JURNAL FASILKOM
Volume 13 No. 3 | Desember 2023: 504-510

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Prediksi Tingkat Kelulusan Menggunakan K-Means Pada Program Studi Informatika Unismuh Makassar

Fahrim Irhamna Rahman[1*], Siti Mujadilah[2], Titin Wahyuni[3], Lukman Anas[4]
[1,2,3,4]Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar
[1]fachrim141020@unismuh.ac.id*, [2]mujadilah@student.unismuh.ac.id, [3]titinwh@gmail.com, [4]lukmananas@unismuh.ac.id

Abstract

The Informatic Program at Muhammadiyah University Makassar requires several factors to be met for students to graduate, namely a Cumulative Grade Point Average (CGPA) of at least 2.0, an accumulation of 156 Credit Units, and a maximum study period of 14 semesters or 7 years. Often, the Study Program itself is late in detecting at the outset which students are struggling in their courses and may graduate late. Therefore, this research aims to develop a predictive model for the graduation rate of students in the Informatic Program at Muhammadiyah University Makassar. This method is used to cluster students based on attributes such as total credits taken, semester Grade Point Average (GPA), and overall Cumulative Grade Point Average (CGPA). The clustering aims to identify patterns and characteristics of student graduation. Data from several semesters is collected and preprocessed, including data normalization and transformation. The research steps involve data preprocessing, cluster labeling, distance calculation to cluster centers, and result analysis. The analysis shows that the K-means method can generate student clusters with varying graduation rate patterns. The formed clusters can be interpreted as groups of students with potential for timely graduation or groups needing more attention to achieve on-time graduation. Empirical validation is performed by comparing K-means prediction results with actual graduation data. Accuracy measurement involves calculating the percentage of similarity between predictions and actual data. Empirical validation results demonstrate the accuracy level, which can serve as a benchmark for assessing the performance of this prediction model. This study aims to provide deeper insights into factors influencing student graduation and potentially support decision-making at the academic level.

Keywords: Graduation Prediction, Data Mining, K-Means, Analysis, Clustering, Empirical Validation.

Abstrak

Program Studi Informatika Universitas Muhammadiyah Makassar memerlukan beberapa faktor yang harus terpenuhi agar mahasiswa dapat lulus yaitu dari Indeks Prestasi Kumulatif (IPK) dengan minimal IPK keseluruhan yaitu 2.0, Satuan Kredit Semester (SKS) yang harus dicapai yaitu 156 SKS, dan Lama studi maksimal 14 semester atau 7 tahun, dimana sering sekali pihak Program Studi sendiri telat dalam mendeteksi di awal bahwa mahasiswa-mahasiswa mana saja yang tergolong bermasalah dalam perkuliahan dan telat lulus. Maka dengan ini penelitian ini bertujuan untuk mengembangkan model prediksi tingkat kelulusan mahasiswa pada program studi Informatika di Universitas Muhammadiyah Makassar menggunakan metode data mining K-means. Metode ini digunakan untuk mengelompokkan mahasiswa berdasarkan atribut seperti total SKS yang diambil, indeks prestasi semester (IPS), serta IPK keseluruhan. Pengelompokan ini bertujuan untuk mengidentifikasi pola dan karakteristik kelulusan mahasiswa. Langkah-langkah penelitian meliputi pemrosesan data, pelabelan kluster, perhitungan jarak data ke pusat kluster, serta analisis hasil. Hasil analisis menunjukkan bahwa metode K-means dapat menghasilkan kelompok mahasiswa dengan pola tingkat kelulusan yang bervariasi. Kluster yang terbentuk dapat diartikan sebagai kelompok mahasiswa yang memiliki potensi kelulusan tepat waktu atau kelompok mahasiswa yang memerlukan perhatian lebih untuk mencapai kelulusan tepat waktu. Validasi empiris dilakukan dengan membandingkan hasil prediksi K-means dengan data aktual kelulusan. Pengukuran akurasi dilakukan dengan menghitung persentase kemiripan antara prediksi dan data aktual. Hasil validasi empiris menunjukkan tingkat akurasi yang dapat digunakan sebagai acuan dalam menilai kinerja model prediksi ini. Penelitian ini diharapkan dapat memberikan pandangan lebih mendalam terhadap faktor-faktor yang mempengaruhi kelulusan mahasiswa dan berpotensi mendukung pengambilan keputusan di tingkat akademik.

Kata Kunci: prediksi kelulusan, data mining, k-means, analisis, kluster, validasi empiris

.

1. Pendahuluan

Kemajuan pesat dalam teknologi informasi pada saat ini menandakan bahwa tingkat ketepatan sebuah Informasi atau data memiliki peran yang sangat signifikan dalam kehidupan sehari-hari, di mana setiap bit data menjadi sangat penting untuk membimbing keputusan dalam berbagai situasi. Oleh karena itu, penyediaan informasi menjadi alat yang dapat dianalisis dan diringkas untuk membentuk pengetahuan berharga dari data, yang nantinya dapat bermanfaat saat proses pengambilan keputusan dilakukan. [1]

Kelulusan mahasiswa yang tepat waktu membawa banyak manfaat tidak hanya bagi mahasiswa, tetapi juga bagi universitas itu sendiri. Karena kelulusan merupakan penilaian dalam proses akreditasi perguruan tinggi, maka dengan lulus nya mahasiswa

P-ISSN : 2089-3353
E-ISSN : 2808-9162

tepat waktu tentu akan membantu dalam penilaian akreditasi perguruan tinggi. Tidak selalu dapat memprediksi kelulusan mahasiswa sejak permulaan, ada potensi terjadinya keterlambatan dalam memproses kelulusan. Untuk menghadapi tantangan tersebut, diperlukan penerapan teknik yang memungkinkan perkiraan kelulusan mahasiswa. Salah satu cara yang umum digunakan untuk hal ini adalah pemanfaatan teknik penambangan data. Dan cara yang diterapkan yaitu untuk prediksi kelulusan mahasiswa adalah metode K-Means.

Prediksi merupakan suatu proses sistematis untuk mengevaluasi kemungkinan terjadinya suatu kejadian Pada masa yang akan datang, berdasarkan informasi dari masa lalu dan saat ini, dapat diproyeksikan kejadian mendatang sehingga kesalahan atau perbedaan antara hasil yang diharapkan dan kenyataan dapat diminimalkan. Walaupun prediksi tidak selalu memberikan jawaban yang pasti mengenai masa yang akan datang., upaya dilakukan untuk mencari jawaban yang seakurat mungkin terhadap perkembangan yang akan terjadi [2]. Siswa mengetahui bagaimana menghadapi kecerdasannya dan secara sistematis mencari masalah, yang kemudian diterapkan dalam kehidupan sehari-hari untuk bersaing dalam kehidupan profesional. Kelulusan mahasiswa merupakan isu penting untuk dipertimbangkan karena tingkat kelulusan mempengaruhi evaluasi dewan dan mempengaruhi status akreditasi program.

Pada Program Studi Informatika Universitas Muhammadiyah Makassar memerlukan beberapa faktor yang harus terpenuhi agar mahasiswa dapat lulus yaitu dari Indeks Prestasi Kumulatif (IPK) dengan minimal IPK keseluruhan yaitu 2.0, Satuan Kredit Semester (SKS) yang harus dicapai yaitu 156 SKS, dan Lama studi maksimal 14 semester atau 7 tahun. Menurut kutipan dari David Hand, Heikki Mannila, dan Padhraic Smyth yang disitir dalam penelitian oleh Widodo dan kawan-kawannya pada tahun 2013, data mining adalah proses analisis data, terutama data yang berskala besar, dengan tujuan menemukan hubungan yang jelas dan menyimpulkan informasi yang sebelumnya tidak diketahui. Proses ini dilakukan dengan cara yang dapat dimengerti dan memberikan manfaat yang signifikan bagi pemilik data. Dalam penelitian ini penulis akan membuatkan prediksi tingkat kelulusan mahasiswa menggunakan data mining dengan metode K-means pada program studi informatika Universitas Muhammadiyah Makassar yang akan mempercepat prediksi kelulusan mahasiswa agar mahasiswa dapat membuat strategi atau merencanakan sesuatu untuk dapat melihat apa yang harus dilakukan suatu mahasiswa agar ke depannya bisa lulus tepat waktu dan memperkirakan tidak terjadinya Drop Out (DO) pada suatu mahasiswa.

Penelitian yang berjudul "Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik Dan Non Akademik Menggunakan Metode K-Means (Studi Kasus: Universitas Catur Insan Cendekia)" Pada penelitian ini mampu memberikan prediksi ketepatan kelulusan mahasiswa dengan metode K-means dimana dalam sistemnya terdapat hasil perhitungan prediksi kelulusan dan dapat memberitahukan jumlah lulus berdasarkan tahun angkatan.

Penelitian yang berjudul "Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa" Pengimplementasian algoritma Neural Network ini memberikan hasil yang diinginkan yaitu dapat memberikan prediksi Dengan temuan yang menunjukkan bahwa keterlambatan kelulusan mahasiswa kemungkinan disebabkan oleh adanya data latihan (data training)., dan untuk algoritma Neural Network ini akan berpengaruh keakuratan nya jika jumlah input seperti nodes, training cycle, dll [3].

Penelitian yang berjudul "Implementasi Data mining Dengan Algoritma Apriori Untuk Memprediksi Tingkat Kelulusan Mahasiswa" Algoritma Apriori mampu memprediksi tingkat kelulusan mahasiswa dimana aplikasi ini hanya akan berjalan untuk mahasiswa semester 5 keatas dan juga aplikasi ini dapat memberitahukan kepada mahasiswa matakuliah mana yang menjadi keunggulan dari mahasiswa tersebut (Kurniawan, 2018) [4].

Penelitian yang berjudul "Pembuatan aplikasi *data mining* untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma apriori" pengimplementasian algoritma apriori ini memberikan hasil bahwa penelitian Menampilkan informasi tingkat kelulusan dapat dimungkinkan melalui penerapan aplikasi Data Mining. Informasi yang disajikan mencakup Nilai dukungan (support) dan keyakinan (confidence) berkaitan dengan hubungan antara tingkat kelulusan dan data induk mahasiswa. Semakin tinggi nilai confidence dan support, semakin kuat keterkaitan antar atribut. Proses penambangan data pada data induk mahasiswa mencakup asal sekolah dan informasi kota tempat mahasiswa berasal. Hasil dari proses Data Mining ini dapat dijadikan pertimbangan dalam proses pengambilan keputusan. [5].

Penelitian yang berjudul "Implementasi *Data mining* Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5" Temuan dari penelitian ini menunjukkan bahwa algoritma C4.5 terbukti lebih akurat daripada evaluasi yang dilakukan oleh analis mahasiswa dalam mengukur tingkat ketepatan waktu penyelesaian studi. Pernyataan ini didukung oleh hasil penelitian yang menunjukkan bahwa algoritma C4.5 mampu secara efektif menganalisis tingkat keakuratan prediksi waktu penyelesaian studi mahasiswa[6].

Penelitian yang berjudul "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA" dengan pengimplementasian metode Naïve Bayes pada prediksi kelulusan diketahui bahwa Dalam penelitian ini, terdapat tiga model, dan model terunggul adalah model 3. Model ini terdiri dari atribut jenis kelamin dan indeks prestasi semester 1 hingga

P-ISSN : 2089-3353
E-ISSN : 2808-9162

semester 4, dengan tingkat keakuratan mencapai 80,19%, ketepatan presisi 92,75%, daya ingat recall 80,26%, dan ukuran F-Measure 86,05%. Untuk meningkatkan ketepatan prediksi, disarankan untuk menambah jumlah data mahasiswa agar hasil prediksi menjadi lebih akurat. Selain itu, penambahan atribut yang relevan dengan prediksi kelulusan mahasiswa, seperti alamat, lulusan sekolah, dan faktor-faktor lainnya, juga dapat memberikan kontribusi positif, dapat memperkaya model dan meningkatkan keakuratannya[7].

Penelitian yang berjudul "Merancang sistem prediksi ketepatan kelulusan mahasiswa berdasarkan data akademik dan non akademik menggunakan metode k-means". Dengan menerapkan metode K-means, dapat dilakukan prediksi terkait kelulusan mahasiswa dengan kategori "tepat waktu" dan "terlambat" di Universitas Catur Insan Cendekia. Selain itu, metode ini juga berguna untuk mengidentifikasi jumlah mahasiswa yang berhasil lulus berdasarkan tahun penerimaan mereka di institusi tersebut. [3]

Berdasarkan penelitian-penelitian di atas maka salah satu fungsi penerapan data mining dengan metode K-means dalam perguruan tinggi untuk memprediksi tingkat kelulusan mahasiswa dalam hal ini diambil dari beberapa universitas dan institut. Banyak penelitian yang telah dilakukan berkaitan dengan metode K-means tersebut. Kebanyakan dari penelitian tersebut mengarah kepada informasi mengenai data statistika sehingga pembahasannya lebih ditekankan mengenai nilai hasil prediksi untuk suatu metode K-means yang baru dalam menentukan prediksi kelulusan mahasiswa tepat waktu[7].

Pengembangan metode K-means dianggap cocok digunakan untuk klasifikasi prediksi kelulusan mahasiswa tepat waktu. maka penelitian ini akan di implementasikan ke metode K-means tersebut dan dicari hasil akurasi dengan menggunakan validasi empiris dan menggunakan bahasa pemrograman python serta hasil pengelompokan dari tingkat asurasi tersebut, sehingga dapat menggali informasi mengenai hasil data mining pada prediksi hasil tingkat kelulusan mahasiswa lulusan tepat waktu, dan menghasilkan prediksi data terhadap mahasiswa yang aktif yang menjalankan studi di Universitas Muhammadiyah Makassar Fakultas Teknik Prodi Informatika. Informasi pengelompokan ini bertujuan untuk mengurangi set fungsi Maksud dari proses pengelompokan biasanya adalah untuk mengurangi perbedaan dalam setiap kelompok dan meningkatkan perbedaan antara satu kelompok dengan kelompok lainnya.[8].

2. Metode Penelitian

2.1 Data Mining

Metode utama yang diterapkan dalam topik penelitian ini adalah *Data Mining* yang merupakan kegiatan analisis data, terutama pada data yang berskala besar, dengan tujuan untuk menemukan keterkaitan yang nyata dan menyimpulkan informasi yang sebelumnya tidak teridentifikasi. Proses ini dilakukan dengan metode yang dapat dipahami saat ini dan memberikan keuntungan bagi pemilik data [9]. Di awal data yang dikumpulkan merupakan data mahasiswa Program Studi Informatika Universitas Muhammadiyyah Makassar dalam 3 tahun terakhir dan akan digunakan sebagai data sampel training untuk kemudian ditambang atau Dimana proses *Data Mining* akan dijalankan. Data Mahasiswa ini memiliki atribut-atribut parameter yang akan digunakan dalam proses analisis *Data Mining* dalam menemukan keterkaitan yang penting, pola, dan arah kecenderungan dengan menyelidiki kumpulan besar data yang tersimpan, mengaplikasikan teknik pengenalan pola seperti metode Statistik dan Matematika.[10].

Kerangka kerja data mining terdiri dari tiga tahap, yaitu Pengumpulan data (data collection), pengubah data (data transformation), dan analisis data (data analysis) merupakan bagian integral dari proses. Proses ini dimulai dengan tahap pra-pemrosesan, yang melibatkan pengumpulan data untuk menghasilkan data mentah yang diperlukan dalam aktivitas penambangan data. Langkah selanjutnya adalah transformasi data, Di dalam tahap ini, data mahasiswa prodi informatika beserta atribut-atributnya akan mengalami transformasi menjadi format yang dapat diolah oleh instrumen penambangan data, seperti melalui proses filtrasi atau agregasi. Hasil transformasi data mahasiswa-mahasiswa dan atributnya iini kemudian digunakan dalam proses analisis data untuk menghasilkan pengetahuan, dengan memanfaatkan model *Machine Learning* dalam bentuk Klustering seperti K-Means yang akan digunakan di dalam penelitian ini.

2.2 K-means dan Klustering

Algoritma K-means yang digunakan dalam penelitian ini termasuk dalam kategori algoritma distribusi karena berasal dari penentuan jumlah kelompok awal dari atribut data mahasiswa setelah pra-pemrosesan data awal dengan menetapkan mean awal. Seperti yang kita ketahui Metode pengelompokan K-means clustering merupakan pendekatan non-hierarkis yang mengelompokkan data ke dalam satu atau lebih kluster atau kelompok. Yang nantinya setelah diproses dengan K-Means display data mahasiswa atributnya yang menunjukkan kesamaan karakteristik dikelompokkan bersama dalam satu kluster, sementara data dengan karakteristik yang berbeda ditempatkan dalam kluster lainnya. Sehingga, data yang tergabung dalam satu kluster memiliki variasi yang minim [11].

Berikut beberapa langkah yang dilakukan untuk mengelompokkan data mahasiswa menggunakan K-means Langkah-langkahnya adalah sebagai berikut:

P-ISSN : 2089-3353
E-ISSN : 2808-9162

a. Mengidentifikasi jumlah cluster k yang diinginkan berdasarkan atribut data mahasiswa di atas.
b. Menetapkan pusat cluster secara acak dengan inisialisasi k.
c. Menyelaraskan semua data ke dalam cluster terdekat. Kedekatan data dengan suatu cluster diukur melalui jarak antara data tersebut dan pusat cluster. Data akan diatribusikan ke dalam cluster berdasarkan jarak terpendek, dengan menggunakan perhitungan jarak Euclidean sesuai dengan rumus yang ditunjukkan dalam Persamaan di bawah.

$$d(i,j) = \sqrt{\sum_{l=1}^{p}(x_{il} - x_{jl})^2} \quad (1)$$

Keterangan:

d(i,j) = jarak antara data ke-i dan pusat kluster ke-j;

p = jumlah atribut;

l = indeks yang berjalan dari 1 hingga p;

x_il = nilai data ke-i pada atribut ke-l;

x_jl = nilai pusat kluster ke-j pada atribut ke-l;

d. Lakukan perhitungan ulang pusat kluster dengan mempertimbangkan keanggotaan yang baru.

e. Kalkulasikan kembali jarak data dari pusat kluster yang baru. Jika tidak ada perubahan anggota kluster dari satu kluster ke kluster lainnya, langkah pengelompokan dihentikan. Namun, jika terdapat anggota cluster yang mengalami perpindahan antar cluster, kembali ke langkah 3 dan ulangi proses tersebut hingga tidak ada lagi anggota cluster yang mengalami perpindahan [12].

Pada awalnya, Algoritma K-Means mengambil sebagian komponen dari populasi sebagai titik pusat cluster awal. Pada tahap ini, pusat cluster dipilih secara acak dari seluruh data populasi. Setelah itu, K-Means memeriksa setiap komponen untuk menentukan kluster yang paling sesuai, dengan mempertimbangkan jarak minimum antara komponen tersebut dan pusat cluster yang telah ditentukan. Posisi pusat cluster kemudian dihitung ulang hingga semua komponen data terklasifikasi ke dalam kluster yang paling sesuai. Akhirnya, posisi kluster baru akan terbentuk.[13].

Jadi implementasi secara spesifik sebagaimana yang dipaparkan di atas yakni pada proses awal data yang akan digunakan ada 4 sebagai atribut yaitu Sks, Ipk, Lama studi dan matakuliah wajib atau tidak. Memiliki kluster 2 kluster yaitu kluster 1 lulus tepat waktu dan kluster 2 lulus tidak tepat waktu. Proses selanjutnya menentukan jumlah kluster (k) dimana data akan dibagi ke dalam 2 kluster dan akan terbentuk 2 pusat kluster. Dimana dalam menentukan pusat awal kluster dilakukan dengan cara acak atau menggunakan total sks, nilai ipk, lama studi dan matkul wajib atau tidak wajib mahasiswa.

Selanjutnya melakukan penghitungan jarak dari masing-masing objek dari pusat kluster, dalam penghitungan jarak ini dihitung sesuai banyaknya kluster awal. Perhitungan jarak terdekat setiap objek dihitung sebanyak jumlah sample kluster berdasarkan atribut yang telah ditempuh dengan titik pusatnya dan dilakukan untuk semua mahasiswa.

Selanjutnya pengelompokan objek berdasarkan jarak terdekat atau minimum, dimana dalam keputusan ini hasil yang diperoleh sebelumnya semakin minimum nilai jaraknya maka semakin besar kemiripan terhadap suatu kluster. Sehingga jarak terdekat itulah yang menentukan setiap mahasiswa termasuk ke dalam kluster tersebut.

Setelah menghasilkan pusat kluster baru dan hasil nilai pusat kluster baru ternyata memiliki nilai berbeda maka perlu menghitung kembali jarak dan mengelompokkan data sehingga perlu pusat kluster sebelumnya sama dengan hasil pusat kluster dan jika hasil dari pusat kluster lama dan pusat kluster baru sudah sama atau sudah tidak memiliki selisih nilai maka muncullah hasil akhir sebuah kluster selanjutnya telah selesai.

Jadi dari pemaparan Langkah-langkah di atas sesuai dengan pengertiannya bahwa Klustering merujuk pada proses pengelompokan objek atau data ke dalam kelas-kelas yang menunjukkan kesamaan tertentu. Setiap kelas dalam klustering disebut sebagai cluster, yang merupakan satu set data yang serupa di dalamnya dan berbeda dari cluster lainnya [14]. Teknik klustering dapat menganalisis data dengan tujuan mengelompokkan data serupa ke dalam satu cluster berdasarkan kemiripan. Fungsi utama klustering adalah mengidentifikasi pola dalam data, yang mungkin tidak mudah ditemukan melalui analisis manual.

3. Hasil dan Pembahasan

Dalam bab ini, akan dipaparkan hasil dan pembahasan terkait prediksi kelulusan mahasiswa dengan penerapan metode K-means. Hasil eksperimen mencakup proses pengelompokan data mahasisw berdasarkan atribut Nama, masa studi, total SKS, IPK dan Nilai IPS semester 1 hingga Nilai IPS semester 7 menggunakan algoritma K-means.

3.1. Deskripsi Dataset

Data yang di olah dari penelitian ini merupakan data mahasiswa informatika fakultas teknik universitas muhammadiyah Makassar angkatan 2018 sampai dengan 2021. Dataset mahasiswa terdiri dari atribut data mata kuliah dan data KRS Mahasiswa (kartu rencana studi).

No	Nama	masa_studi	total_sks	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	TOTAL_IPK
1	MUNAWAR PUJA DEWANTARA	10	31	2,50	1,77	0,00	0,00	0,00	0,00	0,00	2,19
2	ADIWIRATMAN	10	31	2,64	2,43	2,36	2,00	2,00	3,29	3,39	2,69
3	MUH. ICHZAN	10	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
4	MUH. SYAHRIAN SURYA	10	73	3,09	2,38	2,30	1,80	0,00	0,00	0,00	2,49
5	ANUGRAH DANIALDY ANWAR	10	74	0,00	1,79	2,09	2,24	1,67	2,33	0,00	2,04
6	RESKI ABBAS	10	156	3,36	3,43	3,53	3,90	3,91	4,00	3,30	3,66
	.....										
46	ZULFIRMAN	10	104	3,73	3,35	3,67	0,00	3,57	2,84	0,00	3,31

Gambar 1. Data Mahasiswa Alumni 2018

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Atribut yang dijadikan dalam prediksi kelulusan mahasiswa yang menggunakan metode pengelompokan K-means ini di sederhanakan menjadi atribut yang terdiri dari Nama, IPK, Nilai Induk Prestasi Semester (IPS) 1 sampai Nilai Induk Prestasi Semester (IPS) 7, Total SKS mahasiswa dan masa studi.

Dataset pada gambar 1 berisikan data mahasiswa angkatan 2018 sebagai data training (latih) untuk metode K-means. Dataset pada gambar 2 berisikan data mahasiswa angkatan 2020-2021 sebagai data uji yang berisikan atribut data yang akan diolah pada klustering dengan menggunakan metode K-means.

No	Nama	masa_studi	total_sks	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	TOTAL_IPK
1	ASWAR	6	26	3,26	3,71	0,00	0,00	0,00	0,00	0,00	3,38
2	Sulastri	6	110	3,55	3,91	3,40	3,14	3,27	3,00	3,00	3,46
3	FATIMAH AZZAHRA . Z	6	110	3,41	3,83	3,51	3,33	3,54	3,00	3,00	3,53
4	JIHAN FAHIRA	6	22	3,36	0,00	0,00	0,00	0,00	0,00	0,00	3,36
5	Rifky Darmawan Syah	6	108	3,77	3,65	3,45	3,10	3,35	3,00	3,00	3,47
6	Khairun Nisha	6	110	3,77	4,00	3,58	3,39	3,43	3,00	3,00	3,64
	.....										
263	MUH. NUR IKHSAN. R	4	42	3,36	3,50	3,17	3,00	3,00	3,00	3,00	3,38
264	Muhammad Erdian	4	55	3,10	3,08	3,03	3,00	3,00	3,00	3,00	3,07

Gambar 2. Dataset Angkatan 2020-2021

3.2 Metode K-means

Dalam perhitungan metode k-means lustering ini digunakan untuk mendapatkan hasil yang di inginkan yaitu memprediksi kelulusan mahassiswa dengan menggunakan atribut pada gambar dimana untuk menentukan nilai kelompok atau nilai kluster nya yaitu dengan sesuai kebutuhan.

No	Nama	total_sks	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPK
1	MUNAWAR PUJA DEWANTARA	31	2,50	1,77	0,00	0,00	0,00	0,00	0,00	2,19
2	ADIWIRATMAN	122	2,64	2,43	2,36	2,00	2,00	3,29	3,39	2,69
3	MUH. ICHZAN	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
4	MUH. SYAHRIAN SURYA	73	3,09	2,38	2,30	1,80	0,00	0,00	0,00	2,49
5	ANUGRAH DANIALDY ANWAR	74	0,00	1,79	2,09	2,24	1,67	2,33	0,00	2,04
6	RESKI ABBAS	156	3,36	3,43	3,53	3,90	3,91	4,00	3,30	3,66
7	MUH. IRHAM PATTA	18	1,71	2,00	0,00	0,00	0,00	0,00	0,00	1,77
8	NURWAHYU	27	2,30	2,00	2,00	0,00	0,00	0,00	0,00	2,22
9	MASNIA	145	3,36	3,70	3,50	3,71	3,00	3,60	3,29	3,46
10	RESKI AWALIA S	150	3,68	3,70	3,83	3,81	3,58	4,00	3,46	3,72

Gambar 3. Sampel Data Mahasiswa Informatika

Data sampel yang diguanakan dalam perhitungan metode K-means clustering yaitu sebanyak 10 data mahasiswa informatika fakultas teknik unismuh makassar dengan atribut IPS,IPK dan SKS seperti yang ada pada gambar 3 Sample data mahasiswa informatika.

a. Jarak antara data siswa pertama dengan pusat kluster pertama.

$$c_1\sqrt{(31-122)^2+(2{,}50-2{,}64)^2+(1{,}77-2{,}43)^2+(0{,}00-2{,}36)^2+(0{,}00-2{,}00)^2+(0{,}00-2{,}00)^2+(0{,}00-3{,}29)^2+(0{,}00-3{,}39)^2+(2{,}19-2{,}69)^2}=91{,}20$$

b. Jarak antara data siswa pertama dengan pusat kluster kedua.

$$c_2\sqrt{(31-73)^2+(2{,}50-3{,}09)^2+(1{,}77-2{,}38)^2+(0{,}00-2{,}30)^2+(0{,}00-1{,}80)^2+(0{,}00-0{,}00)^2+(0{,}00-0{,}00)^2+(0{,}00-0{,}00)^2+(2{,}19-2{,}49)^2}=42{,}11$$

c. Jarak antara data siswa pertama dengan pusat kluster ketiga.

$$c_3\sqrt{(31-156)^2+(2{,}50-3{,}36)^2+(1{,}77-3{,}43)^2+(0{,}00-3{,}53)^2+(0{,}00-3{,}90)^2+(0{,}00-3{,}91)^2+(0{,}00-4{,}00)^2+(0{,}00-3{,}30)^2+(2{,}19-3{,}66)^2}=125{,}30$$

Gambar 4. Perhitungan Jarak Menggunakan Rumus *Euclidean Distance Space*

Pada gambar 4 merupakan proses perhitungan jarak menggunakan rumus *Euclidean Distance Space* untuk menentukan jarak setiap data yang ada terhadap setiap pusat kluster. Berikut ini merupakan hasil dari perhitungan keseluruhan data mahasiswa terhadap tiap pusat kluster awal diatas hasil untuk perhitungan iterasi 1

data i	C1	C2	C3
1	91,20	42,11	125,30
2	0,00	49,27	34,17
3	122,23	73,20	156,34
4	49,27	1,80	83,30
5	48,21	47,20	82,24
6	34,17	83,30	0,00
7	104,18	132,30	138,28
8	95,16	46,05	129,25
9	23,17	72,28	11,05
10	28,22	77,33	6,03

Gambar 5 Hasil Perhitungan Iterasi 1

Pada Gambar 5 merupakan perhitungan iterasi pertama yang digunakan pada data sample sebanyak 10 dataset. Dimana perhitungan jarak menggunakan *Euclidean Distance Space* sudah dapat terlihat mulai data pertama sampai data ke 10 dimana untuk data pertama menempati C2 dengan nilai atau jarak terkecil sebesar 42,11 untuk data ke dua menempati C1 untuk hasil nilai jarak terkecil nya. Dari perhitungan untuk iterasi pertama ini masih akan ada perubahan atau perpindahan centeroid. Perhitungan iterasi akan berhenti saat tidak ada nya perpindahan centeroid.

data i	C1	C2	C3	NILAI KLUSTER
1	91,20	42,11	125,30	C2
2	0,00	49,27	34,17	C1
3	122,23	73,20	156,34	C2
4	49,27	1,80	83,30	C2
5	48,21	47,20	82,24	C2
6	34,17	83,30	0,00	C3
7	104,18	132,30	138,28	C1
8	95,16	46,05	129,25	C2
9	23,17	72,28	11,05	C3
10	28,22	77,33	6,03	C3

data i	C1	C2	C3	NILAI CLUSTER
1	39,12	10,18	119,64	C2
2	52,09	81,17	28,52	C3
3	70,17	41,13	150,68	C2
4	4,29	32,09	77,64	C1
5	5,21	33,16	76,59	C1
6	86,21	115,26	5,69	C3
7	52,09	23,06	132,62	C2
8	43,10	14,08	123,58	C2
9	75,19	104,25	5,37	C3
10	80,23	109,29	0,51	C3

data i	C1	C2	C3	NILAI KLUSTER
1	42,64	12,06	112,52	C2
2	48,69	103,18	21,39	C3
3	73,66	19,19	143,56	C2
4	2,21	54,10	70,53	C1
5	2,21	55,17	69,49	C1
6	82,75	137,28	12,80	C3
7	55,60	1,27	125,50	C2
8	46,57	8,21	116,47	C2
9	71,73	126,27	1,85	C3
10	76,78	131,31	6,84	C3

Gambar 6. Hasil Perhitungan K-means

Terdapat ilustrasi pada gambar 6 yang mencerminkan hasil dari perhitungan K-means dengan menggunakan metode tersebut, dengan jumlah iterasi sebanyak 3 ini memiliki nilai kluster yang sama dengan nilai kluster pada iterasi ke 2 sebelumnya atau nilai kluster untuk

P-ISSN : 2089-3353
E-ISSN : 2808-9162

setiap data sudah sama, bisa dilihat juga dari tanda berwarna kuning yang penempatan warnanya sudah sama tidak ada perubahan, maka perhitungan iterasi berhenti karena jika melanjutkan perhitungan nilai centeroid nya sudah akan sama dari nilai centeroid baru sebelumnya begitu pun dengan nilai jarak clusternya.

Index	Nama	Kluster	label
1	ASWAR	4	Tidak Tepat
2	Sulastri	9	Lulus Tepat
3	FATIMAH AZZAHRA . Z	9	Lulus Tepat
4	JIHAN FAHIRA	10	Tidak Tepat
5	Rifky Darmawan Syah	9	Lulus Tepat
6	Khairun Nisha	9	Lulus Tepat
.....			
261	Andi Fadel Aditya	7	Tidak Tepat
262	MUHAMMADKHADAFI	1	Tidak Tepat
263	MUH. NUR IKHSAN. R	7	Tidak Tepat
264	Muhammad Erdian	1	Tidak Tepat

Gambar 7. Hasil Akhir

Hasil akhir dapat di lihat dari gambar 7 dimana dengan data sampel sebanyak 10 dataset mahasiswa dengan menggunakan metode K-means dimana untuk mahasiswa yang Lulus Tidak Tepat dengan nilai yang sangat rendah dengan kemungkinan membutuhkan 4 atau lebih tambahan semester lagi ada sebanyak 4 mahasiswa, mahasiswa dengan Lulus Tidak Tepat dengan kemungkinan membutuhkan 1 atau 2 semester tambahan untuk lulus tepat ada sebanyak 2 mahasiswa dan terakhir untuk mahasiswa yang Lulus Tepat Waktu ada sebanyak 4 mahasiswa.

3.3. Proses Klustering K-means Dengan Python

Proses kustering K-means dengan python menggunakan dataset angkatan 2018 sebanyak 47 mahasiswa sebagai dataset latih. Dimana klustering pada penelitian ini dengan melihat nilai dari rata-rata keseluruhan kluster sebelumnya.

	Nama	Kluster	Label
0	MUNAWAR PUJA DEWANTARA	8	Tidak Tepat
1	ADIWIRATMAN	5	Tidak Tepat
2	MUH. ICHZAN	0	Tidak Tepat
3	MUH. SYAHRIAN SURYA	2	Tidak Tepat
4	ANUGRAH DANIALDY ANWAR	2	Tidak Tepat
5	RESKI ABBAS	1	Lulus Tepat
6	MUH. IRHAM PATTA	8	Tidak Tepat
7	NURWAHYU	8	Tidak Tepat
.....			
38	MUH. FAHRIL K	0	Tidak Tepat
39	ST. NURWAHIDA ARIF	1	Lulus Tepat
40	ALFITRA HASKAR	3	Tidak Tepat
41	HARPANJI PIAN	6	Tidak Tepat
42	M. RANDI HARIADI	10	Tidak Tepat
43	SUHUDI IBRAHIM	0	Tidak Tepat
44	MUH. KHAEDAR ASYARI K	0	Tidak Tepat
45	ZULFIRMAN	10	Tidak Tepat
46	MUH. GALANK ALBANISART	5	Tidak Tepat

Gambar 8. Hasil Akhir K-means Angkatan 2018

Hasil akhir yang diperlihatkan pada gambar 8 didapatkan dari hasil Klustering sampai akhir dan menentukan nama label yang tepat untuk setiap mahasiswa. Dimana data di atas merupakan dataset tahun angkatan 2018 yang merupakan dataset alumni informatika unismuh makassar, digunakan sebagai dataset untuk menghitung keakurasian metode ini, dengan tingkat keakurasian sebesar 100% dengan menggunakan nilai Kluster sebanyak 11. Dengan jumlah iterasi sebanyak 3 iterasi.

No	Nama	TOTAL_IPK	CLUSTER	LABEL
1	MUNAWAR PUJA DEWANTARA	2,19	C2	TIDAK TEPAT
2	ADIWIRATMAN	2,69	C3	LULUS TEPAT
3	MUH. ICHZAN	0,00	C2	TIDAK TEPAT
4	MUH. SYAHRIAN SURYA	2,49	C1	TIDAK TEPAT
5	ANUGRAH DANIALDY ANWAR	2,04	C1	TIDAK TEPAT
6	RESKI ABBAS	3,66	C3	LULUS TEPAT
7	MUH. IRHAM PATTA	1,77	C2	TIDAK TEPAT
8	NURWAHYU	2,22	C2	TIDAK TEPAT
9	MASNIA	3,46	C3	LULUS TEPAT
10	RESKI AWALIA S	3,72	C3	LULUS TEPAT

Gambar 9. Hasil Akhir K-means Angkatan 2018

Hasil akhir yang ditunjukkan pada gambar 9 merupakan hasil prediksi dari metode k-means dengan jumlah nilai kluster sebanyak 11 dan untuk iterasi ada sebanyak 5 kali. Untuk metode ini ke akurasian dalam prediksi mahasiswa ini belum tentu 100% akurat, dalam prediksi ini hanya memprediksi kemungkinan lulus tepat dan tidaknya mahasiswa yang di lihat dari beberapa atribut menggunakan metode K-means. Penggunaan nilai kluster yaitu 11 kluster atau kelompok digunakan karena mengikuti nilai kluster dari data training sebelumnya yaitu data dari angkatan 2018 dengan keakurasian sebesar 100%, dan angkatan 2019 dengan akurasi 89 % dengan harapan ke akurasiannya juga berlaku untuk dataset 2020 sampai 2021.

4. Kesimpulan

Kelompok Mahasiswa Tidak Lulus Kelompok ini terdiri dari mahasiswa-mahasiswa yang cenderung tidak menyelesaikan studi tepat waktu dan memiliki performa akademik yang kurang baik. Mahasiswa dalam kelompok ini memiliki rata-rata masa studi yang lebih lama dan nilai IPS yang lebih rendah dengan kemungkinan untuk menambah semester tambahan. Dari penelitian yang di lakukan menggunakan data mahasiswa informatika angkatan 2019 sampai 2021 ada sebanyak 103 yang lulus tepat waktu berdasarkan penggunaan metode K-means ini, dimana dalam prediksi kelulusan ini menggunakan nilai kluster sebanyak 11 dikarenakan untuk data training sebelumnya yang dilakukan pada data alumni 2018 itu memiliki ke akurasian sebesar 100%. Metode K-means dapat digunakan untuk melakukan klasterisasi atau pengelompokan mahasiswa berdasarkan pola akademik mereka, seperti atribut yang digunakan p ada penelitian ini yaitu masa studi, total SKS, IPK dan indeks prestasi semester (IPS) 1 hingga 7.

Daftar Rujukan

[1] N. M. Huda, "APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA (Studi Kasus di Fakultas MIPA Universitas Diponegoro)," *Univ. Stuttgart*, 2010.

[2] M. Ayub, U. K. Maranatha, and A. N. Networks, "Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer Proses Data Mining dalam Sistem Pembelajaran

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Berbantuan Komputer," no. January 2012, 2018.

[3] M. R. A. Fernanda, P. Sokibi, and R. Fahrudin, "Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik Dan Non Akademik Menggunakan Metode K-Means (Studi Kasus : Universitas Catur Insan Cendekia)," *J. Digit*, vol. 11, no. 1, p. 89, 2021, doi: 10.51920/jd.v11i1.182.

[4] R. Ridwan, H. Lubis, and P. Kustanto, "Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 286, 2020, doi: 10.30865/mib.v4i2.2035.

[5] Nurjoko and H. Kurniawan, "Aplikasi Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori di IBI Darmajaya, Bandar Lampung," *J. TIM Darmajaya*, vol. 02, no. 01, pp. 79–93, 2016.

[6] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.

[7] D. Anugrah Putra and M. Kamayani, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA," *Pros. Semin. Nas. Teknoka*, vol. 5, no. 2502, pp. 34–40, 2020, doi: 10.22236/teknoka.v5i.331.

[8] Agusta Yudi, "K-Means – Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, no. Februari, pp. 47–60, 2007.

[9] P. P. Widodo, *Penerapan data mining dengan Matlab*, Cet. 1. Bandung: Bandung : Rekayasa Sains, 2013, 2013.

[10] J. Fürnkranz, "A Comparison of Pruning Methods for Relational Concept Learning," *Knowl. Discov. Databases Pap. from 1994 AAAI Work.*, pp. 371–382, 1994, [Online]. Available: http://www.ofai.at/cgi-bin/tr-online?number+94-03

[11] E. D. Cahyati, D. Herawatie, and E. Wuryanto, "Implementasi K-Means Clustering Untuk Pemetaan Desa Dan Kelurahan Di Kabupaten Bangkalan Berdasarkan Contraceptive Prevalence Rate Dan Tingkat Pendidikan," *Semin. Nas. Mat. dan Apl.*, pp. 341–348, 2017.

[12] J. O. Ong, "Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.

[13] Y. W. Syaifudin and R. A. Irawan, "Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means," *J. Inform. Polinema*, vol. 4, no. 3, p. 189, 2018, doi: 10.33795/jip.v4i3.205.

[14] J. Han, M. Kamber, and J. Pei, "Third Edition : Data Mining Concepts and Techniques," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2012, [Online]. Available: http://library.books24x7.com/toc.aspx?bkid=44712