

Penerapan Algoritma K-Means dan Analisisnya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa

Rina Novita¹, Agus Nur Khomarudin², Romy Aulia³, Adesia Yudithwa⁴, Arti Ayuri⁵

^{1,2,3} Politeknik Pertanian Negeri Payakumbuh, Sumatera Barat, Indonesia

^{4,5} UIN Sjech. M. Djamil Djambek Bukittinggi, Sumatera Barat, Indonesia

Email: ¹rinanovita12345@gmail.com, ^{2,*}agusnurkhumarudin@gmail.com, ³romysinggalang@gmail.com,

⁴yudithwa@gmail.com, ⁵arthiyuri01@gmail.com

Email Penulis Korespondensi: agusnurkhumarudin@gmail.com

Article History:

Received Jun 12th, 202x

Revised Aug 20th, 202x

Accepted Aug 26th, 202x

Abstrak

Pertambahan jumlah mahasiswa yang menyelesaikan studi setiap tahun pada suatu program studi telah menghasilkan banyak tumpukan data dan tentunya menimbulkan sebuah pertanyaan besar, yaitu "pengetahuan apa yang dapat dihasilkan dari tumpukan data tersebut?". Menjawab pertanyaan besar tersebut peneliti mencoba untuk menerapkan suatu teknik *data mining* yang merupakan sebuah tahapan dalam *Knowledge Discovery in Database* (KDD). Usaha meningkatkan kualitas mahasiswa yang dilakukan oleh prodi PTIK UIN Bukittinggi menghadapi beberapa kendala, diantaranya yaitu secara umum tugas akhir mahasiswa dirasakan mengalami penurunan jika dibandingkan dengan beberapa tahun sebelumnya. Selain itu, prodi belum mensosialisasikan kepada dosen dan mahasiswa tentang judul penelitian tugas akhir mahasiswa dan juga *roadmap* atau arah penelitian ke depan. Masalah lainnya yaitu terdapat mahasiswa yang mengajukan judul penelitian tugas akhir namun kurang sesuai dengan empat domain skill pada prodi PTIK. Metodologi pengembangan *data mining* yang digunakan dalam penelitian ini adalah *CRIPS-DM* merupakan singkatan dari *Cross-Industry Standart Process for Data Mining*. Hasil dari penerapan *data mining* metode *clustering* dengan *algoritma K-Means* yang diterapkan dalam penelitian ini memperoleh tiga buah *cluster* atau kelompok dengan beberapa analisis setiap *cluster*. Analisis dari setiap *cluster* menghasilkan rekomendasi untuk program studi dalam mengambil kebijakan strategis terhadap mahasiswa yang sedang dalam penyelesaian studi pada prodi PTIK UIN Bukittinggi.

Kata Kunci : Algoritma k-Means, Data Mining, Penyelesaian Studi Mahasiswa

Abstract

The increase in the number of students completing studies each year in a study program has resulted in many piles of data and of course raises a big question, namely "what knowledge can be generated from these piles of data?". In answering this big question the researcher tries to apply a data mining technique which is a stage in *Knowledge Discovery in Database* (KDD). Efforts to improve the quality of students carried out by the PTIK UIN Bukittinggi study program faced several obstacles, including that in general, student final assignments were felt to have decreased when compared to several previous years. In addition, the study program has not disseminated to lecturers and students about the research title of the student's final project and also the *roadmap* or direction of future research. Another problem is that there are students who submit research titles for their final project but are not in accordance with the four skill domains in the PTIK study program. The data mining development methodology used in this research is *CRIPS-DM* which stands for *Cross-Industry Standard Process for Data Mining*. The results of applying the data mining clustering method with the *K-Means* algorithm applied in this study obtained three clusters or groups with several analyzes for each cluster. The analysis of each cluster produces recommendations for study programs in taking strategic policies for students who are currently completing their studies at the PTIK UIN Bukittinggi study program.

Keyword : k-Means Algorithm, Data Mining, Completion of Student Studies

1. PENDAHULUAN

Pertambahan jumlah mahasiswa yang menyelesaikan studi setiap tahun pada suatu program studi telah menghasilkan banyak tumpukan data. Pertambahan data mahasiswa yang banyak tersebut tentunya menimbulkan sebuah pertanyaan besar, yaitu “pengetahuan apa yang dapat dihasilkan dari tumpukan data tersebut?”. Menjawab pertanyaan besar tersebut peneliti mencoba untuk menerapkan suatu teknik data mining yang merupakan sebuah tahapan dalam *Knowledge Discovery in Database* (KDD).

Knowledge Discovery in Database (KDD) adalah proses menemukan pengetahuan yang berguna dari sebuah data yang bervolume besar, dan sering disebut sebagai data mining [1]. KDD adalah proses yang terorganisir untuk mengidentifikasi pola-pola yang berlaku, berguna dan mudah dipahami dari kumpulan data yang besar dan kompleks. Data yang dikumpulkan dari berbagai sumber heterogen yang terintegrasi ke dalam penyimpanan data tunggal yang disebut sebagai data target. Data yang relevan diputuskan untuk dianalisis dan diperoleh dari pengumpulan data [2]. Kemudian, itu adalah pra-diproses dan berubah menjadi format standar yang sesuai. Data mining adalah langkah yang paling inti dalam algoritma/teknik kecerdasan yang diterapkan untuk mengekstrak pola atau aturan yang bermakna [3].

Adapun *Data mining* adalah inti dari proses KDD, yang melibatkan dalam menyimpulkan algoritma yang menjelajahi data, mengembangkan model dan menemukan pola-pola yang sebelumnya tidak diketahui [4]. Menurut Tacbir, *data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar [5]. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Proses menggali informasi dalam *data mining* melibatkan integrasi teknik dari berbagai disiplin ilmu, seperti teknologi *database* dan *data warehouse*, statistik, *machine learning*, komputasi dengan kinerja tinggi, *pattern recognition*, *neural network*, visualisasi data dan sebagainya [6].

Terdapat dua teknik yang dapat digunakan dalam *data mining* yaitu prediksi dan deskripsi [7]. Teknik prediksi menggunakan data historis untuk menyimpulkan sesuatu tentang kejadian di masa depan. Sedangkan teknik deskripsi bertujuan untuk menemukan pola dalam data yang menyediakan beberapa informasi tentang hubungan interval yang tersembunyi [8]. Menurut Kumar dan Saurabh, terdapat beberapa teknik yang digunakan dalam *data mining*, yaitu: *classification*, *clustering*, *predication*, *association rule*, *neural network*, *decision trees* dan *nearest neighbor method* [9].

Penelitian ini mengambil salah satu teknik *data mining* tersebut di atas yaitu *clustering*. *Clustering* adalah proses pengorganisasian objek data ke dalam *set* kelas yang saling berhubungan, yang disebut *cluster*. *Clustering* merupakan contoh dari klasifikasi tanpa arahan (*unsupervised*). Klasifikasi merujuk kepada prosedur yang menetapkan objek data set kelas. *Unsupervised* berarti bahwa pengelompokan tidak tergantung pada standar kelas dan pelatihan atau *training* [10]. Deka menjelaskan, *Clustering* merupakan salah satu teknik *data mining* yang digunakan untuk mendapatkan kelompok-kelompok dari objek-objek yang mempunyai karakteristik yang umum di data yang cukup besar [11]. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* atau grup sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin [12].

Terdapat algoritma yang cukup populer dalam teknik data *clustering* yaitu algoritma *K-means Clustering*. *K-means clustering* merupakan salah satu metode *cluster analysis* non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain [13].

Penelitian tentang penerapan algoritma *k-means clustering* sebelumnya telah dilakukan oleh Penelitian tentang penerapan algoritma *k-means clustering* sebelumnya telah dilakukan oleh Ramdani Budiman dan Rudianto pada tahun 2019 dalam menentukan lokasi promosi penerimaan mahasiswa baru ada Universitas Banten Jaya sehingga dapat diketahui daerah – daerah yang memiliki potensi untuk dilakukan promosi agar meningkatkan jumlah mahasiswa baru [14]. Penelitian serupa juga telah dilakukan oleh Fitri Yunita dalam menerapkan data mining pada penerimaan mahasiswa baru di Universitas Islam Indagiri yang menghasilkan sebuah pengetahuan untuk menentukan strategi dalam mempromosikan masing-masing program studi yang terdapat pada universitas tersebut [15]. Selain itu, penelitian terkait penerapan algoritma *k-mean clustering* juga telah dilakukan oleh Rony Setiawan pada Politeknik LP3I Jakarta untuk menghasilkan *cluster* penerimaan mahasiswa baru yang dapat membantu memberikan informasi kebijakan strategis dalam penerimaan mahasiswa baru [16].

Penelitian lain tentang penerapan algoritma *k-means clustering* sebelumnya juga telah dilakukan oleh Johan Oscar Ong, dalam mengelompokkan data mahasiswa yang telah lulus dari *President University*. Sehingga bagian *marketing President University* mendapatkan strategi pemasaran dalam menerima calon mahasiswa baru [17]. Penelitian lain yang terkait dengan penerapan algoritma *K-means Clustering* dilakukan oleh Gibran Satya Nugraha, yaitu Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means. Penelitian

tersebut di atas menerapkan algoritma *k-means clustering* dalam bidang pendidikan, akan tetapi belum ada peneliti yang menerapkan algoritma *k-means clustering* untuk menentukan *cluster* mahasiswa yang sedang menyelesaikan studi [18].

Perguruan Tinggi Keagamaan Islam Negeri (PTKIN) merupakan perguruan tinggi islam yang memiliki jumlah peminat yang tinggi. Oleh karena itu, PTKIN perlu mengembangkan inovasi – inovasi yang dapat meningkatkan mutu dan memberikan informasi serta strategi pengembangan yang berguna untuk kemajuan PTKIN tersebut. Universitas Islam Negeri (UIN) Bukittinggi sebagai salah satu PTKIN yang terus berbenah untuk terus berkembang baik dari sisi kuantitas maupun kualitas. Terdapat beberapa program studi dan program studi Pendidikan Teknik Informatika dan Komputer (PTIK) menjadi pilihan program studi pada Fakultas Tarbiyah dan Ilmu Kependidikan yang memiliki jumlah mahasiswa dan peminat yang cukup banyak. PTIK UIN Bukittinggi adalah salah satu prodi yang cukup banyak menghasilkan lulusan dengan bidang keahlian IT yang dibekali empat domain skill yaitu pendidikan, pemrograman, jaringan, multimedia. Saat ini prodi PTIK telah berupaya untuk meningkatkan kualitas mahasiswa, khususnya mahasiswa yang sedang menyelesaikan tugas akhir. Diantara terobosan yang dilakukan oleh prodi PTIK diantaranya adalah membuat format penulisan tugas akhir yang sesuai dengan tuntutan perkembangan teknologi saat ini, seperti format pengutipan atau sitasi wajib menggunakan mendeley. Kemudian tugas akhir wajib dijadikan sebagai artikel dan harus diusahakan untuk dapat diterbitkan dalam jurnal. Kewajiban membuat artikel jurnal ini merupakan syarat untuk mengambil ijazah dan tentunya tidak hanya menjadi output saja, namun juga menjadi outcome dari tugas akhir yang sangat berguna bagi mahasiswa dan dosen pembimbing.

Dibalik usaha meningkatkan kualitas mahasiswa yang dilakukan oleh prodi PTIK tersebut di atas, terdapat beberapa kendala yang dihadapi. Diantara kendala tersebut yaitu secara umum tugas akhir mahasiswa dirasakan mengalami penurunan jika dibandingkan dengan beberapa tahun sebelumnya, seperti dari judul penelitian yang berkulat seputar perancangan program atau media yang tidak mencoba mengeksplor bidang komputer sains lain seperti sistem pakar, sistem pendukung keputusan, jaringan syaraf tiruan, logika fuzzy, data mining, multimedia 3 dimensi, augmented reality dan sebagainya. Selain itu, mahasiswa cukup kesulitan mengajukan judul penelitian tugas akhir disebabkan telah ada penelitian yang sudah dilakukan oleh pendahulunya. Disisi lain, prodi belum mensosialisasikan kepada dosen dan mahasiswa tentang judul-judul penelitian tugas akhir mahasiswa dan juga roadmap atau arah penelitian selanjutnya untuk mahasiswa. Masalah lainnya yaitu terdapat mahasiswa yang mengajukan judul penelitian tugas akhir namun kurang sesuai dengan empat domain skill pada prodi PTIK.

Salah satu solusi yang akan peneliti lakukan untuk mengatasi permasalahan yang telah tertulis di atas adalah dengan menerapkan teknik *data mining* yaitu *clustering* menggunakan algoritma *k-means clustering*. Hasil penerapan algoritma *k-means clustering* adalah berupa analisis pola *cluster* mahasiswa yang telah dan sedang dalam penyelesaian studi. Dengan demikian dapat memberikan informasi kepada prodi terkait dengan cara atau kebijakan strategis yang dapat diambil oleh prodi untuk ke depannya.

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

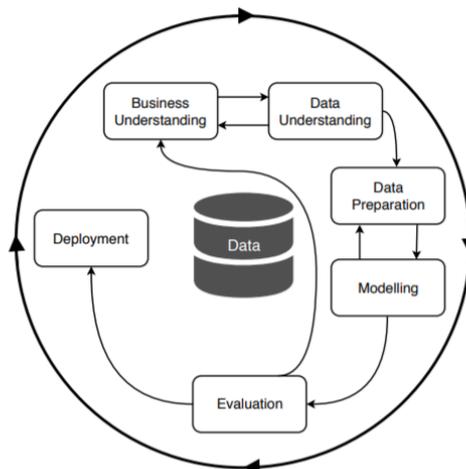
Sumber Data, Jenis sumber data yang didapatkan peneliti adalah data primer dan data sekunder. Data primer dalam penelitian peneliti peroleh dari wawancara secara langsung dengan kepala dan staff program studi pendidikan teknik informatika dan komputer Fakultas Tarbiyah dan Ilmu Keguruan serta staff Unit Pelaksana Teknis (UPT) Teknologi Informasi dan Pangkalan Data UIN Bukittinggi. Data sekunder pada penelitian ini dikategorikan menjadi 2 kategori diantaranya adalah kategori data mahasiswa yang sudah lulus serta data pengajuan judul yang akan telah disetujui oleh ketua program studi PTIK. Masing-masing terdiri dari data NIM, nama mahasiswa, IPK/IPs, judul tugas akhir, tema tugas akhir, nilai mata kuliah sesuai kelompoknya.

Pengumpulan Data, Metode pengumpulan data pada penelitian ini dilakukan dengan beberapa cara diantaranya: *Wawancara*, Wawancara dilakukan terhadap ketua program studi pendidikan teknik informatika dan komputer UIN Bukittinggi dengan tujuan untuk memperoleh gambaran dan konsep dari tahapan penyelesaian studi mahasiswa. *Studi Pustaka*, Studi pustaka dilakukan dengan tujuan untuk memperoleh referensi terkait penelitian yang sedang dilakukan. Studi pustaka dalam penelitian ini didasarkan dari jurnal, buku, internet, dan sebagainya. *Survey*, Survey penelitian dilakukan dengan meminta ijin untuk pengambilan data mengenai data mahasiswa dan transkrip nilai mahasiswa terhadap mahasiswa yang dalam masa penyelesaian studi, pengajuan skripsi, dan terhadap mahasiswa yang sudah menyelesaikan studi.

Jenis Data, Jenis data yang digunakan dalam penyelesaian permasalahan dalam penelitian ini adalah data kuantitatif. Data kuantitatif dalam penelitian ini meliputi : NIM, Nama Mahasiswa, Judul Tugas Akhir, Tema Tugas Akhir, Rerata nilai matakuliah bidang multimedia, Rerata nilai matakuliah bidang pemrograman, Rerata nilai matakuliah bidang jaringan, Rerata nilai matakuliah bidang pendidikan.

2.2 Metode Analisis Data

Metodologi pengembangan data mining yang digunakan dalam penelitian ini adalah CRISP-DM. CRISP-DM merupakan singkatan dari *Cross- Industry Standart Process for Data Mining*. Tahapan dalam CRIPS-DM terdapat enam tahapan berurutan yang digambarkan pada gambar 1 [19].



Gambar 1. Tahapan CRIPS-DM

Tahapan pengembangan CRIPS-DM ini adalah: *Business Understanding*, Tahapan pertama ini dimulai dengan memahami permasalahan yang terdapat dalam penyelesaian studi mahasiswa pada Program Studi Pendidikan Teknik Informatika dan Komputer UIN Bukittinggi. Tujuan penelitian ini adalah untuk menggali pengetahuan yang tersembunyi dari rutinitas penyelesaian studi mahasiswa yang telah selesai pada periode-periode sebelumnya. Hal ini diharapkan mampu memberikan gambaran kepada mahasiswa-mahasiswa yang sedang atau akan menyelesaikan studi untuk dapat menyelesaikan studi lebih cepat dan tepat waktu. Selain itu juga memberi rekomendasi kepada program studi dalam membuat keputusan strategis untuk siswa yang dalam proses penyelesaian studi.

Data Understanding, *Data understanding* dilakukan dengan mendapatkan pemahaman yang baik terhadap kebutuhan data yang diperlukan dalam penyelesaian masalah dalam penelitian ini. Adapun kebutuhan data dalam penelitian ini yaitu data mahasiswa, data judul tugas akhir, data tema tugas akhir, nilai mata kuliah bidang multimedia, nilai mata kuliah bidang pemrograman, nilai mata kuliah bidang jaringan dan nilai mata kuliah bidang pendidikan.

Data Preparation, Pada tahapan data preparation penulis mengolah data yang didapat dengan menggunakan beberapa tahapan yang terdapat dalam KDD. KDD merupakan singkatan dari Knowledge Discovery in Database. Beberapa tahapan KDD yang digunakan dalam penelitian ini adalah melakukan pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), dan melakukan transformasi data (*data transformation*) [20].

Modeling, Fase pemodelan pada penelitian ini secara langsung melibatkan teknik data mining. Tahapan yang kita lakukan pada fase ini yang pertama adalah memilih dan aplikasikan teknik pemodelan yang sesuai, kemudian kalibrasi aturan model untuk mengoptimalkan hasil. Perlu diperhatikan bahwa beberapa teknik mungkin digunakan untuk permasalahan data mining yang sama. Dan jika diperlukan, proses dapat kembali ke fase pengolahan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu.

Evaluation, Pada fase evaluasi ini, kita mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarakan untuk digunakan, kemudian menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal, kemudian menetapkan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik, dan terakhir Mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.

Deployment, Pada tahap ini penulis mencoba menyajikan hasil dari penerapan data mining metode clustering dengan algoritma K-Means yang telah dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Business Understanding (Pemahaman Terhadap Bisnis)

Program Studi Pendidikan Teknik Informatika dan Komputer (PTIK) Fakultas Tarbiyah dan Ilmu Kependidikan UIN Bukittinggi telah dimulai sejak tahun 2017. Prodi PTIK ini merupakan salah satu program studi yang sangat diminati. Hal ini dapat dilihat jumlah peminat yang kian meningkat setiap tahunnya. Lulusan dari prodi PTIK ini dibekali dengan empat domain utama bidang keahlian IT diantaranya adalah bidang multimedia, bidang pemrograman, bidang jaringan, dan bidang pendidikan. Saat ini prodi PTIK telah berupaya untuk meningkatkan kualitas mahasiswa, khususnya mahasiswa yang sedang menyelesaikan tugas akhir.

Berbagai upaya dan inovasi terus dilakukan oleh prodi PTIK untuk meningkatkan kualitas lulusan diantaranya adalah dengan membuat format penulisan tugas akhir yang sesuai dengan tuntutan perkembangan teknologi saat ini, seperti format pengutipan atau sitasi wajib menggunakan mendeley. Kemudian tugas akhir wajib dijadikan sebagai artikel dan harus diusahakan untuk dapat diterbitkan dalam jurnal. Kewajiban membuat artikel jurnal ini merupakan syarat untuk mengambil ijazah dan tentunya tidak hanya menjadi output saja, namun juga menjadi outcome dari tugas akhir yang sangat berguna bagi mahasiswa dan dosen pembimbing.

Tingkat penyelesaian studi mahasiswa program studi pendidikan teknik informatika dan komputer UIN Bukittinggi beberapa tahun terakhir mengalami sedikit kemunduran di banding dengan tahun-tahun awal berdirinya program studi ini. Motivasi mahasiswa cenderung menurun dalam menyelesaikan studinya serta lambatnya pengajuan judul skripsi yang merupakan syarat dari kelulusan pada tingkat sarjana pada program studi ini. Judul skripsi yang diajukan oleh mahasiswa pun dinilai kurang memenuhi standard dari program studi dan cenderung mirip atau sama dengan judul-judul skripsi yang telah ada sebelumnya. Oleh karena itu, hal ini juga merupakan penyebab utama penolakan judul skripsi yang diajukan oleh mahasiswa.

3.1.2 Data Understanding

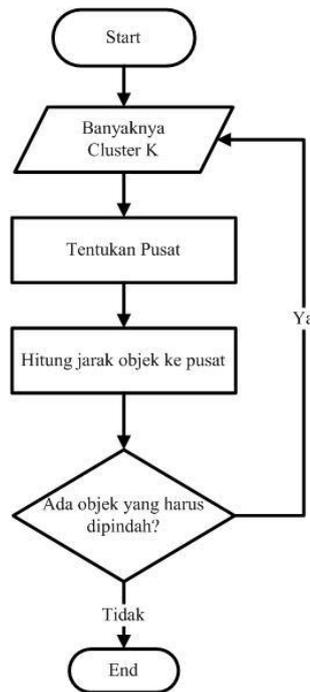
Data yang dibutuhkan dalam penelitian ini diantaranya adalah : Data mahasiswa yang meliputi nama, nim, judul tugas akhir, dan tema tugas akhir; Data nilai mata kuliah bidang multimedia yang merupakan nilai rata-rata dari nilai mata kuliah aplikasi grafika komputer, nilai mata kuliah multimedia dan nilai mata kuliah perancangan grafik 3D; Data nilai mata kuliah bidang pemrograman yang merupakan nilai rata-rata dari nilai mata kuliah bahasa pemrograman I (C++), nilai mata kuliah Bahasa pemrograman II (VB NET), nilai mata kuliah Bahasa pemrograman III (PHP/MYSQL), nilai mata kuliah web programming, dan nilai mata kuliah object oriented programming (JAVA); Data nilai mata kuliah bidang jaringan yang merupakan nilai rata-rata dari nilai mata kuliah jaringan komputer, nilai mata kuliah jaringan nirkabel, dan nilai mata kuliah *software installation network communication* dan; Data nilai mata kuliah bidang pendidikan yang merupakan nilai rata-rata dari nilai mata kuliah ilmu pendidikan, nilai mata kuliah pengembangan kurikulum, nilai mata kuliah pengelolaan kelas, nilai mata kuliah evaluasi pembelajaran TIK, nilai mata kuliah perencanaan pengajaran TIK, dan nilai mata kuliah strategi pembelajaran TIK.

3.1.3 Data Preparation

Data yang telah diperoleh pada tahapan data understanding selanjutnya dilakukan diskusi terfokus dengan dilaksanakannya kegiatan Forum Grup Discussion (FGD) tahap I. Pada kegiatan FGD 1 ini peneliti bersama tim mengundang pemateri yang kompeten dalam bidang ini dan dihadiri oleh 5 orang peserta. Hasil diskusi dari kegiatan FGD tahap 1 ini berfokus kepada persiapan data yang digunakan dalam penelitian ini dan selanjutnya dilakukan pengolahan dengan menggunakan beberapa tahapan yang terdapat pada *Knowlegde Discovery in Database* (KDD). Tahapan tersebut diantaranya adalah melakukan pembersihan terhadap data (*data cleaning*), melakukan integrasi data (*data integration*), melakukan pemilihan data (*data selection*), dan melakukan transformasi pada data (*data transformasi*). Pengolahan data dilakukan dengan menggunakan aplikasi *microsoft excel*. Pada tahapan persiapan data, penulis melakukan inisialisasi dan perhitungan untuk mendapatkan nilai rerata dari setiap bidang keilmuan dalam prodi PTIK.

3.1.4 Modeling

Setelah semua data ditransformasikan, selanjutnya data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *k-means clustering*. Pengolahan data menggunakan algoritma *k-means clustering* dilakukan dengan langkah-langkah yang dapat dilihat pada flowchart berikut ini :



Gambar 2. Flowchart Algoritma *K-means Clustering*

Berdasarkan gambar 2, maka langkah-langkah pengolahan data sampel mahasiswa menggunakan algoritma *k-means clustering* adalah sebagai berikut : Menentukan jumlah *cluster*, pada penelitian ini penulis akan mengelompokkan data yang ada menjadi tiga *cluster*. Menentukan nilai awal titik tengah atau pusat *cluster*, penentuan nilai awal titik tengah atau pusat *cluster* dilakukan secara acak dan didapat pusat dari setiap *cluster* [21]. Menghitung jarak objek ke *centroid* dengan menggunakan rumus jarak *Euclidean* :

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

Selanjutnya melakukan *clustering* objek dengan memasukkan setiap objek ke dalam *cluster* berdasarkan pada jarak minimumnya. Suatu data akan menjadi anggota dari suatu *cluster* (C1, C2 atau C3) yang memiliki jarak atau nilai terkecil dari pusat *cluster*-nya. Langkah berikutnya yaitu *menghitung pusat cluster baru*. Pusat *cluster* yang baru ditentukan berdasarkan pengelompokan anggota masing-masing *cluster*. *Mengulangi iterasi dengan dimulai dari langkah 3*, sehingga *cluster* yang baru memiliki angka yang tetap [22].

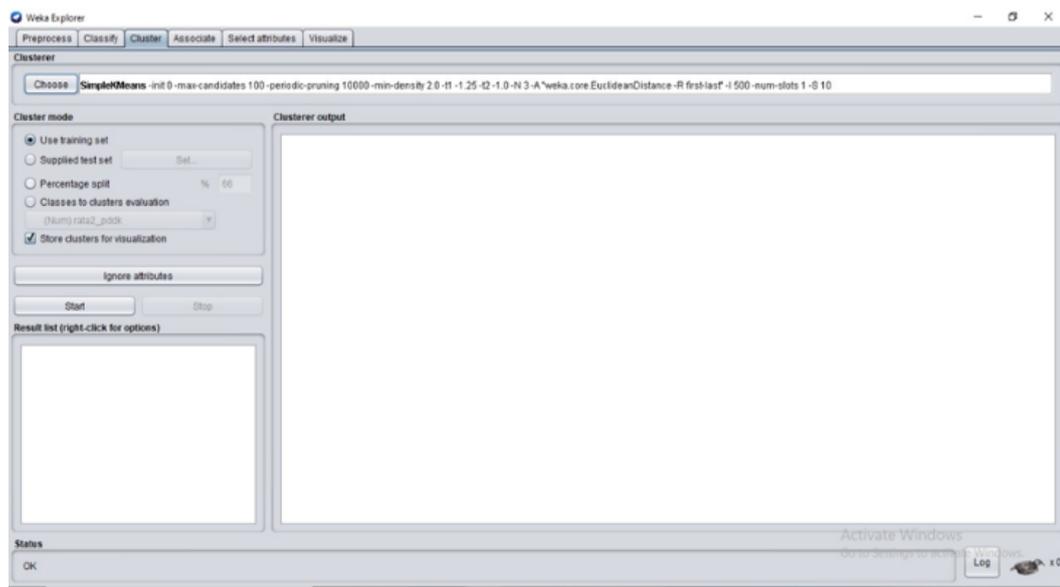
3.1.5 Evaluation

Berdasarkan data yang diperoleh dalam tahapan modeling dilakukan evaluasi terhadap kesesuaian proses yang dilakukan secara manual yang selanjutnya dibandingkan dengan pengujian dengan menggunakan *software*. Hal ini dilakukan agar memperoleh hasil yang valid. Gambar 3 merupakan tampilan awal dari aplikasi *Waikato Environment for Knowledge Analysis (WEKA)* yang sebelumnya telah di install pada *personal computer (PC)*. Versi *WEKA* yang digunakan dalam penelitian ini adalah versi 3.8.5.



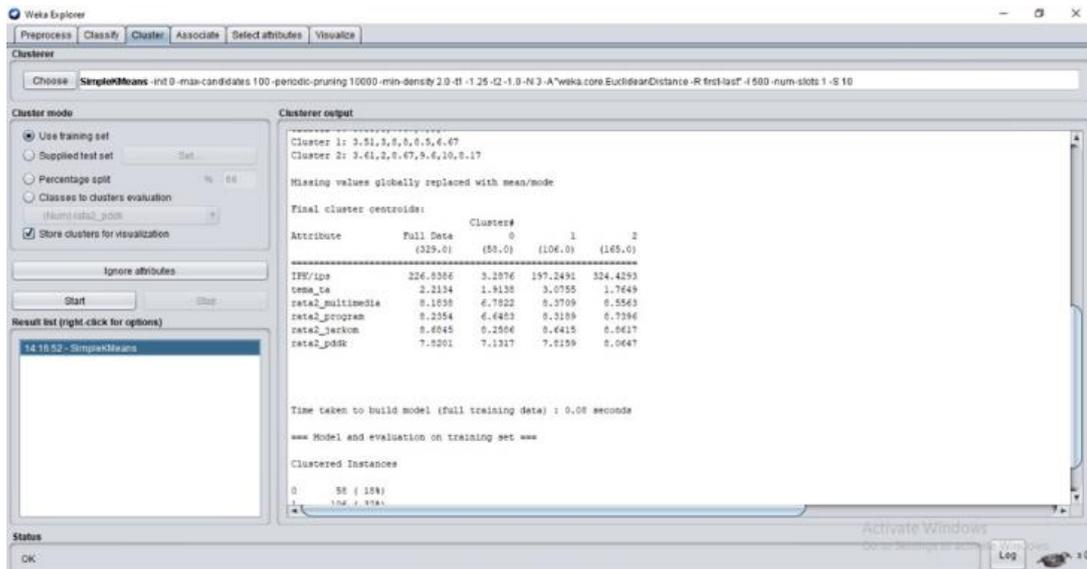
Gambar 3. Tampilan awal WEKA

Tahapan selanjutnya adalah meng-*upload* data yang akan diuji. Data yang dapat diproses dalam aplikasi ini adalah data dengan *format* .CRV. Data yang sudah diupload ini kemudian ditentukan teknik yang akan digunakan. Dalam penelitian ini digunakan teknik *clustering* dengan menggunakan algoritma *k-means clustering*. Gambar 4 menunjukkan pemilihan teknik data mining yang akan digunakan.

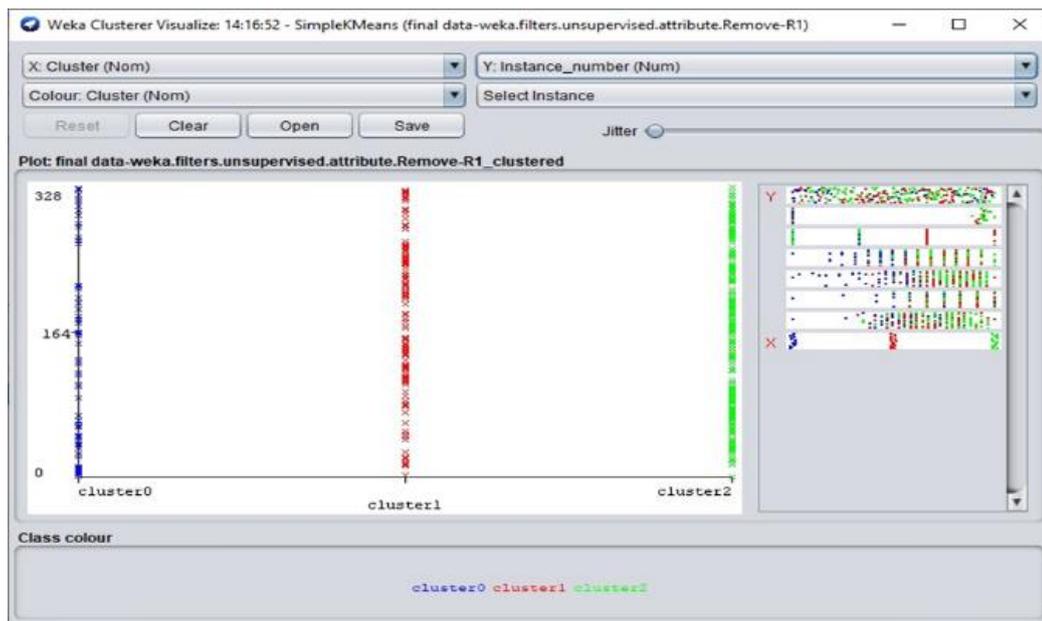


Gambar 4. pemilihan teknik data mining dalam aplikasi WEKA

Hasil pengolahan data mining teknik *clustering* dengan menggunakan aplikasi WEKA akan menghasilkan 3 *cluster*. Gambar 5 dan 6 menunjukkan hasil *cluster* yang terbentuk pada aplikasi WEKA.



Gambar 5. Hasil Clustering pada Aplikasi WEKA



Gambar 6. Hasil Pola Clustering pada Aplikasi WEKA

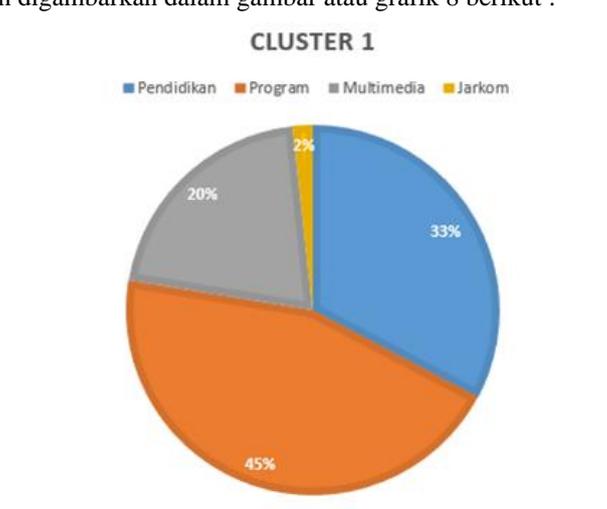
Langkah terakhir dalam proses pengujian dengan aplikasi WEKA ini adalah dengan mengambil data yang masih dalam format .rtf untuk diolah kembali dengan menggunakan tool Microsoft excel. Gambar 7 menunjukkan hasil clustering dalam format .rtf pada aplikasi weka.

Instance Number	rata2_ipk	rata2_multimedia	rata2_program	rata2_jaricom	rata2_pdid	Cluster		
1	9.0	3.28	1.0	8.33	7.4	9.5	7.0	cluster2
2	1.0	3.13	3.0	6.67	4.8	9.5	7.17	cluster0
3	2.0	3.25	3.0	6.67	7.6	9.0	6.17	cluster1
4	3.0	3.14	3.0	6.33	5.6	9.5	7.17	cluster0
5	4.0	3.24	2.0	5.67	7.8	9.0	7.0	cluster0
6	5.0	3.17	3.0	6.0	4.8	9.0	6.33	cluster1
7	6.0	3.2	2.0	6.33	6.4	9.5	6.17	cluster0
8	7.0	3.17	1.0	6.0	5.8	9.0	6.5	cluster0
9	8.0	3.14	1.0	7.0	5.9	9.5	6.83	cluster0
10	9.0	3.69	1.0	6.67	5.4	9.0	6.33	cluster0
11	10.0	3.29	1.0	6.67	6.0	7.0	7.5	cluster0
12	11.0	3.23	4.0	5.33	5.8	8.0	7.17	cluster0
13	12.0	3.57	3.0	6.67	9.0	9.0	8.33	cluster1
14	13.0	3.6	3.0	9.0	9.4	9.5	9.33	cluster1
15	14.0	3.45	2.0	8.33	7.4	8.5	7.5	cluster2
16	15.0	3.37	3.0	7.67	6.6	7.5	8.33	cluster1
17	16.0	3.57	2.0	7.67	7.8	9.5	8.5	cluster2
18	17.0	3.48	3.0	9.0	8.8	9.5	8.5	cluster1
19	18.0	3.55	2.0	7.67	8.8	9.0	8.33	cluster2
20	19.0	3.35	1.0	8.33	8.0	9.0	7.83	cluster2
21	20.0	3.55	2.0	9.0	9.0	9.5	6.17	cluster2
22	21.0	3.27	3.0	8.0	7.4	7.5	7.33	cluster1
23	22.0	3.32	3.0	8.67	7.8	6.0	7.33	cluster0
24	23.0	3.3	2.0	7.0	7.6	8.5	8.0	cluster0
25	24.0	3.37	2.0	7.0	8.2	8.5	8.5	cluster2
26	25.0	3.13	1.0	7.0	5.4	9.0	7.83	cluster0
27	26.0	3.53	2.0	7.33	8.4	9.0	8.33	cluster2
28	27.0	3.37	2.0	9.0	9.0	9.0	8.33	cluster2
29	28.0	3.45	3.0	8.0	8.2	9.0	8.33	cluster1
30	29.0	3.5	2.0	7.0	9.6	7.0	8.5	cluster2
31	30.0	3.45	2.0	9.0	8.4	8.5	7.8	cluster2
32	31.0	3.62	2.0	9.0	8.8	9.5	9.17	cluster2
33	32.0	3.29	2.0	7.33	7.8	8.5	7.33	cluster0
34	33.0	3.43	2.0	8.33	8.8	9.5	9.0	cluster2
35	34.0	3.49	2.0	9.33	8.4	9.5	9.33	cluster2
36	35.0	3.45	2.0	9.33	8.6	9.5	7.83	cluster2
37	36.0	3.16	2.0	6.0	6.0	7.5	7.33	cluster0
38	37.0	3.45	3.0	8.0	8.6	8.0	8.33	cluster0

Gambar 7. Hasil Clustering dalam Format .rtf pada Aplikasi WEKA

Berdasarkan data yang diperoleh dengan pengolahan data yang dilakukan secara manual yang kemudian dibandingkan dengan pengolahan data melalui aplikasi WEKA ini diperoleh hasil yang sama yaitu terdapat 3 buah cluster atau kelompok mahasiswa dalam penyelesaian studi yang terdapat pada program studi pendidikan teknik informatika dan komputer fakultas tarbiyah dan ilmu keguruan UIN Bukittinggi.

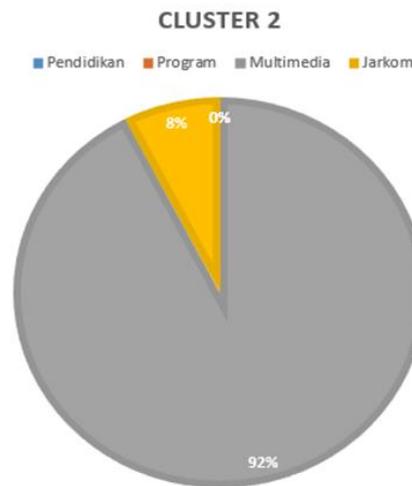
Hasil cluster 1 selanjutnya digambarkan dalam gambar atau grafik 8 berikut :



Gambar 8. Hasil cluster 1

Dapat disimpulkan bahwa pada cluster 1 mahasiswa dengan IPK ≤ 3.29 , rerata nilai mata kuliah bidang multimedia 6.78, rerata nilai mata kuliah bidang pemrograman adalah 6.78, rerata nilai mata kuliah bidang jaringan komputer adalah 8.26 dan rerata nilai mata kuliah bidang pendidikan adalah 7.13 sebesar 45% memilih tema tugas akhir tentang pemrograman. Persentase kedua tertinggi berikutnya adalah pada bidang pendidikan yaitu sebesar 33%. Selanjutnya mahasiswa memilih tema tugas akhir pada bidang multimedia yaitu sebesar 20 % dan 2% untuk bidang jaringan komputer.

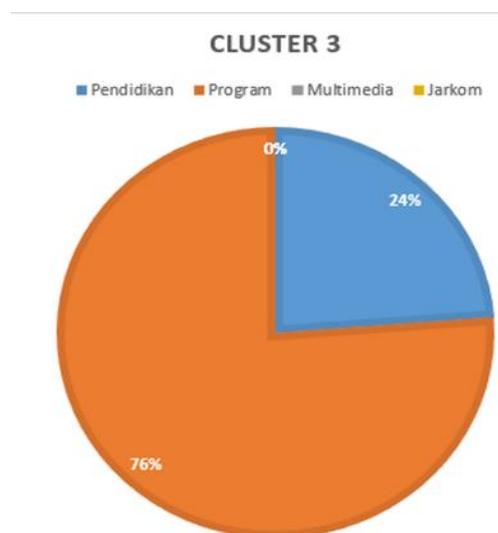
Hasil cluster 2 selanjutnya digambarkan dalam grafik atau gambar 9 berikut :



Gambar 9. Hasil *cluter* 2

Cluster 2 menunjukkan bahwa mahasiswa yang memperoleh IPK 3.50, rerata nilai mata kuliah bidang multimedia 8.37, rerata nilai mata kuliah bidang pemrograman adalah 8.32, rerata nilai mata kuliah bidang jaringan komputer adalah 8.64 dan rerata nilai mata kuliah bidang pendidikan adalah 7.82 hanya memilih 2 bidang saja yaitu bidang multimedia dan jaringan komputer. Persentase tertinggi yang terdapat pada *cluster* 2 ini adalah bidang multimedia sebesar 92% dan sisanya sebesar 8% memilih bidang jaringan komputer.

Hasil *cluster* 3 selanjutnya digambarkan dalam grafik atau gambar 10 berikut :



Gambar 10. Hasil *cluter* 3

Hasil *Cluster* 3 menunjukkan bahwa mahasiswa yang berada pada *cluster* ini memilih tema tugas akhir dalam bidang pendidikan dan pemrograman. Pada *cluster* 3 ini umumnya mahasiswa memperoleh IPK ≥ 3.55 rerata nilai mata kuliah bidang multimedia sebesar 8.56, rerata nilai mata kuliah bidang pemrograman sebesar 8.74, rerata nilai mata kuliah bidang jaringan komputer sebesar 8.86 dan rerata nilai mata kuliah bidang pendidikan sebesar 8.07. Persentase tertinggi pada *cluster* 3 ini adalah bidang pemrograman yaitu sebesar 76 % dan 24 % mahasiswa memilih bidang pendidikan.

Hasil yang telah diperoleh ini selanjutnya dilakukan diskusi dalam sebuah forum melalui kegiatan *Focus Group Discussion* (FGD) tahap 2 dengan mengundang ibu Amelia Permata Sari, M.Kom dan dihadiri sebanyak 15 orang peserta dalam kegiatan ini. Kegiatan FGD tahap 2 ini membahas mengenai *evaluation* dan *development*. Kegiatan FGD

tahap 2 juga mendiskusikan tentang masukan atau rekomendasi kepada program studi untuk perbaikan dan pengembangan ke depannya.

3.1.6 Deployment

Hasil dari penerapan data mining metode *clustering* dengan algoritma K-Means yang diterapkan dalam penelitian ini memperoleh beberapa rekomendasi sebagai berikut : Mahasiswa dengan rata rata IPK 3.29, rata-rata nilai mata kuliah bidang multimedia 6.78, nilai rata-rata mata kuliah pemrograman 6.65, nilai rata-rata mata kuliah jaringan komputer 8.26 dan nilai rata-rata pendidikan 7.13 umumnya mengambil judul tugas akhir dalam bidang pemrograman, pendidikan dan multimedia dengan bobot masing-masing 45%, 33% dan 20%; Mahasiswa yang memperoleh IPK 3.50, rata-rata nilai mata kuliah bidang multimedia 8.37, nilai rata-rata mata kuliah pemrograman 8.32, nilai rata-rata mata kuliah jaringan komputer 8.64 dan nilai rata-rata pendidikan 7.82 dominan mengambil tugas akhir dalam bidang multimedia yaitu sebesar 92 % dan siswanya jaringan; Mahasiswa dengan IPK rata 3.55, rata-rata nilai mata kuliah bidang multimedia 8.56, nilai rata-rata mata kuliah pemrograman 8.74, nilai rata-rata mata kuliah jaringan komputer 8.86 dan nilai rata-rata pendidikan 8.07 memilih tema tugas akhir dalam bidang pemrograman dan pendidikan dengan perbandingan 76% dan 24%; Penelitian yang dilakukan untuk 3 tahun ini yaitu mahasiswa yang lulus pada tahun 2019, 2020 dan yang sedang menyelesaikan studi pada tahun 2021 paling banyak mengambil tugas akhir dalam bidang pemrograman; Tema tugas akhir dalam bidang multimedia sangat banyak di ambil oleh mahasiswa dengan IPK 3.50 pada 3 tahun terakhir ini, oleh karena itu, agar tidak terjadi penumpukan pada satu bidang ini, maka dapat diusulkan pada bidang lain seperti jaringan yang masih sedikit penelitian terkait; Agar mahasiswa yang akan mengajukan judul tugas akhir mendapat gambaran arah penelitian sebaiknya dari program studi pendidikan teknik informatika dan komputer UIN Bukittinggi memiliki peta histori judul skripsi tentang tema dan judul yang sudah pernah di teliti sebelumnya.

3.2 Pembahasan

Penyelesaian studi merupakan dilema yang sering sekali dialami oleh mahasiswa terutama pada mahasiswa yang berada pada semester atas. Kurangnya informasi dan motivasi menjadi kendala yang menghambat penyelesaian studi mahasiswa selesai tepat waktu. Dalam sebuah penelitian yang dilakukan oleh Okta Sulita Sari dalam skripsinya yang berjudul faktor keterlambatan penyelesaian studi mahasiswa prodi bimbingan konseling islam Universitas Islam Negeri Bengkulu tahun 2018 menjelaskan bahwa terdapat beberapa faktor diantaranya disebabkan oleh faktor internal dan faktor eksternal. Faktor kurangnya motivasi dan problem pribadi merupakan faktor dari dalam (internal). Sedangkan faktor sudah menikah saat aktif kuliah, terlambat KKN, faktor ekonomi, dan kurangnya motivasi dari keluarga dalam memberi dukungan dan semangat merupakan faktor eksternal yang menyebabkan sulitnya penyelesaian studi tepat waktu.

Menyikapi banyaknya mahasiswa yang terkendala dalam penyelesaian studi, maka suatu program studi perlu menyiapkan berbagai strategi dan motivasi agar mahasiswa lebih terarah dan bersemangat agar bisa menyelesaikan studi tepat waktu. Bersumber kepada history data yang sudah ada pada program studi Pendidikan Teknik Informatika dan Komputer, maka dapat diperoleh sebuah informasi berupa pengetahuan baru yang dapat menjadi kebijakan pada program studi. Dengan melihat kasus-kasus yang ada dalam proses penyelesaian studi mahasiswa sebelumnya, suatu program studi dapat menyiasati serta memperoleh gambaran terkait penyelesaian studi mahasiswanya. Dengan menganalisa kasus yang terjadi, suatu program studi bisa mendapatkan beberapa kebijakan strategis dalam upaya membantu mahasiswa agar dapat menyelesaikan studi tepat waktu.

Teknologi data mining merupakan suatu teknologi yang dikembangkan untuk memperoleh suatu pengetahuan baru berdasarkan data historis yang telah tertumpuk. Dengan memanfaatkan data mining dihasilkan sebuah informasi yang berharga dan kemudian dapat dijadikan sebagai rekomendasi untuk menentukan dalam mengambil kebijakan untuk masa yang akan datang. Pemanfaatan data mining yang digunakan penulis dalam penelitian ini dilakukan pada program studi pendidikan teknik informatika dan komputer (PTIK) UIN Bukittinggi dalam hal penyelesaian studi mahasiswa. Informasi yang diperoleh diharapkan mampu menjadi dasar dalam penentuan kebijakan dari program studi kepada mahasiswa yang akan menyelesaikan studi mereka.

Hasil pengolahan data mining dengan menggunakan algoritma K-Means Clustering yang telah dilakukan pada program studi PTIK UIN Bukittinggi menghasilkan beberapa kebijakan strategis bagi program studi yang dapat digunakan untuk pengambilan keputusan di masa yang akan datang. Diantara hasil yang diperoleh meliputi nilai yang diperoleh mahasiswa pada setiap bidang keilmuan yang terdiri dari bidang multimedia, pemrograman, jaringan komputer dan pendidikan. Selain itu, program studi juga memperoleh informasi bahwa mahasiswa di angkatan ini dominan membuat tema judul tugas akhir dengan satu bidang yang sama, dengan informasi ini program studi dapat mengarahkan mahasiswa untuk mengambil tema lain yang akan dijadikan sebagai produk tugas akhirnya.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang mengacu pada tujuan penelitian ini, maka dapat disimpulkan bahwa: algoritma *k-means clustering* yang diterapkan pada program studi pendidikan teknik informatika dan komputer UIN Bukittinggi dapat menghasilkan kelompok atau *cluster* yang berguna untuk menentukan kebijakan strategis program studi terhadap mahasiswa yang sedang dalam penyelesaian studi. Teknologi data mining dengan algoritma *k-means clustering* dapat memberikan informasi dan pengetahuan baru yang dapat digunakan oleh program dalam menentukan kebijakan strategis program studi terhadap mahasiswa yang sedang dalam penyelesaian studi pada program studi pendidikan teknik informatika dan komputer UIN Bukittinggi.

Berdasarkan kesimpulan di atas, maka dalam penelitian ini dapat diberikan saran kepada beberapa pihak, yaitu : *Bagi program studi*, agar dapat memberikan peta tugas akhir mahasiswa sehingga dapat membantu mahasiswa dalam menentukan tema tugas akhir yang akan diajukan. *Bagi mahasiswa*, agar lebih termotivasi lagi dalam menyelesaikan studi tepat waktu. *Bagi peneliti selanjutnya*, agar dapat mengkaji dan membahas serta mengembangkan penelitian pada program studi yang lain serta dengan jumlah tahun yang lebih banyak lagi.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada tim penelitian ini dan kepada pihak terkait prodi Pendidikan Teknik Informatika dan Komputer UIN Bukittinggi.

DAFTAR PUSTAKA

- [1] S. Hameetha Begum, "Data Mining Tools and Trends – An Overview," *Int. J. Emerg. Res. Manag. & Technology*, vol. ISSN, no. February, pp. 2278–9359, 2013.
- [2] R. Rosmini, A. Fadlil, and S. Sunardi, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, pp. 22–31, 2018, doi: 10.25299/itjrd.2018.vol3(1).1773.
- [3] A. B. El Seddawy, "Adapted Framework for Data Mining Technique to Improve Decision Support System in an Uncertain Situation," *Int. J. Data Min. Knowl. Manag. Process*, vol. 2, no. 3, pp. 1–9, 2012, doi: 10.5121/ijdkp.2012.2301.
- [4] Y. Ramamohan, K. Vasantharao, C. K. Chakravarti, and a S. K. Ratnam, "A Study of Data Mining Tools in Knowledge Discovery Process," *Int. J. Soft Comput. Eng.*, vol. 2, no. 3, pp. 191–194, 2012.
- [5] Tacbir Hendro Pudjiantoro, F. Renaldi, and A. Teogunadi, "Penerapan Data Mining untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru," *Konf. Nas. Sist. Inform.*, 2011.
- [6] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013, doi: 10.1038/hdy.2009.180.
- [7] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [8] H. Sulastri and A. I. Gufroni, "The Application of Data Mining in the Grouping of Thalassemia Patients," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017.
- [9] B. K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students' Performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 6, pp. 63–69, 2011, doi: 10.1177/039463200201500108.
- [10] M. Yedla, S. R. Pathakota, and T. M. Srinivasa, "Enhancing K-means Clustering Algorithm with Improved Initial Center," *Int. Journal Comput. Sci. Inf. Technol.*, vol. 1, no. 2, pp. 121–125, 2010, doi: 10.1109/WKDD.2009.210.
- [11] D. Dwinavinta, C. Nugraha, M. Fahmi, Z. Naimah, and N. Setiani, "Klasterisasi Judul Buku dengan Menggunakan Metode K-Means," *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, vol. 21, no. 1, pp. 1907–5022, 2014.

- [12] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *J. Sist. dan Teknol. Inf.*, vol. 6, no. 4, p. 161, 2018, doi: 10.26418/justin.v6i4.29024.
- [13] Ediyanto, M. N. Mara, and N. Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133–136, 2013.
- [14] 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.
- [15] 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.
- [16] S. Rony, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," *J. Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.
- [17] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.
- [18] G. S. Nugraha and H. Hairani, "Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means," *J. MATRIK*, vol. 17, no. 2, pp. 13–23, 2018, doi: 10.30812/matrik.v17i2.84.
- [19] C. P. Lopez, *DATA MINING. The CRISP-DM METHODOLOGY. The CLEM language and IBM SPSS MODELER*. Lulu Press, Inc, 2021.
- [20] 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.
- [21] Agus Nur Khomarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>
- [22] J. Hutagalung and U. F. Sari, "Penerapan Metode K-Means dan MOORA dalam Penerimaan Bantuan Stimulan Perumahan Swadaya (BSPS)," *InfoTekJar J. Nas. Inform. ...*, vol. 1, 2021, [Online]. Available: <https://www.jurnal.uis.ac.id/index.php/infotekjar/article/view/4093>