

PAPER • OPEN ACCESS

Forecasting Student Graduation With Classification And Regression Tree (CART) Algorithm

To cite this article: A Maesya and T Hendiyanti 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **621** 012005

View the [article online](#) for updates and enhancements.

Forecasting Student Graduation With Classification And Regression Tree (CART) Algorithm

A Maesya¹ and T Hendiyanti¹

¹Faculty of Mathematics and Natural Sciences, Department Computer Science , Pakuan University
Jalan Pakuan PO.BOX. 452, Ciheuleut Bogor 16143

e-mail : a.maesya@unpak.ac.id , tissa.hendiyanti@gmail.com

Abstract. Forecasting Student Graduation with *Classification And Regression Tree* (CART) Algorithm aims to predict students' graduation in Computer Science Departement of Pakuan University. The method used is the CART Algorithm where this method works by forming a decision tree that can infer the rules of a particular classification of the decision tree. The results of this study resulted ouput in the form of decision trees and determine the students with predetermined criteria are classified in categories such as pass on time or pass not on time. Assessment of the accuracy of the data classification that has been done using confusion matrix validation technique. Once calculated with the matrix, the accuracy rate reached 77.5% in the training data.

1. Introduction

The development of technology and information has a very important role to support all work activities to be more optimal. Application of information technology can be used in data processing such as prediction or forecasting. In the process of processing data there are stages such as cleaning, selection, combination, until the data processed can be presented as a new information, this is called knowledge discovery in database (KDD). KDD is a series of processes to gain knowledge or patterns from data sets known as data mining [2]. Process of data mining include classification, association rule, and clustering. In the classification technique in describing the learning to do prediction on a value of the target in the category variable, one way in the classification that is often used is the decision tree. One of the methods contained in the decision tree is Classification And Regression Tree (CART) algorithm. This method works by forming a decision tree that can infer the rules of a particular classification that has a way of making decisions with regression activities on the data. Relevant studies related to predictions have been made by previous researchers, Mardiani [3] conducted the application of classification with the CART algorithm for predicting lectures for new students (Study Cased : STMIK MDP Palembang). This study uses data classification classification functionality to predict which students will continue to study and which do not continue in the next semester. Based on these studies formed a new study to forecasting students' graduation in Department Computer Science, Pakuan University.

2. Methods

The research method applied to Forecasting Student Graduation With Classification And Regression Tree (CART) Algorithm is the method of KDD (Knowledge Discovery In Database).



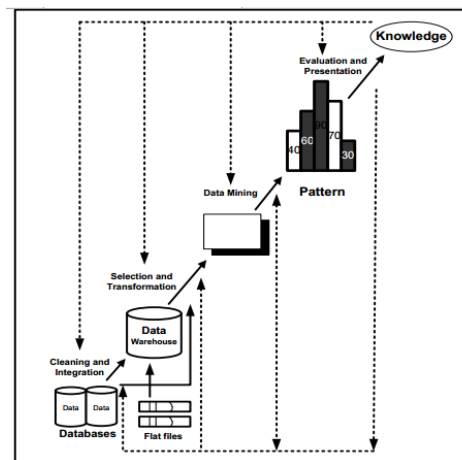


Figure 1. KDD Approach [1]

2.1 Cleaning and Integration Data phase

This process is an action to remove noise and inconsistent data or irrelevant data. The data set of students can be from PUTIK (Center of Information and Communication Technology) based on student data in SIMAK (Information System Academic) Pakuan University.

2.2 Selection and Transformation Data Phase

Data available on the database is often not all can be used, therefore the data to be retrieved is only relevant and relevant data. Later this relevant data will be the task of analysis to be returned into the database. Data is altered or merged into the appropriate format for processing in data mining.

2.3 Data Mining Phase

In this phase, different types of modeling techniques are selected and applied and the parameters calibrated to obtain optimal results. There are usually several techniques for the same type of data mining problem. Some techniques also have a need for a specific data form. Therefore, often the process of data preparation is needed again. The mining process is done by classifying the data with CART algorithm.

2.4 Pattern Evaluation Phase

The process of identifying patterns that represent knowledge based on multiple measures. The evaluation activities of the research measure the accuracy of CART algorithm using coincidence (confusion) matrix.

2.5 Knowledge Presentation Phase

Visualization and representation techniques a knowledge or information to the user. The last stage of the data mining process is how to formulate the decisions or actions of the analysis results obtained. Sometimes this should involve people who do not understand data mining. Therefore the presentation of data mining results in the form of knowledge that can be understood by everyone is a necessary step in the process of data mining.

3. Results

The results obtained from this research is the making of classification system of students' graduation forecasting of Departement of Computer Science, Faculty of Mathematics and Natural Sciences of Pakuan University using CART algorithm function in machine learning of SPSS Modeler application. The data is processed so that it will produce classification in the form of decision making tree when the student graduation is timely or not. The design of this system using Visual Studio 2015 application with

the management of a. NET framework and database system using XAMPP (MySQL) application. Because it is not necessary to process data after predicting, this system can be used by anyone to obtain information on predictions of students graduation.

3.1 Page Input Test Data

The student data entry form will be classified using CART algorithm, resulting in the output in the form of the graduation level of the student pass, ie graduated on time (TW) or not on time (TTW). In the input menu test data can be directly viewed data entered in the table below the form fields. shown in Figure 2.

npm	nama	ssk	jk	asal	agama	sekolah	umur	prediksi	rule
005113007	M. HOFWAN YUS	3.05	L	JAWA	Islam	SMA	19	TW	4
005113007	SONY SETIA RA	2.40	L	JAWA	Islam	SMA	21	TW	1
005113008	MUHAMMAD AL	3.35	L	JAWA	Islam	SMK	18	TTW	13
005113010	NIDA WIDHANN	3.52	P	JAWA	Islam	SMA	19	TTW	15
005113011	TRI MUHAMMAD	3.23	L	JAWA	Islam	SMK	18	TTW	13
005113012	BENNY NORKA	3.23	L	JAWA	Islam	SMK	20	TTW	13
005113013	ANDI EKAP P.	3.02	L	SULAWESI	Islam	SMA	18	TW	4
005113014	REFARDI SEPTI	2.49	L	JAWA	Islam	SMK	19	TW	1
005113015	DWI RANDANI	2.25	L	JAWA	Islam	SMA	20	TW	1

Figure 2. Display Page Input Test Data

In the input view of this test data, the user can retrieve data with the format .xlsx (excel worksheet) by clicking the "Import" button. The display of import menu is shown in Figure 3.

npm	nama	ssk	jk	asal	agama	sekolah	umur	prediksi
005113007	M. HOFWAN YUS	3.05	L	JAWA	Islam	SMA	19	TW
005113007	SONY SETIA RA	2.40	L	JAWA	Islam	SMA	21	TW
005113008	MUHAMMAD AL	3.35	L	JAWA	Islam	SMK	18	TTW
005113010	NIDA WIDHANN	3.52	P	JAWA	Islam	SMA	19	TTW
005113011	TRI MUHAMMAD	3.23	L	JAWA	Islam	SMK	18	TTW
005113012	BENNY NORKA	3.23	L	JAWA	Islam	SMK	20	TTW

Figure 3. Display Page Import Test Data

3.2 Results of Classification Tree

The view of the classification tree provides an overview of the classification results of the previous training data. From the amount of data of 200 students, yielding decision tree and shown in Figure 4.

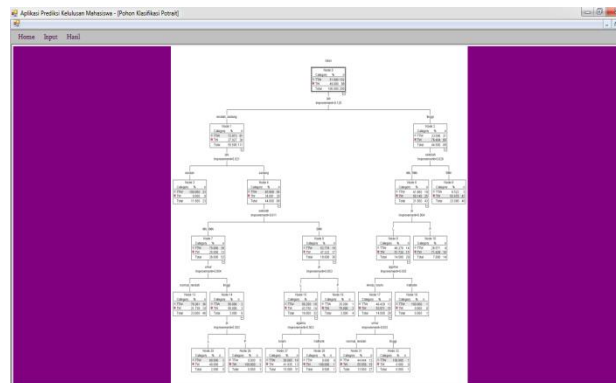


Figure 4. Display Results of Classification Tree

3.3 Result of Rule

The rules generated from the classification of training data 26 rules generated is shown in Figure 5.

No	gk	jk	asal	agama	pendah	umur	pendidikan
1	berand						ITW
2	Seling			RAA	normal		ITW
3	Seling			RAA	berdah		ITW
4	Seling			RAA	normal		ITW
5	Seling			RAA	berdah		ITW
6	Seling	L			inggi		TW
7	Seling	P			inggi		TW
8	Seling	L			RAA		ITW
9	Seling	P			inggi		ITW
10	Seling	L	Meen	DAK			ITW
11	Seling	L	Kathulu	DAK			TW
12	Seling	P		DAK			TW
13	inggi			DAK			TW
14	inggi	P		RAA			TW
15	inggi	P		RAA			TW
16	inggi	L	Meen	RAA	berdah		TW
17	inggi	L	Meen	RAA	normal		TW
18	inggi	L	Meen	RAA	berdah		TW
19	inggi	L	Meen	RAA	normal		TW
20	inggi	L	Meen	RAA	inggi		ITW
21	inggi	L	Meen	RAA	inggi		ITW
22	inggi	L	Hindu	DAK	berdah		TW
23	inggi	L	Hindu	DAK	normal		TW
24	inggi	L	Hindu	DAK	inggi		ITW
25	inggi	L	Kathulu	DAK			ITW

Figure 5. Display Result Rule of Test Data

3.4 Result Forecasting

The results of activities forecasting students' passing of the training data and test data are presented in Figure 6.

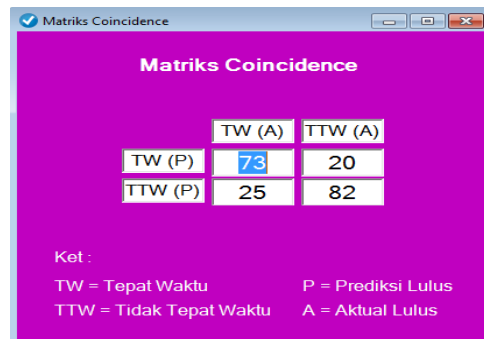
id	nama	th	jk	jenis	agama	sekolah	umur	prediksi	lulus
05010001	DESI PRADITA	213	L	Jawa	Islam	SMA	18	TTW	TTW
05010010	DESI PRADITA	213	L	Jawa	Islam	MA	18	TTW	TTW
05010004	LENDY TRIAN	242	L	Jawa	Islam	SMA	19	TTW	TTW
05010008	UNHARMAN TAR	281	L	Sunda	Katolik	SMA	18	TTW	TTW
05010023	GYAN LEE	279	L	Jawa	Islam	SMK	17	TTW	TTW
05010026	ANISHA KOMAL	272	P	Jawa	Islam	SMA	18	TTW	TTW
05010001	HERLY SAPULU	142	L	Jawa	Islam	SMK	18	TW	TTW
05010009	RAHMAT HOSNY	276	L	Jawa	Sekular	SMK	18	TTW	TTW
05010041	ENDANG PERMA	281	L	Jawa	Islam	SMK	21	TTW	TTW
05010042	DEDE NERIDA	319	L	Jawa	Islam	SMK	18	TTW	TTW
05010051	MUHAMMAD FA	308	L	Jawa	Islam	SMK	18	TTW	TTW
05010054	FABRI WITANTO	303	L	Jawa	Islam	SMA	18	TTW	TTW
05010066	PUJAKRATI AL	106	P	Jawa	Islam	SMA	18	TTW	TTW
05010067	RIZKY DWI SET	230	L	Jawa	Islam	SMA	18	TTW	TTW
05010068	RAMA SCHIR	318	L	Jawa	Islam	SMK	19	TTW	TTW
05010069	NELIA PERTWI	315	P	Jawa	Islam	SMA	17	TTW	TTW
05010075	FANNY ANNISA	305	P	Jawa	Islam	SMA	18	TTW	TTW
05010076	AHMAD YADIN	279	L	Jawa	Islam	SMA	18	TTW	TTW
05010083	RULAL MUMIN	236	L	Jawa	Islam	SMK	18	TTW	TTW
05010084	HOLD MAWAZDI	271	L	Jawa	Islam	MA	19	TTW	TTW
05010087	MUHAMMAD RI	308	L	Jawa	Islam	SMA	18	TTW	TTW
05010089	AHMAD BANU	241	L	Jawa	Islam	SMA	18	TTW	TTW
05010098	RIFKI MULIANA	260	L	Jawa	Islam	SMA	17	TTW	TTW

Figure 6. Display Result Forecasting

3.5 Test Validation

After the training data is processed, the analysis activity is done so that the results obtained in accordance with the desired design, while the analysis on this menu consists of analysis of

coincidence matrix (confusion), accuracy and prediction of test data. Each analysis result is presented in Figures 7, 8, and 9.

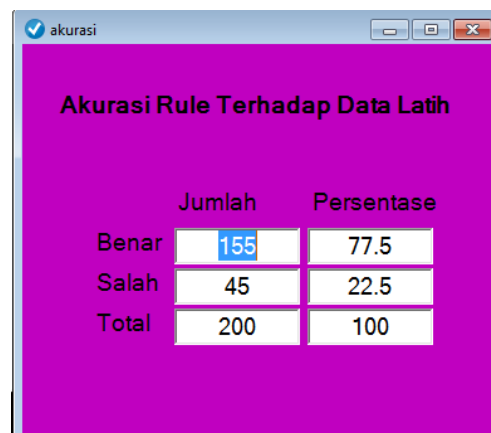


Matriks Coincidence

	TW (A)	TTW (A)
TW (P)	73	20
TTW (P)	25	82

Ket :
 TW = Tepat Waktu P = Prediksi Lulus
 TTW = Tidak Tepat Waktu A = Aktual Lulus

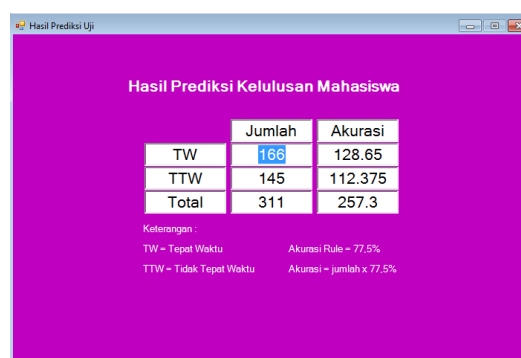
Figure 7. Display *Coincidence (Confusion) Matrix*



Akurasi Rule Terhadap Data Latih

	Jumlah	Persentase
Benar	155	77.5
Salah	45	22.5
Total	200	100

Figure 8. Display *Accuracy of Train Data*



Hasil Prediksi Kelulusan Mahasiswa

	Jumlah	Akurasi
TW	166	128.65
TTW	145	112.375
Total	311	257.3

Keterangan :
 TW = Tepat Waktu Akurasi Rule = 77.5%
 TTW = Tidak Tepat Waktu Akurasi = jumlah x 77.5%

Figure 8. Display *Result of Test Data*

4. Conclusion

Forecasting Student Graduation with Classification And Regression Tree (CART) Algorithm aims to predict students' graduation in Computer Science Departement of Pakuan University. The method used is the CART Algorithm where this method works by forming a decision tree that can infer the rules of a particular classification of the decision tree. The results of this study resulted output in the form of decision trees and determine the students with predetermined criteria are classified in categories such as pass on time or pass not on time. Assessment of the accuracy of the data classification that has been done using confusion matrix validation technique. Once calculated with the matrix, the accuracryate reached 77.5% in the training data.

References

- [1] Han J and Kamber M 2012 *Data Mining Concepts and Techniques 2nd edition*. Mofgan Kaufhann Publishers, San Fransisco.
- [2] Witten, Ian H, Frank, Eibe and Hal MA 2011 *Data Mining: Pratical Machine Learning Tools and Techniques*, Third Edition. Burlington: Morgan Kaufmann Publishers.
- [3] Mardiani 2012 *Application of Classification With CART Algorithm For Lecture Prediction For New Students (Case Study: STMIK MDP Palembang)*. Palembang. Thesis.