. CONCLUSION

Based on the problems, literature review, a review of research, review the object of research and research methodologies, data mining is a method for predicting the CCT Receiver Candidates in Lebak, it can be concluded as follows:

1. The research result prediction prospective beneficiaries Program Keluarga Harapan (PKH), there are still many who are able to still get help PKH, resulting in a survey to collect data not in the house to prospective beneficiaries of PKH, then this study was to determine the effectiveness of the performance of the method in the C4.5 algorithm Receiver Candidates predict PKH case in Lebak.

2. After doing research with C4.5 algorithm method gain accuracy rate above 95%, and the accuracy of the results obtained from the use of Data Mining Algorithm C4.5 method can be concluded, the data PKH including as many as 34 563 Eligible KPM, while the Noneligible the data as much as 454 KPM, of a total of as many as 35 017 KPM PKH recipient.
3. Accuracy of CCT research results using the C4.5 algorithm performs data validation at 95%, after using a prototype system that is created as a tool to be able to predict the level of accuracy of the prospective beneficiaries Program Keluarga Harapan (PKH) as expected.

VII. SUGGESTION

Based on the description of the discussion of research and conclusions obtained, it was suggested a few things for future research are as follows:

1. Application of C4.5 Algorithm method can be used to predict the CCT Receiver Candidates in Lebak.
2. This research can be developed with other classification algorithms, such as Neural Network algorithm, the K-Means or SVM (*Support Vector Machine*) by adding stages of the electoral attribute (feature selection) in order to obtain optimal results.

REFERENCE

- [1] Hastuti, K. & Hidayat, E. Y., 2013. Analisis Algoritma Decision Tree untuk Prediksi Mahasiswa Non Aktif. *Seminar Nasional Teknologi Informatika & Komunikasi Terapan*, p. 211.
- [2] Kementerian Sosial Republik Indonesia. (2015). *Pedoman Umum Program Keluarga Harapan tahun 2015*. Jakarta: Kementerian Sosial RI, Direktorat Jaminan Sosial Keluarga.
- [3] Buku kerja pendamping PKH (Program Keluarga Harapan), 2008. *Direktorat Jenderal Bantuan Sosial Dan Jaminan Sosial* Depsos RI. Jakarta.
- [4] Buku Pedoman umum PKH, 2008. *Direktorat Jenderal Bantuan Dan Jaminan Depsos RI*.
- [5] Direktorat Jaminan Kesejahteraan Sosial, Kementerian Sosial RI, 2009. *Pedoman Umum PKH*. Jakarta.
- [6] Kementerian Sosial Republik Indonesia. (2011). *Pedoman Umum Program Keluarga Harapan tahun 2011*. Jakarta: Kementerian Sosial RI, Direktorat Jaminan Sosial
- [7] Syarif, M., 2015. Penerapan Particle Swarm Optimization Untuk Menentukan Kredit Kepemilikan Rumah Dengan Menggunakan Algoritma C4.5. *Jurnal Techno Nusa Mandiri*, XII(2), p. 55.
- [8] Badrul, M., 2014. Perbandingan Algoritma C4.5 Dan Neural Network Untuk Memprediksi Hasil Pemilu Legislatif DKI Jakarta. *Jurnal Pilar Nusa Mandiri*, X(2), p. 128.
- [9] Cortez, P. & Silva, A., 2008. USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE. *Proceedings of 5th Future Business Technology Conference*, pp. 5-12

- [10] Ariyani, L., 2016. Kajian Penerapan Model C4.5, Support Vector Machine (SVM), Dan Neural Network Dalam Prediksi Kenaikan Kelas. *Faktor Exacta* 9, I(9), p. 85.
- [11] Guritno, S., Sudaryono & Rahardja, U., 2011. Theory and Application of IT Research-Metodologi dan Penelitian Teknologi Informasi, Yogyakarta: Andi.
- [12] Gorunescu, F. 2011. Data Mining Concept Model and Techniques. Berlin: Springer. ISBN 978-3-642-19720-8
- [13] Soewono, R., Gernowo, R. & Sasongko, P. S., 2014. Sistem Pakar Identifikasi Modalitas Belajar Siswa Dengan Implementasi Algoritma C4.5. *Jurnal Sistem Informasi Bisnis*, Volume I, p. 21.
- [14] Han, J., & Kamber, M. 2006. Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3
- [15] Mu'arif, K., 2009. Komparasi Pemodelan Data Menggunakan C4.5 Dan C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Kelulusan Mahasiswa. *Journal of Bionic Engineering*, p. 7.
- [16] Kusriani, & Luthfi, E. T. (2009). Algoritma Data Mining. Yogyakarta: Andi Publishing.
- [17] Larose, D.T. & Larose, C.D., 2014. Discovering Knowledge in data An Introduction to Data Mining.
- [18] Lusiana, S.A., 2014. Algoritma C4.5 dalam menganalisa kelayakan kredit(studi kasus dikoperasi pegawai Republik Indonesia(KP-RI)) Lembang Pesisir Selatan, Painan, Sumatera Barat. , 1(2), pp.6–10.
- [19] Liao, T. W., & Triantaphyllou, E. (2007). Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications. Series on Computers and Operations Research, Vol 6. USA: World Scientific. ISBN-13 978-981-277-985-4, ISBN-10 981-277-985-X
- [20] Larose, T.D., 2005. Discovering Knowledge in Data an Introduction to Data Mining, New Jersey: John Wiley & Sons, Inc.
- [21] Vercellis, C., 2009. Business Intelligence : Data Mining and Optimization for Decision Making, John Wiley & Sons, Inc.
- [22] Widodo, P.J. et al., 2013. Sistem Interkoneksi Data Antar Unit Guna Mendukung Keberhasilan Akreditasi Program Studi Pada Fakultas Teknik UNS. , 3(1), pp.1–9.
- [23] Bramer, Max. 2007. Principles of Data Mining. London: Springer. ISBN-10: 1-84628-765-0, ISBN-13: 978-1-84628-765-7.
- [24] Derick Iskandar, Yoyon K Suprpto Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C4.5 Dan Naïve Bayes Jurusan Teknik Elektro, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember Surabaya 2015