

Implementing CBR on The College Rankings Based on Webometrics with EPSBED's Data and Webometrics Knowledge

Marzuki^{#1}, Maria Shusanti F^{#2}, Ahmad Cucus^{#3}, Agus Sukoco^{#4}

[#]Faculty of Computer Science, University of Bandar Lampung
Jl. ZA.PagarAlamLabuhanRatu Bandar Lampung, Indonesia

¹marzukird@yahoo.com

²cucus_ahmad@yahoo.com

³suzanne_maria1986@yahoo.com

⁴agussukoco19@gmail.com

Abstract—inclusion of the name of a university in a ranking system, especially at the top level according to a specific community is important and worth doing. The consideration, including the entry of a university in the list of top universities, is an excellent starting point for the world to recognize the existence of higher education, to boost the image of the university, will also improve promotion, especially in student acquisition and external support.

Webometrics is one of the world university ranking system that can be enjoyed by universities in Indonesia. Of the publication in July 2009 a total of 39 universities and the publication in January 2010 as many as 57 universities ranked 6000 Indonesia succeeded in the world. This ranking technique to calculate the accessibility of university websites and publications on google scholar as the ranking parameter. However, Webometrics do not Rank its based on a study program of higher education. Indonesia has more than 3000 universities which have an operating permit and each university has a study program. Objectivity in the performance of a university can be controlled through the data formulated academic activities through program Evaluation Based Self-Evaluation Study (EPSBED). EPSBED Data published to the public through the website [http: \ \ evaluasi.or.id](http://\evaluasi.or.id). whereas until now Indonesia has not had a rating system.

Case-Based Approach Computing Reasoning (CBR), combined with machine learning algorithms its Nearest Neighbor (NN) as the ranking methodology that uses data EPSBED, then the ranking of a university system may be based on their individual study program.

Keywords— Case-Based Reasoning, Nearest Neighbor, EPSBED, classification algorithms, AI

I. INTRODUCTION

The recognition that the so-called international institutions, highly desirable range of Higher Education. Not only in Indonesia, but also in different countries. Among other forms of recognition by the Nobel Prize, an international, up to university ranking is based on certain indicators. Higher ranking is based on certain indicators that have been done a few organizations or institutions. The ranking is quite recognized in the world there are several other Academic.

Ranking of World University (ARWU), the Times Higher Education Supplement (THES) dan Webometrics.

According to Romi, [1] inclusion of the name of a university in the rankings, especially at the top level according to a specific community is important and worth doing. The consideration, including the entry of a university in the list of top universities, is an excellent starting point for the world to recognize the existence of higher education, to boost the image of the university, will also improve promotion, especially in the acquisition and external support students in the study. It was also one of the tangible manifestation of the successful efforts of the National Education program.

Currently there is the rating that can be enjoyed-College Education in Indonesia. Of the publication in July 2009 a total of 39 universities in Indonesia managed to occupy the 6000 ranking of the world. College of Engineering ranking is calculated from the accessibility of university websites and publications on google scholar as the ranking parameter. Such ranking is done by Webometrics.

Indonesia has the 3099 Higher Education, but if the Webometrics ranking is based on the data above, shows that only 1.25% Higher Education Indonesia to enter the ranking system. According to Romi [1], low rank Indonesia Higher Education, Higher Education while Indonesia has parameters that can be used as indicators of the ranking.

From the above facts can be identified the problem as follows:

- There are currently no College Ranking System Based on Data EPSBED*
- There are currently no models and criteria to rank universities in Indonesia*

Case-Based Reasoning (CBR) is a problem solving technique based on the knowledge and experience Aamodt and Plaza [2], describes the type of CBR as a circular process that consists of 4 (four) stages.

- Retrieve, get the cases compared with a collection of similar cases in the past.
- Reuse, reuse of existing cases and try to solve a problem now.
- Revise, amend and adopt the proposed solution if necessary.
- Retain, continue to use the last solution as part of the new cases.

Case-Based Reasoning methodology will be used for evaluating indicator-style ranking by Webometrics EPSBED parameter data, so it will get a higher education system has pemeringkaan indicators based on data and facts are valid. The purpose of the research Implementation Case-Based Reasoning to rank the data-based National Universities Webometrics EPSBED and aims to:

- Build and test the Intelligent Systems that use the data and the data EPSBED Webometrics ranking to rank universities in Indonesia.
- Evaluating Case-Based Reasoning Methodology for Ranking the models using data EPSBED as a criteria to rank universities. The benefits of this research are as follows:
 - Practical benefits of this research is that by implementing the Case-Based Reasoning in a data-based Higher Education ranking EPSBED may be an alternative in the ranking universities without re-weighting to each value of the parameter
 - Theoretical benefits of this research are expected to contribute to the development of a method of ranking institutions
 - As a reference for further research, particularly with respect to the ranking of an institution.

II. MODEL, ANALYSIS, DESIGN, AND IMPLEMENTATION

The study was conducted to provide a solution to the lack of university ranking systems, as well as the lack of models and criteria that can be used as an indicator of the ranking.

Case-Based Reasoning (CBR) is a method in solving problems that occur, based on the case (case) has ever seen. This method since it was first coined, has been widely implemented in the computational problems are complicated and complex. Webometrics is one example of an international ranking published in January and July on the internet. Publication month of July 2009 there were 39 Webometrics ranking Universities in Indonesia who entered the world university ranking 6000. Parameters used in evaluating universities Webometrics uses 4 (four).

This study was to implement the CBR method that uses Case-Based form of the model containing the knowledge that there are 39 universities in the Webometrics ranking of the parameters have been added to the data EPSBED. Then the system will test the similarity of each university with Case-Based Model.

All data of Higher Education in the country who have made Webometrics ranking on General Knowledge. While the High

College to be calculated to measure the proximity to the rank of General Knowledge Nearest Neighbor algorithm (NN). NN is an algorithm to calculate the distance matrix between two pieces of data based on attribute data is owned by Siu, [3] So that the research approach in developing the rating system can be seen in Figure 1.

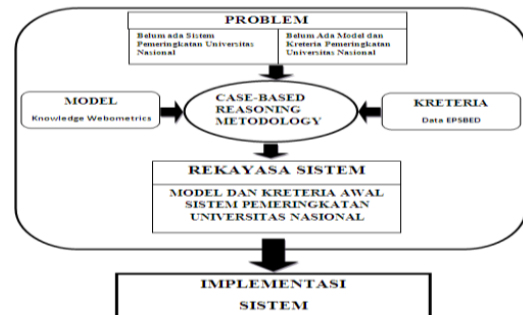


Figure 1. The research approach

This study uses data that is used as a model Webometrics ranking, extracted and stored in the Case-Based. So that the CBR stages rating system can be seen in Figure 2.

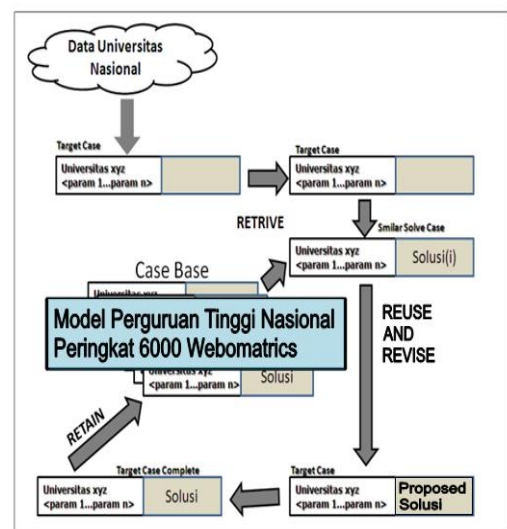


Figure 2. Stages of CBR with Case-Base Model

The final results of this study is a graph of the difference between the Webometrics ranking with the ranking system output with the CBR method.

Analysis data analysis and measurement of research output analysis method, using a comparative technique to analyze the pretest and posttest graphs the difference between before and after the system is applied to the system applied. Before the system is implemented, data ranking universities that go on Webometrics ranking at the input to the system as general knowledge, which referred as a data model. Data model is a list of national universities into the ranks Webometrics month of July 2009 which appears in Figure 3, the X axis is the name of High school and Y axis are ranked on Webometrics.

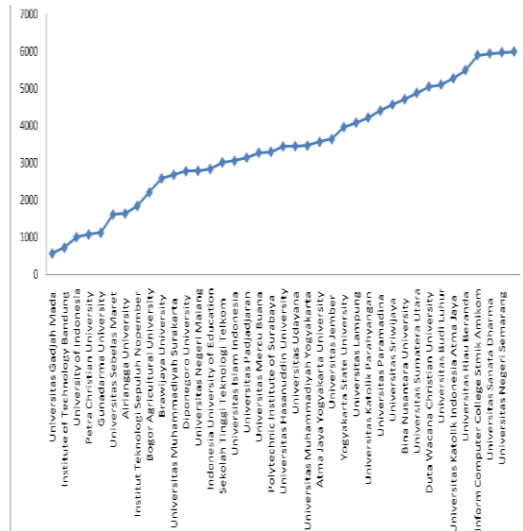


Figure 3. Graph Model Implementation

To retrieve the data to evaluate the system uses the nearest-neighbor algorithm. According to Greg Grudic [4], nearest-neighbor algorithm is an algorithm used to calculate the distance matrix between the input data based on any attribute similarity data. This algorithm has the advantage, among others, implement and simple to use and comprehensive and easy to use for prediction. Formula used:

$$\text{Sim} = D(X, Y)$$

Where:

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

X = data evaluation

Y = the sample data

x_i = the i th parameter of evaluation data

y_i = i -th parameter of evaluation data

n = number of parameters

After the data obtained, to calculate weights based on the similarity of the formula:

$$\text{IF } \sum_{i=1}^n x_i \leq \sum_{i=1}^n y_i, \text{Sim} + 1 \quad (2)$$

$$W_i = \text{Sim} * X_{\text{sol}} \quad (3)$$

X_{sol} = the solution of the data evaluation

w_i = weight of i -th solution

Both the Universities Indonesia who entered the Webometrics ranking or university which is used as the test

sample, is a university that has a data EPSBED His studies for each program. That to describe a working model in this study retrieval process can be described as follows:

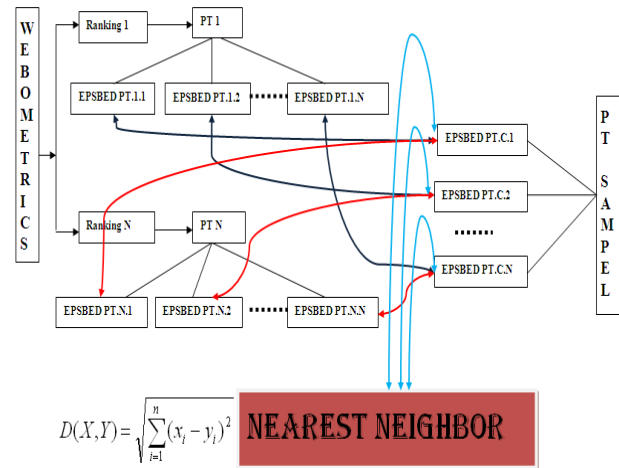


Figure 5. Retrieval process

Pseudo-code implementation of the above algorithm is as follows:

- Pseudo-code : Similarity Test
Narrative : Computing similarity using k-nearest neighbor algorithm
Input : Table CaseBased and Table CurrentCase
Output : Table Retrieve
Method : ujiSimilaritas ()
1. CREATE Object $cCase$ as an instance of using the ArrayList class CurrentCase type;
 2. CREATE Object $cBase$ as an instance of using the ArrayList class CaseBased type;
 3. CREATE Object $cCaseManager$ as an instance of class CurrentCaseManager;
 4. CREATE Object $cBaseManager$ as an instance of class CaseBaseManager;
 5. Comparative Variable INITIAL = 0, $x = 0$, $y = 0$;
 6. $cCase = cCaseManager.getCurrentCase ()$;
 7. $cBase = cBaseManager.getCaseBase ()$;
 8. WHILE $Y \leq cCase.size ()$
CurrentCase CCN = $cCase.get (Y)$;
 9. WHILE $X \leq cBase.size ()$ CaseBased CBN = $cBase.get (X)$;
 10. IF $(cBN.skorAkreditascCN.skorAkreditas) + (cCN.jlPenelitianancBN.jlPenelitian) + (-cCN.rtIPScBN.rtIPS) + (-cCN.rtSKScBN.rtSKS) + (-cCN.rsMHS cBN.rsMHS) + (cBN.rtMHSBaru - cCN.rtMHSBaru) + (cCN.S1cBN.S1) + (cCN.S2cBN.S2) + (cCN.S3cBN.S3) + (-cCN.prof cBN.prof) \leq 0$ then = 1 ELSE Comparative Comparative = 0;
 11. Similarity = $SQRT((cBN.skorAkreditascCN.skorAkreditas)^2 + (cBN.jlPenelitianancBN.jlPenelitian)^2 + (cBN.rtIPScCN.rtIPS)^2 + (cBN.rtSKScCN.rtSKS)^2 + (cBN.rsMHS cCN.rsMHS)^2 + (cBN.rtMHSBaru - cCN.rtMHSBaru)^2)$

$SBarucCN.rtMHSBaru)^2+(cBN.S1-cCN.S1)^2+(cBN.S2cCN.S2)^2+(cBN.S3cCN.S3)^2+(cBN.profcCN.prof)^2$

11. INSERT INTO table result (cBN.id, cCN.KdUniv, similarity, contrast)
12. INCREMENT X
13. END WHILE X
14. INCREMENT Y
15. END WHILE Y
16. END Method;

III. RESULTS AND DISCUSSION

A. Reserch Model

The data sample used in testing the CBR-based ranking method uses data that enter the National Higher Education in the Webometrics ranking and published in January 2010. While the selected study program is the Management Studies Program has data extracted from the site EPSBED Directorate General of Higher Education (DIKTI) is <http://dikti.evaluasi.go.id>, with EPSBED parameters as shown in step 11 in the pseudo-code similariats test.

National Higher Education Data entry Webometrics ranking contained in Table 1, sorted again only at the universities have a Program Management Studies.

Table 1. Sample Data Webometrics

UNIVERSITY	SIZE	VISIBILITY	RIC H FILE S	SCHOLAR	RANK
University of Gadjah Mada	602	421	1,028	827	562
Institute of Technology Bandung	564	657	1,138	654	661
University of Indonesia	903	1,007	741	981	815
Petra Christian University	1,137	1,794	964	59	854
Gunadarma University	883	1,590	1,106	613	1025
University of Negeri Malang*	3,545	1,873	1,544	107	1256
Institute Teknologi Sepuluh Nopember	1,289	885	2,716	2,094	1315
Universitas Sebelas Maret	1,386	1,224	3,070	2,479	1585
Airlangga University	1,287	2,863	3,767	231	1628
Brawijaya University	2,671	1,825	3,542	1,975	2026
Diponegoro University	2,353	3,209	4,343	426	2059
Bogor Agricultural University	3,053	1,912	4,989	1,569	2162
Universitas Padjadjaran	2,902	2,423	3,239	1,880	2236
Indonesia University of	1,436	1,763	4,117	4,423	2298

Education*					
Universitas Sriwijaya	2,352	4,644	4,631	193	2337
Universitas Islam Indonesia	3,026	3,764	4,965	470	2422
Universitas Muhammadiyah Surakarta	3,839	5,423	2,829	185	2471
Informatics and Computer College Stmik Amikom	4,537	3,492	5,928	472	2616

Higher Education in the name of the * mark in Table 1 is a sign of Higher Education has more than one webdomain. In the National College of Management Studies Program and has entered Webometrics rankings are listed in Table 2.

B. Rank Name of University Ranked

Whereas after the implementation of Case-Based Reasoning methods in The ranking system to measure the proximity (similarity) Higher Education (sample data) to the data output system model is obtained as in figure 6.

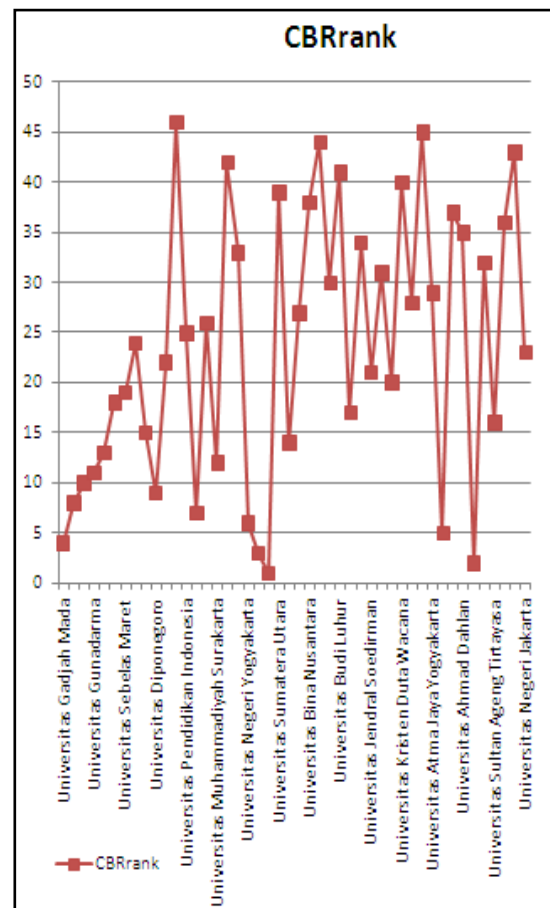


Figure 6. Graph Case-Based Reasoning rating

CONCLUSION

Comparison of the data above using Microsoft Excel, by combining the output above, the data are as follows:

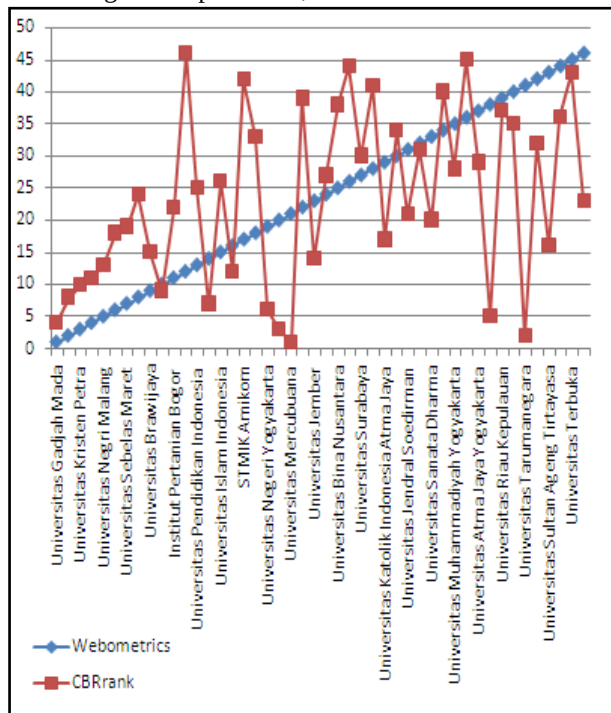


Figure 7. Comparison of results and graphs Webometrics CBR rating

From the data in Figure 7, it is clear that significant differences in output CBR method, although the method is to use machine learning methods to model, but the application will have its own characteristics. Comparative ranking of the above anomalies can be measured at the universities have the highest rank of each rating system. College is located at 572 Webometrics rankings are the University of Gadjah Mada. While the ranking CBR, Management Studies Program University of Mercubuana that was ranked highest as shown in table 3.

Of the value of W_i (weight rating) Program of Management Studies University Mercubuana have a higher weight rating of Management Studies, Gadjah Mada University. Learning algorithm is applied to the revised classification models to calculate the weight (score) to the data by testing encludian evaluation of each university.

Table 3. Comparison Table Parameters between the University of Gajah Mada University and the University Mercubuana

Perguruan Tinggi	SA	PS2	PS1	PS3	RMHS	MhsP	S1	S2	S3	Prof	MhsU
Universitas Mercubuana	1	5	14	0	38	2510	6	40	4	1	17687
Universitas Gadjah Mada	1	71	70	42	1	18	1	15	1	0	3276
Perguruan Tinggi	$\sum_{i=1}^n X_i$		$IF \sum_{i=1}^n x_i \leq \sum_{i=1}^n y_i, Sim+1$								W_i
Gadjah Mada	20306		Sim+1								293.233
Mercubuana	3496		Sim								572.000

Based on the results of the study, which starts from the needs analysis phase, design to testing with mengimplemtasikan Ranking System Methodology Case-Based Reasoning it can be concluded:

- Intelligent systems that use data and data EPSBED Webometrics ranking may rank the universities in Indonesia
- Cased-Based Reasoning methodology can be used as a model and data EPSBED ranking as the criteria to rank universities
- To improve the accuracy pemeringkatan CBR, further research should be done.

ACKNOWLEDGMENT

We would like to thank Coordinator of Research Development & Community Services in Computer Science Faculty, Universitas Bandar Lampung and all my friends.

REFERENCES

- Wahono, R. S. (2007). Webometrics Ranking of university-style techniques. RomiSatriaWahono.Net
- Aamolt, E. P. (1994). A. Aamodt, E. Plaza (1994): Case-Based Reasoning: foundational Issues, Methodological Variations, and System Approches. Catalonia, Spain.: AI Communications. IOS Press
- Siu, S. K. (2004). Fondations Of Soft Case-Based Reasoning. Canada: Published by John Wiley & Sons, Inc., Hoboken, New Jersey
- Grudic, G. (2004). Nearest Neighbour Learning. Borrowed Notes from Thomas G.Dietterich and Tom Mitchell