International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 9, Issue 1, No 1, 2021. ISSN: 2348-6600
Reference ID: IJCS-378 PAGE NO:2580-2585

Prototype of University Directory Application Using K-Means Clustering and TOPSIS

Ariyani Wardhana[#1], Wawan Kurniawan[*2]

[#1] *information system* [#2] *Information technology,*

[#]*Universitas Mercu Buana,*

Jakarta, Indonesia

[1]ariyani.wardhana@mercubuana.ac.id

[*]*Universitas Satya Negara Indonesia*

Jakarta, Indonesia

[2]wawan.krn75@gmail.com

Abstract— **The development of education and educational institutions in Indonesia has led to the emergence of many new higher education institutions and there have been many changes in majors and specializations offered from each higher education institution. Many higher education institutions also expand their campus network to various regions and even establish cooperative relationships with higher education institutions at home and abroad as partners. Prospective students who are looking for higher education institutions to continue their education often consider aspects of the majors offered, accreditation, location, tuition fees, facilities and infrastructure offered by each higher education institution, as well as various other supporting information.**

***Index Terms*—Directory, K Means, Clustering, document, university** *(key words)*

I. INTRODUCTION *(HEADING 1)*

The development and progress of the world of education in Indonesia is very rapid both in terms of the number of institutions and the diversity of study programs. Information about a large number of universities requires good handling so that information obtained by users is more optimal and useful. Utilization of information technology, especially by utilizing a web portal is a very good alternative used in the world of education to distribute information to high school / vocational high school graduates who are looking for information to continue their education to a higher level. From the PT statistics, there are 83 State Universities (PTN) organized by the Ministry of National Education (Depdiknas) spread across 30 provinces, and 2,598 Private Universities (PTS) organized by the public which are divided into 12 Kopertis areas, with so many of them. state and private tertiary institutions in Indonesia and scattered in many regions allow prospective students to continue their studies wherever they wish, the obstacle that often occurs is the lack of information obtained by them related to the destination university (Aditya Ramadhan & Desi Ramayanti, 2019), also with information where they will stay, how much should be prepared to rent a place to stay, where food is delicious and cheap, a cheap photocopy place near campus and other matters related to the need for teaching and learning process that they will live. With the existing problems like this it is very possible to build an application that can contain all the information needed by those who are just about to continue their studies far from their respective homes. The application that is built will provide all information relating to supporting the study of college students, so if new students try to find information relating to rent / boarding houses that are cheap, inexpensive snacks, a photocopy place near the campus can be provided by this application.

II. RELATED WORK

2.1 Information Systems

Information Systems consist of 2 words, namely Systems and Information. The system means a collection of elements (hardware, software, brainware) that are interconnected and interact and work together to achieve certain goals while Information means the results of data processing in a form that is more useful and more meaningful for the recipient who describes a real event used for making a decision. Based on the two meanings above it can be concluded that the Information System is a system consisting of input systems,

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 9, Issue 1, No 1, 2021. ISSN: 2348-6600
Reference ID: IJCS-378 PAGE NO:2580-2585

processes and outputs that are interconnected and work together where the output shows and displays a result of the process in the form of information that is useful and needed by the user. (Yaya Sudarya Triana, Fryda Farizha Adrianti, Fitri Anggri Maharani, 2019).

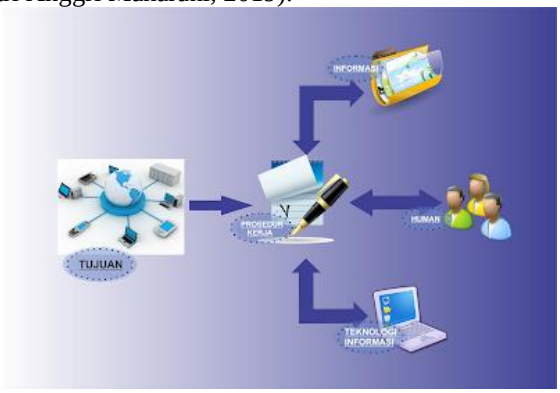

Fig. 1 Flow of Information Systems

With a simpler sentence, Information System is a system that consists of inputs, processes and outputs which will produce information. "Information systems are very much needed in human daily life. Information systems become one of human needs in order to be able to interact with external situations that are always changing."

2.2 Directory

Directory according to the ALA Glossary of Library and Information System is a collection of references that contain names or organizations that are arranged systematically, usually in alphabetical order or group, completed with addresses, activities and other data. Contains a list of names of people or institutions or organizations by providing information in the form of addresses, biographies or brief history and their activities. reluctant to have a directory, each file can be grouped according to its definition, then the directory is used to find information about:

- Address or telephone number, about a person or company or agency.
- The full name of a person, company or organization or agency.
- A description of the agency or certain factory products or services of a particular service bureau.
- Information about who is the head of an agency, director of a company, chancellor of a university, principal and so on at this time or in a certain period.

The directory / manual can be divided into several groups as follows:

1. Local user manuals, such as telephone books, city guides and so on
2. Guidance related to the government, for example instructions on post offices, police stations and other government agencies. The user guide in this group often also contains information about international bodies.
3. A manual containing information about agency agencies, for example schools, foundations, libraries, hospitals, museums and other similar organizations.
4. Guidebooks about a profession, such as legal experts, library experts, doctors, and so on
5. A manual that contains information about trade and industry, for example factories, companies, service bureaus, and others.

Example: Jakarta PPPI Directory 2002-2003, and Indonesian ASKARINDO Association Directory

2.3 Algoorithmic K-Means Clustering

K-Means Clustering is a method of analyzing data or the method of Data Mining that performs the modeling process without supervision (unsupervised) and is one of the methods that group data by partitioning the system. There are two types of data clustering that are often used in the process of grouping data, namely Hierarchical and Non-Hierarchical, and K-Means is a non-hierarchical or Partitional Clustering data clustering method.

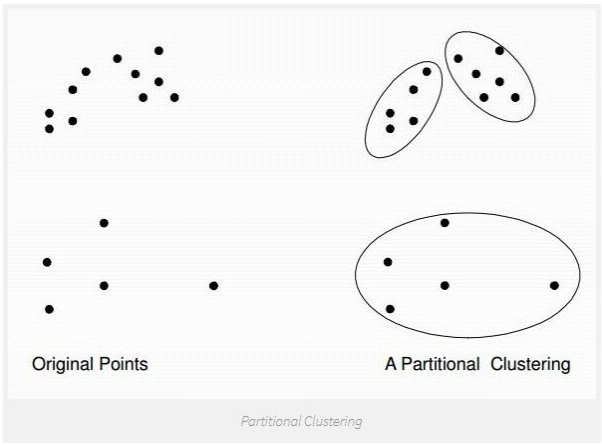

Fig. 2 Partional Clustering

The K-Means Clustering method seeks to group existing data into several groups, where data in one group has the same characteristics with each other and has different characteristics from the data in other groups.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 9, Issue 1, No 1, 2021. ISSN: 2348-6600
Reference ID: IJCS-378 PAGE NO:2580-2585

2.4 TOPSIS

Based on the concept where the best chosen alternative not only has the shortest distance from the positive ideal solution, but also has the longest distance from the negative ideal solution (Kusumadewi, 2006: 87). This concept is widely used in several MADM models to solve practical decision problems. This is because the concept is simple and easy to understand, efficient computation, and has the ability to measure the relative performance of decision alternatives in a simple mathematical form. In general, the TOPSIS procedure follows the steps as follows:

1. Make a normalized decision matrix
2. Make a weighted normalized decision matrix
3. Determine a positive ideal solution matrix and a negative ideal solution matrix
4. Determine the distance between the values of each alternative with the positive ideal solution matrix and the negative ideal solution matrix;
5. Determine the preference value for each alternative

III. METHODOLOGY

The research framework used in the "university directory prototype" uses the Research and Development (R&D) method. According to Sudaryono Research and Development (R&D) is a research method used to produce a product and test its effectiveness (Sudaryono, 2015). The product produced is a university directory application for all tertiary and private universities. Therefore, to get the appropriate product, in the application development the researcher uses the Rational Unified Process (RUP) development model. The reason for choosing the RUP development model is that the RUP uses an iterative and incremental process so that it can accommodate changes in software requirements (Rosa, 2011).

3.1 Development Procedure

Rational Unified Process has 4 stages, namely: Inception, Elaboration, construction, and Transition. Following is an explanation of each stage.

a. Inception

Inception phase is more on modeling business processes (business modeling) and defining the needs of the system that is made (requirements). Here are the steps needed at the inception stage:

- Understand the scope, needs, costs, and time.
- Understand existing business models in the project scope.
- Build a business model.

The expected outcome at this stage is to fulfill the Objective Milestone Lifecycle with the following criteria:

- Feedback on definition of scope, estimated schedule, and estimated cost.
- Needs are in line with primary cases.
- The scope of the prototype will be developed.

b. Elaboration

This stage is more focused on system architecture design planning. This stage is also more about the analysis and design of the risks that may occur from the architecture created. The activities carried out at this stage are:

- Description of the software architecture developed from needs analysis.
- Making use of the system with identified actors.
- Architectural design developed from usecase design.
- Prototype with digital mockup to reduce technical risks.
- Unified Modeling Language (UML) design, interface layout, and icons.

If at the end of this stage the desired target is not achieved then it can be canceled or repeated

C. Construction

This stage focuses on developing components and features used in building applications. This stage is more on the implementation and testing of program code. The activities carried out at this stage are:

- Implementation of user interface according to material design
- Implementation using Android studio tools
- Testing functionality
- Early stage deployment

D. Transition

This stage focuses on the deployment stage or the installation and publication of applications in the playstore. This stage produces a software product which becomes the Initial Operational Capability Milestone. At this stage also carry out training, maintenance, and testing with the user.

3.2 Research Flow

The plot of determining the location of a university or college uses K-Means Clustering and TOPSIS as shown below :

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 9, Issue 1, No 1, 2021. ISSN: 2348-6600
Reference ID: IJCS-378 PAGE NO:2580-2585

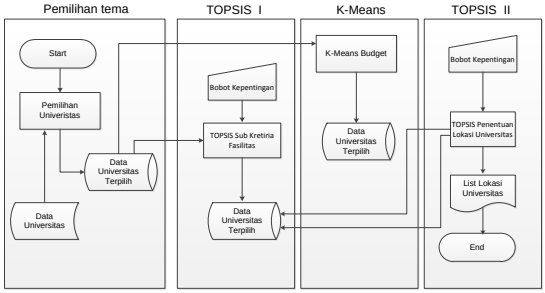

Fig. 3 Research Flow

In this study, using a system model namely K-Means and TOPSIS (Technique for Order Performance by Similarity to Ideal Solution). The application of the K-Means method will provide a value that will be reprocessed by TOPSIS. As an illustration of the course of the process from the selection of kretiria to information on tourism locations, From the data obtained, the first step is to determine the theme of the university location. After the theme is selected, the user is asked to determine the importance of the existing sub-criteria, the weight value is used to determine the value of the criteria. In this study the calculation results obtained from the calculation of phase I, so that the TOPSIS phase I is not displaying the ranking order of the location of the university. In the same process, the calculation of the budget value uses K-Means to determine the budget grouping. After getting the value of facilities and budget in the previous process, it is continued by calculating TOPSIS II to determine the order of recommendation of the location of the boarding house, but before the user is asked to determine the weight of importance between the budget and facilities. In this study the value of profit and cost is determined by the system, with budget as a factor of cost and facilities as profit. So the system will choose the smallest cost range and the facility selects the largest range.

IV. EVALUATION

From the existing design in accordance with the research method described in Figure 2.10, the flow of the system can be explained as follows:

1. Choosing a University Theme

In determining the theme of university location, selection is only done using ordinary queries based on the names entered by the facility.

2. Enter the sub-facility weight values

From the data obtained from the initial query (step 1) the subfac- ture data of each location is taken, then using the formula, the first stage of TOPSIS calculation is weighted with the importance value entered by the user. From this calculation the values obtained from the university location facility criteria. TOPSIS is only used to determine grades not to determine ranking of the University location.

3. Entering the value of the weight of the facilities and costs

There are two processes that occur when the system displays an input menu of budget weights and facilities. The first process is grouping the range of costs generated by the initial search (step 1) using the K-Mean Clustering method and it is determined that the choice always lies in cluster 1 assuming the cost choice is always the smallest. The next process is determining the level of importance between the bugs and the facilities whose value is obtained from the step 2 process.

4. Display the calculation results

Testing the recommendation system for coastal resorts around Malang that were built was divided into 3 types of testing, namely:

In general, the TOPSIS procedure follows the steps as follows:

1. Make a normalized decision matrix
2. Make a weighted normalized decision matrix
3. Determine a positive ideal solution matrix and a negative ideal solution matrix
4. Determine the distance between the values of each alternative with the positive ideal solution matrix and the negative ideal solution matrix;
5. Determine the preference value for each alternative.

TOPSIS requires a work rating for each alternative Ai on each normalized Cj criterion

With i = 1,2, ... m; and j = 1,2, ..., n; Where :

Rij = normalized matrix [i] [j]

Xij = decision matrix [i] [j]

The positive ideal solution A + and the negative ideal solution A- can be determined based on the normalized weight rating (yij) as:

Yij = wi.rij

With i = 1,2, ... m;

And j = 1,2, ... n;

A + = (y1 +, y2 +, ..., yn +)

A- = (y1-, y2-, ..., yn-)

Where :

Yij = weighted normalized matrix [i] [j]

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 9, Issue 1, No 1, 2021. ISSN: 2348-6600
Reference ID: IJCS-378 PAGE NO:2580-2585

Wi = weight vector [i] of the AHP process
Yj + = max yij, if j is the gain attribute and
Yj- = min yij, if j is the cost attribute
J = 1,2, ..., n

The distance between the alternative Ai and the positive ideal solution
I = 1,2, ... m
Where :
D + = Distance of alternative Ai with positive ideal solution
Yi + = positive ideal solution [i]
Yij = weighted normalization matrix [i] [j]

The distance between the alternative Ai, the negative ideal solution
Where :
Di = alternative distance Ai with an ideal negative solution
Yi = positive ideal solution [i]
Yij = weighted normalization matrix [i] [j]

The preference value for each alternative (Vi) can be seen in the formula below:
I = 1,2, ..., m
Where :
Vi = proximity of each alternative to the ideal solution
At + = distance alternative Ai with positive ideal solution
Di- = distance alternative Ai with negative ideal solution
A greater value of Vi indicates that alternative Ai is preferred.

4.1 Testing the Validity of the Algorithm

The algorithm validation test is performed to determine the level of compatibility between the outputs of the system and the outputs from manual calculations. The weight of criteria and alternative data used in testing can be seen in Table 1 and Table 2. The results of the algorithm validation test can be seen in Table 3.

Table 1. Weight of Criteria

No	Criteria	Status	Weight
1.	Price	COST	1
2.	Distance	COST	2
3.	Rating	Benefit	2
4.	Transportation	Benefit	1
5.	Amenities	Benefit	1

Table 2: Criteria Weight

No	Alternative Universities	P	D	R	T	F
1.	University of Indonesia	5000	63	4,3	3	6
2.	Gajah Mada University		62	4,3	3	4
3.	Bandung Institute Of Technology		63	4,3	2	3
4.	The Ten November Institute Of Technology		58	4,2	4	4
5.	Bogor Agricultural Institute		70	4,3	4	4

Table 3. Result of Algorithm Validation

No	Hasil Perhitungan		Status
	Sistem	Manual	
1.	University of Indonesia	University of Indonesia	Valid
2.	Gajah Mada University	Gajah Mada University	Valid
3.	Bandung Institute Of Technology	Bandung Institute Of Technology	Valid
4.	The Ten November Institute Of Technology	The Ten November Institute Of Technology	Valid
5.	Bogor Agricultural Institute	Bogor Agricultural Institute	Valid

V. Conclusions

K-Means algorithm can only be used for calculations in determining buget grouping, this is because of the three criteria (location, facilities and budget) that are the research variables, only the budget which is the continuous data or numbers. In the use of the K-Means and TOPSIS algorithms the criteria used are if the data does not continue then the conversion must be done first. The steps taken to display the university location search using the K-Means and TOPSIS

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 9, Issue 1, No 1, 2021. ISSN: 2348-6600
Reference ID: IJCS-378 PAGE NO:2580-2585

algorithm are as follows, The following are suggestions or input for the next research are as follow, The need for facilities that can be interactive from user input, It needs to be linked to the geographic information system to display a map of campus locations., Can display location search results in groups based on the budget expected by the user, which is obtained from the facility search list., Application examples are only prototypes, it would be better if the actual application was made in the form of a mobile application.

REFERENCES

[1] (t.thn.). Dipetik November 19, 2018, dari https://webstatsdomain.org: https://webstatsdomain.org/d/www.seamolec.org?smode=full

[2] Anwar, A. (2014). A Review of RUP (Rational Unified Process). *International Journal of Software Engineering (IJSE)*.

[3] Aditya Ramadhan & Desi Ramayanti (2019). Aplikasi Pencarian SMA/SMK Di Wilayah Jakarta Barat Berbasis Android, *Jurnal Saintekom*, 9 (2).

[4] Bonvillian, W. B., & Singer, S. R. (2013). The Online Challenge to Higher Education. *science and technology, 29* (4), 23-30.

[5] Melie-Garcia, L., Draganski, B., Ashburner, J., & Kherif, F. (2018). Multiple Linear Regression: Bayesian Inference for Distributed and Big Data in the Medical Informatics Platform of the Human Brain Project. *bioRxiv* .

[6] Preacher, K. J., Preacher, K. J., & Bauer, D. J. (2006). Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling,in Multiple Linear Regression, Multilevel Modeling,in Multiple Linear Regression, Multilevel

[7] Pressman, R. (2012). *Software Engineering: A Practitioner's Approach, Seventh Edition.* (A. Nugroho, G. Nikijuluw, T. Rochadiani, & I. Wijaya, Trans.) Yogyakarta: Penerbit Andi

[8] Prasetyo, E. 2014. Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta. Penerbit ANDI Yogyakarta, 218.

[9] Sein-Echaluce, M. L., Fidalgo-Blanco, A., & Alves, G. (2017). Technology behaviors in education innovation. *computers in human behavior* , 596-598.

[10] Shafagatova, V. L. (2016). Business process performance. *Springer Plus* , 1-24.

[11] Sri Kusumadewi, Sri Hartati, Agus Harjoko, Retantyo Wardoyo, Fuzzy Multi-Atribut Decision Making

[12] Yaya Sudarya Triana, Fryda Farizha Adrianti, Fitri Anggri Maharani (2019), Jurnal RESTI (Rekayasa Sistem dan Tekhnologi Informasi), 163-169.