Pengelompokan Nilai Mahasiswa Untuk Mengoptimalkan Prediksi Kelulusan Tepat Waktu dengan Menggabungkan Algoritma

K-Means dan C4.5 (Studi Kasus : Universitas Potensi Utama)

**Bob Subhan Riza1), Sarjon Defit2)**

1Universitas Potensi Utama, Jl. K.L. Yos Sudarso Km. 6,5 No. 3 A – Tj. Mulia - Medan, Sumatera Utara, 20241

2Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung Nan XX, Padang, Sumatera Barat, 25145

e-mail: 1bob.potensi@gmail.com, 2sarjond@yahoo.co.uk

|  |  |  |
| --- | --- | --- |
| A R T I C L E I N F O  |  | *ABSTRACT* |
| Article History:*Received* *Revised* *Received* *Online*  | *Graduating on time is the dream of every student who studies in tertiary education, there are several factors that cause the passing character is not correct, such as grades, but students are sometimes careless and underestimate these grades, if there are problematic grades it will certainly the student. This study helps the study program predict which students will graduate on time, the method used in this research is K-Means which serves to cluster each student and C4.5 is a decision process that can predict the right student on time, the variables used are: Value, Failure Score, Specialization, PKL, Thesis 1, Thesis 2 and pass using the Rapid Miner software and the results from the processing with the software can predict which students will graduate on time.**Keywords: K-Means, C4.5, graduated on time, student grades* |
| Keywords: *K-Means* *C4.5* *graduated on time* *student grades* |
| Kata Kunci: K-MeansC4.5lulus tepat waktunilai mahasiswa | ABSTRAK |
| Lulus tepat waktu adalah impian setiap mahasiswa yang menempuh pendidikan di perguruan tinggi, ada beberapa faktor yang menyebabkan tidak tepat watunya lulus seperti nilai, akan tetapi mahasiswa kadang-kadang kurang peduli dan anggap sepele terhadap nilai tersebut, jika ada nilai yang bermasalah tentunya akan menghambat mahasiswa, penelitian ini membantu pihak program studi dalam memprediksi mahasiswa yang dapat lulus tepat waktu, metode yang digunakan dalam penelitian ini adalah K-Means yang berfungsi untuk mengklusterkan setiap mahasiswa dan C4.5 adalah proses keputusan yang dapat memprediksi mahasiswa yang tepat waktu, variabel yang digunakan adalah : Nilai, Nilai Gagal, Peminatan, PKL, Skripsi 1, Skripsi 2 dan Lulus dengan menggunakan software RapidMiner dan hasilnya dari proses pengolahan dengan software tersebut dapat memprediksi mahasiswa yang dapat lulus tepat waktu.Kata kunci : K-Means, C4.5, lulus tepat waktu, nilai mahasiswa |
| **Correspondence:**Telephone: +628126013912 E-mail: bob.potensi@gmail.com |
|  |

PENDAHULUAN

Permasalahan yang terjadi saat ini banyak mahasiswa yang tidak menyadari dan kurang perduli terkait dengan nilai bermasalah (nilai gagal). Nilai bermasalah dapat mempengaruhi tidak tepat waktunya mahasiswa tersebut pengambilan matakuliah peminatan, PKL, Skripsi 1 dan Skripsi 2, hal itu berdampak pada kelulusan. Biasanya mahasiswa yang memiliki nilai bermasalah akan lulus tidak tepat waktu. Jadi kriteria-kriteria yang digunakan dalam penelitian ini adalah Nilai, Nilai Gagal. Peminatan, PKL, Skripsi 1 dan Skripsi 2 dan tool yang digunakan dalam penelitian ini adalah RapidMiner, serta menggunakan metode K\_Means yang berfungsi untuk mengkluster mahasiswa untuk mempermudah dalam proses selanjutnya dan C4.5 adalah proses untuk keputusan prediksi mahasiswa yang bisa lulus tepat waktu serta penelitian ini hanya membahas tentang prediksi kelulusan mahasiswa dengan tepat waktu, manfaat yang didapat adalah mahasiswa lebih peduli dengan nilai dan mata kuliah yang akan diambil selanjutnya sehingga dapat lulus tepat waktu, dan program studi dapat mengetahui mahasiswa-mahasiswa yang bermasalah nilainya.

Hendro Priyatman dalam penelitiannya mengungkapkan Prediksi waktu kelulusan mahasiswa bertujuan memberikan sarana untuk mengetahui perkiraan waktu lulus mahasiswa dengan melihat mahasiswa-mahasiswa mana saja yang masuk kedalam suatu kluster tertentu berdasarkan parameter IPK dan kehadiran. Penelitian ini menerapkan algoritma K-Means. Metode ini mempartisi data ke dalam kluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kluster yang sama dan data yang mempunyai karateristik yang berbeda dikelompokkan ke dalam kluster yang lain. Penelitian ini diharapkan dapat membantu pihak kampus maupun mahasiswa untuk memprediksi tingkat kelulusan tepat waktu dan untuk meningkatkan reputasi bagi pihak kampus itu sendiri dan kelulusan tepat waktu bagi mahasiswa supaya kelulusan mereka tidak terlambat, selain itu pihak kampus bisa melakukan hal-hal yang perlu dilakukan apabila mereka diprediksi lulus tidak tepat waktu seperti dengan melakukan bimbingan dan hal lainnya[1].

Gede Aditra Pradnyana meneliti tentang pembagian kelas yakni permasalahan yang terjadi pada saat pembentukan atau pembagian kelas mahasiswa adalah perbedaan kemampuan yang dimiliki oleh mahasiswa di setiap kelasnya yang dapat berdampak pada tidak efektifnya proses pembelajaran yang berlangsung. Pengelompokan mahasiswa dengan kemampuan yang sama merupakan hal yang sangat penting dalam rangka meningkatkan kualitas proses belajar mengajar yang dilakukan. Dengan pengelompokan mahasiswa yang tepat, mereka akan dapat saling membantu dalam proses pembelajaran. Selain itu, membagi kelas mahasiswa sesuai dengan kemampuannya dapat mempermudah tenaga pendidik dalam menentukan metode atau strategi pembelajaran yang sesuai. Penggunaan metode dan strategi pembelajaran yang tepat akan meningkatkan efektifitas proses belajar mengajar. Pada penelitian ini dirancang sebuah metode baru untuk pembagian kelas kuliah mahasiswa dengan mengkombinasikan metode K-Means dan K-Nearest Neighbors (KNN). Metode K-Means digunakan untuk pembagian kelas kuliah mahasiswa berdasarkan komponen penilaian dari mata kuliah prasyaratnya. Adapun fitur yang digunakan dalam pengelompokan adalah nilai tugas, nilai ujian tengah semester, nilai ujian akhir semester, dan indeks prestasi kumulatif (IPK). Metode KNN digunakan untuk memprediksi kelulusan seorang mahasiswa di sebuah matakuliah berdasarkan data sebelumnya. Hasil prediksi ini akan digunakan sebagai fitur tambahan yang digunakan dalam pembentukan kelas mahasiswa menggunakan metode Kmeans. Berdasarkan hasil pengujian yang dilakukan diperoleh kesimpulan bahwa jumlah cluster atau kelas dan jumlah data yang digunakan mempengaruhi dari kualitas kluster yang dibentuk oleh metode K-Means dan KNN yang digunakan. Nilai Silhouette Indeks tertinggi diperolah saat menggunakan 100 data dengan jumlah cluster 10 sebesar 0,534 yang tergolong kelas dengan kualitas medium structure[2].

Evi dalam penelitiannya mengungkapkan bahwa pengayaan materi merupakan salah satu persiapan peserta didik untuk menghadapi Ujian Nasional (UN). Di SMA NEGERI 4 TASIKMALAYA tidak ada pengelompokan khusus untuk pengayaan UN. Tidak adanya pengelompokan khusus, mengakibatkan kurang tepat sasaran dalam mengatur jadwal mata pelajaran apa saja yang benar-benar dibutuhkan oleh siswa tersebut. Untuk itu diperlukan solusi yang dapat mengatasi kesulitan tersebut. Metode clustering untuk pengelompokan siswa dengan menggunakan Algoritma K-Means dan untuk pola mata pelajaran tiap kelompok siswa menggunakan Algoritma FP-Growth. Jumlah kluster ada k=2 dan maksimal itemset adalah 3 itemset, jumlah dataset yang digunakan adalah 190 dataset. Aplikasi ini selain menampilkan pengelompokan siswa pada mata pelajaran UN, juga dapat menampilkan pola mata pelajaran UN yang dibutuhkan dari tiap kelompoknya[3].

METODE PENELITIAN

Analisa data

Data yang digunakan dalam penelitian ini adalah data transkrip nilai mahasiswa, data mahasiswa yang mengambil matakuliah peminatan, data mahasiswa PKL, data mahasiswa skripsi 1, data mahasiswa skripsi 2 dan data yudisium untuk stambuk 2015 program studi SI (Sistem Informasi) Universitas Potensi Utama. Jumlah data yang digunakan ada 25 record dengan 7 variabel yang digunakan yaitu (Nilai, Nilai Gagal, Peminatan, PKL, Skripsi 1, Skripsi 2 dan Lulus).

Tabel 1 : Data Mahasiswa Prodi Sistem Informasi

| **No.** | **NIM** | **Nama** | **Nilai** | **Nilai Gagal** | **Peminatan** | **PKL** | **Skripsi 1** | **Skripsi 2** | **Lulus** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1513000001 | Ahmad Fariz | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1513000002 | Aneke Aldwi | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1513000003 | Arya Fatwa Gumilang | 2 | 3 | 1 | 1 | 2 | 2 | 2 |
| 4 | 1513000004 | Aulia Ulfa | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1513000005 | Baqizzarqoni | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 6 | 1513000006 | Beby Faradina Anggraeny Hrp | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 7 | 1513000007 | Bismar Fadli | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | 1513000009 | David H Christian Surbakti | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 9 | 1513000010 | Desy Arista | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 10 | 1513000011 | Dimas Rahmawan | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 11 | 1513000012 | Dinda Novita Sari | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 12 | 1513000013 | Erlina Hapni | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 13 | 1513000014 | Gusdinda Pramiptha | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 14 | 1513000015 | Hardiansyah | 2 | 3 | 1 | 1 | 2 | 2 | 2 |
| 15 | 1513000016 | Haris Fadillah Nasution | 2 | 3 | 1 | 2 | 2 | 2 | 2 |
| 16 | 1513000017 | Kezia Veronika Sinaga | 2 | 2 | 1 | 1 | 1 | 2 | 2 |
| 17 | 1513000018 | Lestina Fraskawita Silalahi | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 18 | 1513000019 | M. Alfin Fauzi | 2 | 3 | 2 | 2 | 2 | 2 | 2 |
| 19 | 1513000020 | M. Fauzi Syahputra | 2 | 3 | 1 | 2 | 2 | 2 | 2 |
| 20 | 1513000024 | Novia Anjelia | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 21 | 1513000025 | Novita Sari | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 22 | 1513000026 | NURUL HUSNAH | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 23 | 1513000027 | Oktarina Wulandari | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 24 | 1513000030 | Reza Riyanto | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 25 | 1513000031 | Riani Ilyas | 2 | 2 | 1 | 1 | 1 | 1 | 1 |

Tabel 2 : Variabel yang digunakan

|  |  |  |
| --- | --- | --- |
| **Variabel** | **Konversi** | **Keterangan** |
| Nilai | 1 | Nilai OK |
| 2 | Nilai Gagal |
| Nilai Gagal | 1 | Tidak Ada |
| 2 | 0 > x <=20 |
| 3 | >20 SKS |
| Peminatan | 1 | Tepat Waktu |
| 2 | Tidak Tepat Waktu |
| PKL | 1 | Tepat Waktu |
| 2 | Tidak Tepat Waktu |
| Skripsi 1 | 1 | Tepat Waktu |
| 2 | Tidak Tepat Waktu |
| Skripsi 2 | 1 | Tepat Waktu |
| 2 | Tidak Tepat Waktu |
| Lulus | 1 | Tepat Waktu |
| 2 | Tidak Tepat Waktu |

Menghitung dengan Algoritma K-Means

Langkah-langkah algoritma K-Means :

1. Tentukan banyak/jumlah kelompok data
2. Bangkitkan *centroid* awal secara acak
3. Hitung jarak data ke tiap-tiap *centroid*.

 $d\_{euclidean }\left(x,y\right)\sqrt{\sum\_{i=1}^{n}(xi-yi)^{2}}$

1. Pengelompokan data

Dari hasil tahap ke 3 tersebut maka dilakukan pengelompokan data.

1. Jika ada kelompok data yang tidak sama atau nilai kelompoknya masih berubah lalu lakukan tahap yang ke 3.

Formula membangkitkan *centroid* baru :

$C=\frac{∑m}{n}$

Apabila hasil pengelompokan data pada iterasi pertama tidak sama dengan kelompok di awal, maka dilanjutkan langkah selanjutnya yaitu tentukan kembali nilai pusat data sementara. Gunakan formula ini :

$$C=\frac{∑m}{n}$$

Menghitung *jarak* ke *centroid* yang baru untuk iterasi selanjutnya.

Algoritma c4.5

Tahapan C4.5 :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus sebagai berikut:



(1)

Di mana :

*S* : Himpunan Kasus

*A* : Atribut

*n* : Jumlah Partisi Atribut A

*|Si|* : Jumlah Kasus pada Partisi ke-i

*|S|* : Jumlah Kasus dalam S

Sementara itu, perhitungan nilai *entropy* dapat dilihat pada persamaan berikut ini:



(2)

Di mana :

*S* : Himpunan Kasus

*A* : Atribut

*n* : Jumlah Partisi *S*

*pi* : Proporsi dari *Si* terhadap *S*

Tabel 3 : Menghitung Jarak dan Pembuatan Group

| **Jarak ke C1** | **Jarak ke C2** | **Jarak ke C3** | **Group** |
| --- | --- | --- | --- |
| 0.00 | 2.83 | 1.41 | 1 |
| 0.00 | 2.83 | 1.41 | 1 |
| 2.83 | 0.00 | 2.00 | 2 |
| 0.00 | 2.83 | 1.41 | 1 |
| 3.16 | 1.41 | 2.45 | 2 |
| 3.16 | 1.41 | 2.45 | 2 |
| 0.00 | 2.83 | 1.41 | 1 |
| 3.16 | 1.41 | 2.45 | 2 |
| 0.00 | 2.83 | 1.41 | 1 |
| 3.16 | 1.41 | 2.45 | 2 |
| 1.41 | 2.00 | 0.00 | 3 |
| 0.00 | 2.83 | 1.41 | 1 |
| 1.41 | 2.00 | 0.00 | 3 |
| 2.83 | 0.00 | 2.00 | 2 |
| 3.00 | 1.00 | 2.24 | 2 |
| 2.00 | 1.41 | 1.41 | 3 |
| 3.16 | 1.41 | 2.45 | 2 |
| 3.16 | 1.41 | 2.45 | 2 |
| 3.00 | 1.00 | 2.24 | 2 |
| 0.00 | 2.83 | 1.41 | 1 |
| 0.00 | 2.83 | 1.41 | 1 |
| 0.00 | 2.83 | 1.41 | 1 |
| 0.00 | 2.83 | 1.41 | 1 |
| 1.41 | 2.00 | 0.00 | 3 |
| 1.41 | 2.00 | 0.00 | 3 |

Tabel 4 : Hasil Perhitungan Cluster Baru (Iterasi 1)

| **No.** | **NIM** | **Nama** | **Nilai** | **Nilai Gagal** | **Peminatan** | **PKL** | **Skripsi 1** | **Skripsi 2** | **Lulus** | **Cluster Awal** | **Cluster Baru** | **Pencocokan** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1513000001 | Ahmad Fariz | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 2 | 1513000002 | Aneke Aldwi | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 3 | 1513000003 | Arya Fatwa Gumilang | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 0 | 2 | -2 |
| 4 | 1513000004 | Aulia Ulfa | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 5 | 1513000005 | Baqizzarqoni | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 6 | 1513000006 | Beby Faradina Anggraeny Hrp | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 7 | 1513000007 | Bismar Fadli | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 8 | 1513000009 | David H Christian Surbakti | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 9 | 1513000010 | Desy Arista | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 10 | 1513000011 | Dimas Rahmawan | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 11 | 1513000012 | Dinda Novita Sari | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 3 | -3 |
| 12 | 1513000013 | Erlina Hapni | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 13 | 1513000014 | Gusdinda Pramiptha | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 3 | -3 |
| 14 | 1513000015 | Hardiansyah | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 0 | 2 | -2 |
| 15 | 1513000016 | Haris Fadillah Nasution | 2 | 3 | 1 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 16 | 1513000017 | Kezia Veronika Sinaga | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 0 | 3 | -3 |
| 17 | 1513000018 | Lestina Fraskawita Silalahi | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 18 | 1513000019 | M. Alfin Fauzi | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 19 | 1513000020 | M. Fauzi Syahputra | 2 | 3 | 1 | 2 | 2 | 2 | 2 | 0 | 2 | -2 |
| 20 | 1513000024 | Novia Anjelia | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 21 | 1513000025 | Novita Sari | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 22 | 1513000026 | NURUL HUSNAH | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 23 | 1513000027 | Oktarina Wulandari | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | -1 |
| 24 | 1513000030 | Reza Riyanto | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 3 | -3 |
| 25 | 1513000031 | Riani Ilyas | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 3 | -3 |

Tabel 5 : Hasil Perhitungan Cluster Baru (Iterasi 2)

| **No.** | **NIM** | **Nama** | **Nilai** | **Nilai Gagal** | **Peminatan** | **PKL** | **Skripsi 1** | **Skripsi 2** | **Lulus** | **Cluster Awal** | **Cluster Baru** | **Pencocokan** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1513000001 | Ahmad Fariz | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 2 | 1513000002 | Aneke Aldwi | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 3 | 1513000003 | Arya Fatwa Gumilang | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 0 |
| 4 | 1513000004 | Aulia Ulfa | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 5 | 1513000005 | Baqizzarqoni | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 6 | 1513000006 | Beby Faradina Anggraeny Hrp | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 7 | 1513000007 | Bismar Fadli | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 8 | 1513000009 | David H Christian Surbakti | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 9 | 1513000010 | Desy Arista | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 10 | 1513000011 | Dimas Rahmawan | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 11 | 1513000012 | Dinda Novita Sari | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 0 |
| 12 | 1513000013 | Erlina Hapni | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 13 | 1513000014 | Gusdinda Pramiptha | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 0 |
| 14 | 1513000015 | Hardiansyah | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 0 |
| 15 | 1513000016 | Haris Fadillah Nasution | 2 | 3 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 16 | 1513000017 | Kezia Veronika Sinaga | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 0 |
| 17 | 1513000018 | Lestina Fraskawita Silalahi | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 18 | 1513000019 | M. Alfin Fauzi | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 19 | 1513000020 | M. Fauzi Syahputra | 2 | 3 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 0 |
| 20 | 1513000024 | Novia Anjelia | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 21 | 1513000025 | Novita Sari | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 22 | 1513000026 | NURUL HUSNAH | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 23 | 1513000027 | Oktarina Wulandari | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 24 | 1513000030 | Reza Riyanto | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 0 |
| 25 | 1513000031 | Riani Ilyas | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 0 |

HASIL DAN PEMBAHASAN

Pengujian dari data yang digunakan dari Program Studi Sistem Informasi adalah sebanyak 25 record dengan menggunakan tool RapidMiner.

1. Analisa dengan algoritma K-Means

Analisa data yang pertama dilakukan untuk mengelompokkan data atau proses clustering data. Proses clustering data menggunakan metode K-means. Jumlah kluster yang akan dibentuk ada 3 cluster. Dengan pengambilan centroid awal secara acak. Hasil dari analisa menggunakan algoritma K-Means diperoleh pusat kluster data sesuai gambar dibawah ini :



Gambar 1 : Nilai Centroid terakhir dari proses ujicoba dengan software RapidMiner



Gambar 2 : Grafik Clustering dari hasil ujicoba dengan software RapidMiner



Gambar 3 : Gambar: Retrieve data dan pemasangan data dengan operator Clustering

1. Analisa data dengan C4.5

Berdasarkan hasil pengelompokan data dengan menggunakan algoritma K-Means selanjutnya dianalisa dengan algoritma C4.5 untuk mendapatkan pengetahuan yang dapat digunakan untuk prediksi kelulusan.

Data yang digunakan sebanyak 25 record dengan sub variabel output ada 3 yaitu Cluster 1, cluster 2 dan Cluster 3. Dimana variable output diperoleh dari hasil pengolahan dengan algorima K-Means.



Gambar 4 : Hasil Tampilan Pohon Keputusan pada software RapidMiner

Pengetahuan yang dihasilkan:

1. If Nilai <=1.5 Then Cluster 1(Mahasiswa Lulus Tepat Waktu dengan Nilai OK)
2. If Nilai >1.5 and Nilai Gagal >2.5 Then Cluster 2 (Mahasiswa Lulus Tidak Tepat Waktu dengan Nilai Bermasalah)
3. If Nilai >1.5 and Nilai Gagal <=2.5 Then Cluster 2 (Mahasiswa Lulus Tepat Waktu dengan Nilai Bermasalah)

KESIMPULAN DAN SARAN

1. Kesimpulan
2. Algorima k-means mampu mengelompokkan data kelulusan mahasiswa. Pada penelitian ini dikelompokkan menjadi 3 kelompok yaitu mahasiswa lulus tepat waktu dengan nilai tidak bermasalah, kelompok kedua mahasiswa tidak lulus tepat waktu karena nilai bermasalah dan kelompok ketiga mahasiswa lulus tepat waktu dengan nilai bermasalah.
3. Algoritma c4.5 mampu menghasilkan pengetahuan yang dapat digunakan sebagai dasar prediksi kelulusan mahasiswa. Berdasarkan pengolahan algorima c4.5 yang menjadi kriteria utama menentukan kelulusan mahasiswa adalah nilai, selanjutnya diikuti dengan nilai gagal. Dimana diperoleh persentase hasil pengujian kecocokan pengetahuan dengan data pengujian >=90%.
4. Saran
5. Sebaiknya untuk menentukan centroid awal sebelum melakukan perhitungan menggunakan algoritma tambahan begitu juga dalam menentukan jumlah kelompok yang akan dihasilkan, agar hasil pengelompokan yang dihasilkan lebih optimal.

**DAFTAR PUSTAKA**

[1] Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *5*(1), 62.

[2] Pradnyana, G. A., & Permana, A. A. J. (2018). Sistem Pembagian Kelas Kuliah Mahasiswa dengan Metode K-Means dan K-Nearest Neighbors untuk Meningkatkan Kualitas Pembelajaran. *JUTI: Jurnal Ilmiah Teknologi Informasi*, *16*(1), 59-68.

[3] Mulyani, E. D. S., Agustin, Y. H., Surgawi, N. M., & Susanto, S. (2018). Implementasi Algoritma K-Means Dan Fp-Growth Untuk Rekomendasi Bimbingan Belajar Berdasarkan Segmentasi Akademik Siswa. *IT (Informatic Technique) Journal*, *6*(2), 160-173.

[4] Helilintar, R., & Farida, I. N. (2018). Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiwa. *Jurnal Sains dan Informatika*, *4*(2), 80-87.

[5] Kusrini, E. T. L. (2009). Algoritma data mining. *Yogyakarta: Andi Offset*.

[6] Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, *1*(1), 72-77.

[7] Hairani, H., Saputro, K. E., & Fadli, S. (2020). K-means-SMOTE untuk menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4. 5, SVM, dan naive Bayes. *Jurnal Teknologi dan Sistem Komputer*, *8*(2), 89-93.

[8] Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *IT Journal Research and Development*, *3*(1), 22-31.

[9] Nasari, F., & Sianturi, C. J. M. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran Diare Di Kabupaten Langkat. *Cogito Smart Journal*, *2*(2), 108-119.

[10] Haryanto, E. V., & Nasari, F. (2015). Penerapan Metode Fuzzy Mamdani Dalam Memprediksi Tingginya Pemakaian Listrik (Studi Kasus Kelurahan Abc). *SEMNASTEKNOMEDIA ONLINE*, *3*(1), 2-2.

[11] Kurniawan, H., & Saleh, A. (2017, October). Analisa Pola Data Penyakit Rumah Sakit Dengan Menerapkan Metode Association Rule Menggunakan Algoritma Apriori. In *Seminar Nasional Informatika (SNIf)* (Vol. 1, No. 1, pp. 195-201).