

STUDENT CLUSTER ANALYSIS AS AN EFFORT TO OPTIMIZE CAMPUS PROMOTION

Romy Aulia¹, Agus Nur Khomarudin^{1*}, Indra Laksmna¹, Jamaluddin¹, Rina Novita¹

¹Teknologi Rekayasa Komputer, Politeknik Pertanian Negeri Payakumbuh

email: *agusnurkhumarudin@gmail.com

Abstract: This research tries to describe student cluster analysis, as an effort to optimize campus promotion to various schools and regions. It is known that every year, Politeknik Pertanian Negeri Payakumbuh, abbreviated as PPNP, brings in students from various regions in Indonesia. Regarding the campus promotion strategy process, the PPNP promotion section has not been based or referred to the results of processing existing student data. So that the budget used by the campus promotion team has not been right on target with the results of students who can be brought to campus. In addition, the existing student database has not been processed or explored further, so that it has not produced knowledge that is very useful as material to support the decisions of the academic and student affairs department and the campus promotion team. The method used in this research is CRISP-DM which stands for Cross- Industry Standard Process for Data Mining. Based on the characteristics of each cluster, the PPNP Promotion Team in conducting the next socialization is advised to prioritize provinces such as West Sumatra and North Sumatra. Currently, managerial circles in this context, university leaders are expected to be able to make data-based decisions. Data-based decision making can foster a culture of sustainable innovation, produce customer-centric offerings and drive long-term business growth.

Keywords: cluster analysis; student data; k-means clustering; campus promotion

Abstrak: Penelitian ini mencoba untuk mendeskripsikan analisis cluster mahasiswa, sebagai upaya optimalisasi dalam melakukan promosi kampus ke berbagai sekolah dan daerah. Diketahui bahwa setiap tahunnya, Politeknik Pertanian Negeri Payakumbuh disingkat PPNP mendatangkan mahasiswa dari berbagai daerah di Indonesia. Terkait dengan proses strategi promosi kampus, bagian promosi PPNP belum didasarkan pada hasil pengolahan data mahasiswa yang ada. Sehingga anggaran yang digunakan tim promosi belum tepat sasaran dengan hasil mahasiswa yang dapat didatangkan ke kampus. Selain itu database mahasiswa yang ada selama ini belum diolah atau digali secara jauh, sehingga belum menghasilkan pengetahuan yang bermanfaat sebagai bahan untuk mendukung keputusan bagian akademik dan kemahasiswaan serta tim promosi kampus. Metode yang digunakan dalam penelitian ini yaitu CRISP-DM merupakan singkatan dari Cross-Industry Standart Process for Data Mining. Berdasarkan karakteristik setiap cluster, maka untuk Tim Promosi PPNP dalam melakukan sosialisasi berikutnya disarankan memprioritaskan pada provinsi seperti Sumatera Barat dan Sumatera Utara. Saat ini kalangan manajerial yaitu pimpinan perguruan tinggi diharapkan dapat melakukan pengambilan keputusan berbasis pada data. Pengambilan keputusan berbasis data dapat menumbuhkan budaya inovasi yang berkelanjutan, menghasilkan penawaran yang berpusat pada pelanggan dan mendorong pertumbuhan bisnis jangka panjang.

Kata kunci: analisis cluster; data mahasiswa; k-means clustering, promosi kampus

INTRODUCTION

The development of information technology, which is increasingly sophisticated today, has produced a lot of data. The increase in data will raise a big question mark, namely “what value can we get from the data warehouse?”. This research examines a solution to the big question mark, namely by applying a database technology that is often referred to as data mining [1]. Data mining can be applied to extract added value from a data set in the form of knowledge that has not been known manually [2]. In this study, data processing of Payakumbuh State Agricultural Polytechnic students will be carried out, with attributes including: NIM, student name, Department, Study Program, regional origin, and school origin data from 2021 to 2023 batch student data. The results of this data processing aim to assist and provide recommendations to the academic and student affairs and campus promotion teams in determining marketing or campus promotion strategies to SMA / MA / SMK / equivalent in various regions.

Every year, Payakumbuh State Agricultural Polytechnic brings in students from various regions in Indonesia, such as from: Aceh, North Sumatra, Nias, Kep. Mentawai, Riau, Jambi, Bengkulu, Papua and eastern Indonesia (through the Afrimasi program), and West Sumatra in particular. Because Payakumbuh State Agricultural Polytechnic students come from various regions, a special strategy is needed by the academic and student affairs department, as well as the promotion team to find prospective students so that the promotion carried out becomes more effective and efficient.

During the preliminary research

process, it was found that the campus promotion strategy process had not been based or referred to the results of processing existing student data. So that the budget used by the campus promotion team has not been right on target with the results of students who can be brought to campus. In addition, related to the database, existing student data has not been processed or explored further, so that it has not produced knowledge that is very useful as material to support the decisions of the academic and student affairs department and the campus promotion team.

Campus promotion strategies to be more effective and efficient are certainly carried out by considering various existing student data. Existing student data needs to be processed to obtain knowledge, patterns or data clusters so as to produce recommendations in decision making. In processing data in the database, there are several techniques used in data mining, one of the data mining techniques is clustering. There are two types of clustering methods used in clustering data, namely hierarchical clustering and non-hierarchical clustering [3]. K-means clustering as a non-hierarchical data clustering method partitions existing data into one or more clusters or groups. The resulting groups or clusters are knowledge/information that is useful for policy users in the decision-making process [4].

Some previous studies related to the application of clustering techniques in supporting decision making are research by: 1) Triyansyah, et al in 2018 on “Data Mining Analysis Using K-Means Clustering Algorithm to Determine Marketing Strategy” [5]; 2) Rosmini, et al 2018 on “Implementation of the K-Means Method in Mapping Student Groups Through

Lecture Activity Data” [6]; 3) Asroni, et al 2018 on “Application of the Clustering Method with the K-Means Algorithm on Grouping Prospective New Student Data at Muhammadiyah University of Yogyakarta (Case Study: Faculty of Medicine and Health Sciences, and Faculty of Social and Political Sciences)” [5]; 4) Novita, et al in 2023 on “Application of the K-Means Algorithm and Its Analysis to Determine Student Study Completion Strategy Policies” [7]. Based on several previous studies, it shows that every decision-making process must be based on data. This culture is an effort by companies or educational institutions to achieve competitive advantage by optimizing resources, minimizing risks, and quickly adapting to market changes [8]. In addition, Data-Based Decision Making can foster a culture of continuous innovation, produce customer-centric offerings and drive long-term business growth [9].

The results of data processing in the form of grouping or clustering data, are expected to help provide recommendations to the academic and student affairs department and the promotion team in the process of making decisions on campus promotion strategies to various schools and regions.

METHOD

Data Collection Methods

Data Source, The types of data sources obtained by researchers are primary data and secondary data. Primary data in research researchers obtained from direct interviews with sub-section heads, student affairs academic staff and several promotional teams of the Payakumbuh State Agricultural Polytechnic. Secondary data in this study are categorized into 2 categories,

including the category of student data for the class of 2021 to 2023 and data on the identity of students of the Payakumbuh State Agricultural Polytechnic. Each consists of NIM data, student names, majors, study programs, regional origins, and data on school origins.

Data Collection, The data collection method in this study was carried out in several ways including: 1) Interviews, Interviews were conducted with sub-section heads, student affairs academic staff and several promotional teams of the Payakumbuh State Agricultural Polytechnic with the aim of obtaining an overview and concept of campus promotion carried out so far. 2) Literature Study, Literature study is carried out with the aim of obtaining references related to the research being conducted. Literature studies in this study are based on journals, books, the internet, and so on. 3) Survey, The research survey was carried out by asking permission to collect data regarding data on students from batch 2021 to 2023 as well as data on the identity of Payakumbuh State Agricultural Polytechnic students, including: NIM, student name, Department, Study Program, regional origin, and school origin data.

Data Type, The type of data used in solving the problems in this study is quantitative data. Quantitative data in this study include: NIM, student name, Department, Study Program, regional origin, and school origin data from 2021 to 2023 batch student data.

Data Analysis Method, The data mining development methodology used in this research is CRISP-DM. There are six sequential stages in CRIPS-DM which are described as follows [5]:

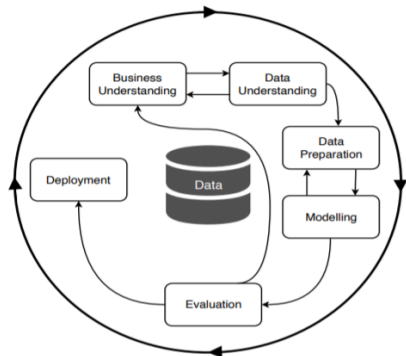


Figure 1. Stages of CRIPS-DM

The stages of CRIPS-DM development are: Business Understanding, This first stage begins with understanding the problems contained in the marketing strategy or campus promotion. The purpose of this research is to explore knowledge in the form of cluster result analysis with each study program pattern taken and school origin and region of origin, where this knowledge is hidden from the pile or student database.

Data Understanding, Data understanding is done by getting a good understanding of the data needs needed in solving the problems in this study. The data needs in this study include: NIM, student name, Department, Study Program, regional origin, and school origin data from 2021 to 2023 batch student data.

Data Preparation, at the data preparation stage the author processes the data obtained using several stages contained in KDD. KDD stands for Knowledge Discovery in Database. Some of the KDD stages used in this research are data cleaning, data integration, data selection, and data transformation [6].

Modeling, The modeling phase in this research directly involves data mining techniques. The steps we take in this phase are first to select and apply the appropriate modeling technique, then titrate the model rules to optimize the results. Note that multiple techniques may be

used for the same data mining problem. And if needed, the process can go back to the data processing phase into a form that matches the specific needs of the particular data mining technique.

Evaluation, In this evaluation phase, we evaluate one or more models used in the modeling phase for quality and effectiveness before they are deployed for use, then determine whether there are models that meet the objectives of the initial phase, then determine whether there are important business or research issues that are not properly addressed, and finally make decisions regarding the use of the results of data mining.

Deployment, At this stage the author tries to present the results of the application of the data mining clustering method with the K-Means algorithm that has been carried out. The stages in the K-Means Clustering algorithm are: 1). Select the number of clusters k ; 2). Initialize k cluster centers can be done with random numbers; 3). Determine the Euclidean distance with the formula:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

formula description:

$D(i,j)$ = Distance of data i to the center of cluster j ; X_{ki} = The i -th data on the k -th data attribute; X_{kj} = The j -th center point on the k -th attribute.

4). Recalculate the cluster center with the current cluster membership; 5). Reassign each object using the new cluster center.

The types of data sources obtained by researchers are primary data and secondary data. Primary data in research researchers obtained from direct interviews, secondary data in this study are categorized into 2 categories, including the category of student data for the class of 2021 to 2023 and the self-identity data of Payakumbuh State Agri-

cultural Polytechnic students.

RESULTS AND DISCUSSION

Research Results

Business Understanding, The first stage begins with understanding the problems contained in the marketing strategy or campus promotion. The purpose of this research is to explore knowledge in the form of analyzing cluster results with each study program pattern taken and school origin and region of origin, where this knowledge is hidden from the pile or student database. The results obtained are the data needs of Payakumbuh State Agricultural Polytechnic students, the need for applications / software that will be used in data processing.

Data understanding, at this stage is done by getting a good understanding of the data needs needed in solving the problems in this study. The data requirements in this study include: NIM, student name, Study Program, origin of city / regency and province, as well as data on the origin of schools from student data from batch 2021 to 2023. The results obtained are in the form of a list of biodata of Payakumbuh State Agricultural Polytechnic students from 2021 to 2023 in Excel format.

Data Preparation, At this stage, researchers process the data obtained using several stages contained in KDD. KDD stands for Knowledge Discovery in Database. Some of the KDD stages used in this research are:

Data cleaning, incomplete student data is refined and also cleaned from abnormal forms. The focus of data refinement and cleaning is carried out on the NIM, gender, study program, city / regency origin, provincial origin and

school origin data columns.

Data integration, All student data needed has been integrated, the data obtained comes from the SIAKAD application of the Payakumbuh State Agricultural Polytechnic.

Data selection, student data obtained in excel form and has gone through a data cleaning process. Some of the data columns used are NIM data, gender, study program, city / regency origin, provincial origin and school origin, for other data, it can be discarded because it is not needed.

Data transformation, data on NIM, gender, study program, city / regency origin, provincial origin and school origin are then transformed so that they can be processed using the WEKA application / software.

Modeling, This phase directly involves data mining techniques. The steps we take in this phase are first to select and apply the appropriate modeling technique, then calibrate the model rules to optimize the results. Note that multiple techniques may be used for the same data mining problem. And if necessary, the process can return to the data processing phase into a form that matches the specifications of the needs of a particular data mining technique.

The data mining technique used for data processing in this research is clustering using the k-means clustering algorithm. In this study, data processing of Payakumbuh State Agricultural Polytechnic from student data from class 2021 to 2023.

The data mining technique applied in this software is k-means clustering. Visualization of the data processing process with the k-means cluster algorithm is shown in Figure 2.

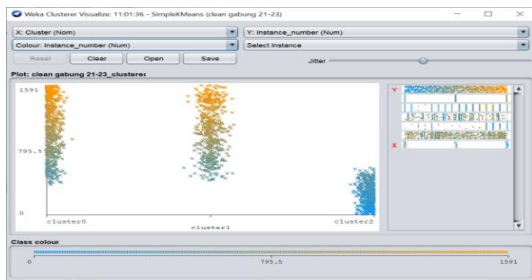


Figure 2. Cluster visualization view with WEKA software

Based on the cluster results that have been obtained, then the cluster result data is withdrawn. In this study, 3 clusters were obtained with the amount of data for each cluster, namely: cluster 1 = 557 (35%), cluster 2 = 558 (35%), cluster 3 = 477 (30%). The cluster result data obtained will then be visualized in various types of tables to make it easier to study the characteristics of each cluster. The following is a detailed description of cluster 1 - cluster 3:

Characteristics in Cluster 1: Students mostly take study programs: TPTP, TRK, TPT, TP, Tekben, TRP. Students mostly come from the Province: West Sumatra: 78% and North Sumatra: 14%. Students mostly come from Lima Puluh Kota Regency and Payakumbuh City. School origin: SMKN 3 Payakumbuh, SMAN 1 Harau, SMAN 1 Padang Gelugur, SMKN 2 Kec. Guguak.

Characteristics of Cluster 1

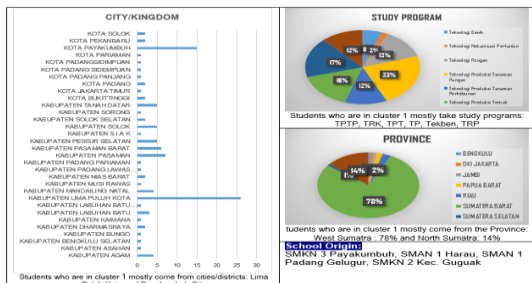


Figure 3. Display of data characteristics in cluster 1

Characteristics in Cluster 2: Students mostly take study programs: PP, TPTP, TPT, Pavet, TMP, TRK. Most

students come from the provinces: West Sumatra: 67%, North Sumatra: 21% and Riau: 8%. Students mostly come from Lima Puluh Kota district, Payakumbuh city, Mandailing Natal district. School origin: SMKN PP Padang Mengatas, SMAN 1 Harau, SMAN 5 Bagan Sinembah, SMKN 2 Payakumbuh, SMKN Pertanian Pembangunan 1 Kualuh Selatan and SMAN 1 Ranah Batahan.

Characteristics of Cluster 2

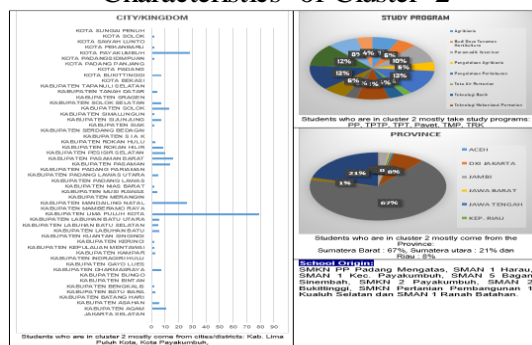


Figure 4. Display of data characteristics in cluster 2

Characteristics of Cluster 3

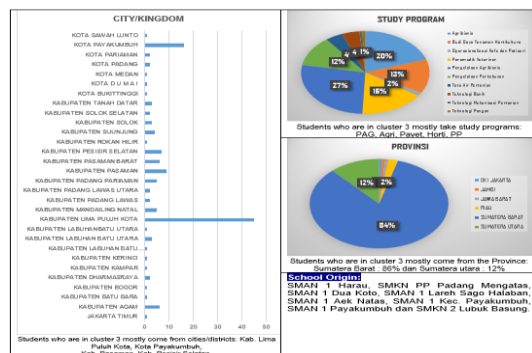


Figure 5. Display of data characteristics in cluster 3

Characteristics in Cluster 3: Students mostly take study programs: PAG, Agri, Pavet, Horti, PP. Students mostly come from the Province: West Sumatra: 86% and North Sumatra: 12%. Students mostly come from Lima Puluh Kota district, Payakumbuh city, Pasaman district, Pesisir Selatan district. School origin: SMAN 1 Harau, SMKN PP

Padang Mengatas, SMAN 1 Dua Koto, SMAN 1 Lareh Sago Halaban, SMAN 1 Aek Natas, SMAN 1 Kec. Payakumbuh, SMAN 1 Payakumbuh and SMKN 2 Lubuk Basung.

Discussion

Data processing in this study uses the WEKA application with a clustering algorithm, namely k-means clustering. The cluster results in this study are 3 clusters, where each cluster has its own characteristics. The results of data processing in the form of grouping or clustering data, can be utilized as support or recommendations to the academic and student affairs department and the promotion team in the process of making decisions on campus promotion strategies to various schools and regions.

Along with the development of information technology and various data growth, managerial circles in this context, university leaders are expected to be able to make data-based decisions [7]. This culture is an effort by companies or educational institutions to gain a competitive advantage by optimizing resources, minimizing risks, and adapting quickly to market changes [8]. Additionally, data-driven decision making can foster a culture of continuous innovation, create customer-centric services, and drive long-term business growth [9].

In addition, Data-Driven Decision Making can foster a culture of continuous innovation, produce customer-centric offerings and drive long-term business growth [10]. This concept is in line with the purpose of data mining, which is part of KDD, which stands for Knowledge Discovery in Database, which is to extract interesting knowledge from a pile of databases. KDD is an organized process for identifying applicable, useful and understandable patterns from large and

complex data sets. Data mining is the core of the KDD process, which involves inferring algorithms that explore data, develop models and discover previously unknown patterns [11].

Based on the characteristics of each cluster, the PPNP Team in conducting the next socialization is advised to prioritize the following areas and schools: 1) West Sumatra Province, covering cities / regencies, namely: kab. Lima Puluh Kota, Payakumbuh City, West Pasaman district, Pasaman district, Solok district and Pesisir Selatan district. SMA/SMK/equivalent: SMAN 1 Harau, SMKN PP Padang Mengatas, SMAN 1 Dua Koto, SMAN 1 Lareh Sago Halaban, SMAN 1 Payakumbuh, SMAN 1 Padang Gelugur, SMKN 3 Payakumbuh, SMKN 2 Kec. Guguak. 2) North Sumatra Province, covering cities/districts namely: Mandailing Natal district, Labuhan Batu district. Labuhan Batu, kab. Labuhan Batu Utara, and kab. Labuhan Batu Selatan. SMA/SMK/equivalent: SMAN 1 Aek Natas, SMKN Pertanian Pembangunan 1 Kualuh Selatan.

CONCLUSION

This research tries to describe student cluster analysis, as an effort to optimize campus promotion to various schools and regions. Along with the development of information technology and various data growth, managerial circles in this context, university leaders are expected to be able to make data-based decisions. Data-Based Decision Making can foster a culture of sustainable innovation, produce customer-centric offerings and drive long-term business growth.

BIBLIOGRAPHY

- [1] Agus Nur Khomarudin, “Teknik Data Mining: Algoritma K-Means Clustering,” *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>
- [2] R. Budiman and R. Anto, “Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering),” *ProTekInfo (Pengembangan Ris. dan Obs. Tek. Inform.)*, vol. 6, no. 1, p. 6, 2019, doi: 10.30656/protekinfo.v6i1.1691.
- [3] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [4] W. Purba, S. Tamba, and J. Saragih, “The effect of mining data k-means clustering toward students profile model drop out potential,” *J. Phys. Conf. Ser.*, vol. 1007, no. 1, 2018, doi: 10.1088/17426596/1007/1/012049
- [5] C. P. Lopez, *DATA MINING. The CRISP-DM METHODOLOGY. The CLEM language and IBM SPSS MODELER*. Lulu Press, Inc, 2021.
- [6] A. Nur Khomarudin, S. Zakir, R. Novita, Endrawati, M. Zahiri Bin Awang Mat, and E. Maiyana, “K-Mean Clustering Algorithm in Grouping Prospective Scholarship Recipients,” *J. Phys. Conf. Ser.*, vol. 1779, no. 1, 2021, doi: 10.1088/1742-6596/1779/1/012007.
- [7] A. Rustandy *et al.*, “Empowering Business Growth: Harnessing Data-Driven Decision Making for Resource Optimization and Innovation in Agile Manufacturing Mendorong Pertumbuhan Bisnis: Memanfaatkan Pengambilan Keputusan Berbasis Data untuk Optimalisasi Sumber Daya dan Inovasi da,” *Manag. Stud. Entrep. J.*, vol. 4, no. 6, pp. 9459–9468, 2023, [Online]. Available: <http://journal.yrpiiku.com/index.php/msej>
- [8] K. Schildkamp, “Data-based decision-making for school improvement: Research insights and gaps,” *Educ. Res.*, vol. 61, no. 3, pp. 257–273, 2019, doi: 10.1080/00131881.2019.1625716.
- [9] I. K. H. Oktaviarosa, “Penggunaan Big Data Dalam Pengambilan Keputusan Kebijakan Publik,” *Triwikrama J. Ilmu Sos.*, vol. 3, no. 7, pp. 70–89, 2024.
- [10] P. Zagoto and E. Murniarti, “Analisis Kebijakan Pendidikan dalam Membangun Kerangka Pengambilan Keputusan Berbasis Data untuk Meningkatkan Mutu Pendidikan Berdasarkan Uu No . 20 Tahun 2003 Pasal 3,” *J. Pendidik. Tambusai*, vol. 8, no. 2, pp. 27668–27673, 2024.
- [11] R. Novita, A. N. Khomarudin, R. Aulia, J. Jamaluddin, A. Yudithwa, and A. Ayuri, “Penerapan Algoritma K-Means dan Analisisnya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa,” *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 22, no. 2, p. 401, 2023, doi: 10.53513/jis.v22i2.8461.