

**PENGAMATAN STUDI MAHASISWA RENTAN *DROP OUT*
MENGUNAKAN ALGORITMA C4.5 DAN *K-MEANS*
(STUDI KASUS FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH YOGYAKARTA)**

*The Monitoring of Student Study of Drop out Susceptible using C4.5 and K-Means Algorithm
(Study Case Engineering Faculty Universitas Muhammadiyah Yogyakarta)*

Lita Gusti Pertiwi¹, Eko Prasetyo² dan Asroni³
¹Mahasiswa (NIM. 20140140067)
²Dosen Pembimbing 1 ³Dosen Pembimbing 2

INTISARI

Pendidikan memiliki peranan penting dalam kelangsungan hidup bangsa dan negara. Umumnya mahasiswa di perguruan tinggi memiliki kualitas tersendiri pada setiap individunya. Kelancaran studi menjadi salah satu tolak ukur keberhasilan dalam sistem pendidikan. *Drop out* sebagai bentuk *punishment* diberlakukan bagi mahasiswa yang dianggap kurang dalam potensi akademik. Untuk memonitor dan meminimalisir hal tersebut sebuah *Bussiness Intelligence* dapat digunakan. Menggunakan algoritma C4.5 dan *K-Means* yang digunakan pada klasifikasi data yang memiliki atribut-atribut numerik dan kategorial dari data. Sehingga menghasilkan proses klasifikasi berupa aturan yang dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* baru. Hasil dari penelitian ini yakni dari 19527 *record* data mahasiswa, 12% (2360 *record* data mahasiswa) berpotensi *drop out* berdasar algoritma C4.5 dan 3.6% (696 *record* data mahasiswa) berpotensi *drop out* berdasar algoritma *K-Means*. Potensi tertinggi terjadi pada angkatan 2013. Pengaruh tertinggi adalah masa studi semester 3 dan 4.

Kata Kunci: *Drop out*, Algoritma C4.5, Algoritma *K-Means*, Data mining

ABSTRACT

Education has an essential role in the survival of the nation and state. Generally, students in universities have their own quality in each individual. The continuity of study is one of the benchmarks of success in the education system. Drop out as a punishment form applied to students who are considered lacking in the academic potential. To monitor and to minimize it, a Bussiness Intelligence can be used. Using the C4.5 and K-Means algorithms which are used in data classification that have numerical and categorical attributes of the data. So that to produce a classification process of rules that can be applied to predict the value of the discrete type attribute of the new record. The results of this study are from 19527 students data record, 12% (2360 Students data record) have the potential to drop out based on algorithms C4.5 and 3.6% (696 Students data record) potentially drop out based on K-Means algorithm. The highest potential occurs in the year 2013. The highest effect is the study period in semester 3 and 4.

Keywords: *Drop out*, Algorithm C4.5, K-Means Algorithm, Data mining

1. PENDAHULUAN

Pendidikan memiliki peranan penting dalam kelangsungan hidup bangsa dan negara. Negara Indonesia menjadi salah satu negara dengan angka pengangguran yang terus meningkat setiap tahunnya. Peningkatan angka tersebut karena sumber daya manusia yang ada kurang memenuhi standar. Perguruan Tinggi sebagai strata tertinggi pendidikan merupakan salah satu pereka cipta sumber daya manusia yang sesuai dengan kebutuhan dunia profesi.

Umumnya mahasiswa di perguruan tinggi memiliki kualitas tersendiri pada setiap individunya. Maka dari itu diperlukan studi untuk mengamati perkembangan akademik mahasiswa agar sesuai dengan peraturan akademik yang berlaku sehingga tidak terjadi *drop out* atau tercabut statusnya sebagai mahasiswa.

Untuk memudahkan pengamatan tersebut dapat dilakukan dengan menggunakan metode data mining. Data mining adalah proses yang menggunakan teknik perhitungan statistika untuk mengidentifikasi informasi penting dalam berbagai *database* besar (Kamagi, 2014). Adapun metode yang digunakan yaitu metode klasifikasi dan *clustering* menggunakan algoritma C4.5 dan *K-Means*.

Hasil analisis ini akan menjadi bahan tinjauan kebijakan universitas dalam menanggulangi dan menekan angka *drop out* mahasiswa di Fakultas Teknik Universitas Muhammadiyah Yogyakarta. Sehingga pembuatan analisis ini dapat mempermudah universitas dalam melakukan

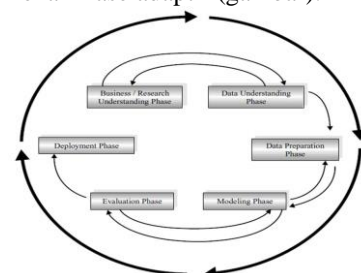
crossing dan peningkatan kualitas universitas. Faktor penyebab *drop out* difokuskan pada faktor akademik dengan menggunakan *macine learning* WEKA 3.8.

2. LANDASAN TEORI

2.1 Data Mining

Data mining adalah proses untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dengan berbagai *database* besar menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning (Turban, 2005).

Menurut (Kusrini, 2009) *Cross-Industry Standard Process for Data Mining* (CRISP-DM) menyediakan standar proses data mining sebagai stategi dalam memecahkan masalah secara umum dari bisnis atau penelitian. Dalam CRISP-DM proyek data mining memiliki siklus hidup yang terbagi dalam enam fase adaptif (gambar).



Gambar 2.1 Proses data mining menggunakan CRISP-DM

Enam fase CRISP-DM (Larose,2005):

- Fase Pemahaman Bisnis (*Business Understanding Phase*)
Fase ini merupakan fase penentuan tujuan dan kebutuhan secara detail dan keseluruhan.
- Fase Pemahaman Data (*Data Understanding Phase*)
Fase ini adalah fase untuk mengumpulkan data dan menggunakan analisis penyelidikan data untuk pengenalan lebih lanjut dan pencarian pengetahuan awal.
- Fase Pengolahan Data (*Data Preparation Phase*)
Fase ini merupakan fase persiapan dari data awal untuk keseluruhan fase berikutnya dan untuk perangkat pemodelan. Memilih kasus dan mengubah *variabel* yang dibutuhkan untuk dianalisis.
- Fase Pemodelan (*Modelling Phase*)
Ini merupakan fase memilih dan mengaplikasikan teknik pemodelan yang sesuai, kalibrasi aturan model untuk mengoptimalkan hasil.
- Fase Evaluasi (*Evaluation Phase*)
Fase untuk mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
- Fase Penyebaran (*Deployment Phase*)
Fase penggunaan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesainya proyek, penerapan data mining secara parallel pada departemen lain dan pembuatan laporan juga dibutuhkan untuk mendukung penyempurnaan proyek.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005):

- Deskripsi
Terkadang peneliti atau analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
- Estimasi
Estimasi hampir sama dengan klasifikasi kecuali variabel target. Estimasi lebih kearah numerik daripada kearah kategori. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.
- Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi kecuali dalam prediksi nilai dari hasil akan ada di masa mendatang.
- Klasifikasi
Pada klasifikasi terdapat target variabel kategori yang merupakan teknik mengklasifikasi data. Pada klasifikasi diharuskan ada variabel dependen sedangkan pada klastering variabel dependen tidak ada.
- Pengklusteran
Pengklusteran merupakan pengelompokkan *record*, pengamatan, atau memperhatikan sehingga membentuk kelas objek-objek yang memiliki kemiripan. Perbedaan pengklusteran dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran melakukan pembagian atau pengelompokkan terhadap keseluruhan data yang memiliki kemiripan (homogen),

yang mana kemiripan *record* dalam suatu kelompok akan bernilai maksimal sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

- Asosiasi
Asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu atau dalam dunia bisnis.

2.2 Algoritma C4.5

(Kamagi, 2014) Secara umum untuk membangun pohon keputusan Algoritma C4.5 adalah sebagai berikut :

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses masing-masing cabang hingga semua kasus memiliki kelas yang sama.

Pemilihan atribut sebagai akar didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti berikut:

$$\text{Gain}(S,A) = \text{Entropy}(s) - \sum_{i=1}^n \frac{|s_i|}{|S|} * \text{Entropy}(s_i) \quad \dots(2.1)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Untuk mendapatkan nilai Gain sebelumnya mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif atribut masukan untuk menghasilkan sebuah atribut. Rumus dasar dari entropi adalah sebagai berikut:

$$\text{Entropy}(s) = - \sum_{i=1}^n p_i * \log_2 p_i \quad \dots\dots\dots(2.2)$$

Keterangan:

S : Himpunan kasus

n : Jumlah partisi S

pi : Proporsi dari Si terhadap S

2.3 Algoritma K-Means

Algoritma *K-Means* merupakan algoritma klusterisasi dimana data dikelompokkan berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Untuk mengelompokkan data dengan memaksimalkan kemiripan data dalam suatu kluster dan meminimalkan kemiripan data antar kluster. Ukuran kemiripan yang digunakan dalam kluster yaitu fungsi jarak. Sehingga kemiripan data yang maksimal didapatkan dari jarak terpendek antara data terhadap titik *centroid*. Tahapan awal proses klusterisasi menggunakan algoritma *K-Means* adalah menentukan titik awal *centroid* yang pada umumnya dibangkitkan secara acak sesuai dengan jumlah kluster yang ditentukan diawal. Setelah *centroid* terbentuk kemudian dihitung jarak tiap datanya. Perhitungan jarak *Euclidean* adalah sebagai berikut (Asroni, 2016):

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \dots\dots\dots(2.3)$$

Keterangan :

d : jarak

x : data

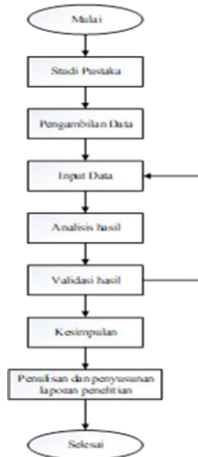
y : centroid
 i : banyaknya data
 n : jumlah data

3. METODOLOGI PENELITIAN

Metodologi penelitian disusun untuk mengarahkan langkah-langkah penelitian agar tujuan penelitian dapat tercapai.

3.1 Alur Penelitian

Alur penelitian ini menjadi pedoman peneliti dalam melakukan tahap-tahap penelitian agar penelitian menjadi terarah.



Gambar 3.1 Bagan alur penelitian

Pada penelitian ini dilakukan studi pustaka untuk mengetahui penelitian-penelitian yang telah dilakukan sebelumnya terkait dengan data mining, penggunaan algoritma C4.5 dan *K-Means*. Hal ini dijadikan sebagai acuan untuk memperkuat teori-teori yang digunakan dan memvalidasi penelitian ini.

Selanjutnya dilakukan studi pustaka, selanjutnya pengambilan data mahasiswa Universitas Muhammadiyah Yogyakarta dari *database* milik Biro Sistem Informasi Universitas Muhammadiyah Yogyakarta. Data tersebut merupakan data mahasiswa Fakultas Teknik UMY yang akan diolah dengan metode klasifikasi menggunakan algoritma C4.5 dan klustering menggunakan algoritma *K-Means*. Data yang olahan tersebut digunakan untuk mendapatkan pengetahuan dan hasil sesuai dengan penelitian ini.

Selanjutnya adalah proses penginputan data. Data yang diinputkan dalam penelitian ini adalah data yang telah disaring dan diseleksi kembali untuk mengambil beberapa atribut dengan menggunakan *SQL Server Management Studio 2013* yang nantinya dapat diolah dengan metode klasifikasi dan klustering menggunakan algoritma C4.5 dan *K-Means*.

Selanjutnya adalah proses analisa hasil, proses ini merupakan proses dimana data tersebut diimplementasikan dalam masing-masing metode dan algoritma untuk mendapatkan hasil akhir yang diharapkan.

Selanjutnya mendapatkan hasilnya, dilakukan validasi hasil dimana hasil dari pengimplementasian metode-metode dan algoritma-algoritma yang digunakan memiliki kesesuaian dengan data-data yang sebenarnya.

Berdasarkan hasil analisis dan validasi hasil dapat ditarik kesimpulan yang mengacu pada tujuan dan rumusan masalah dalam penelitian. Saran untuk pengembangan penelitian yang akan datang dimasukkan untuk meningkatkan kualitas penelitian dan penyelesaian masalah yang belum terselesaikan dalam penelitian ini.

3.2 Bahan Penelitian

Dalam penelitian ini bahan yang dibutuhkan adalah *record* data mahasiswa Fakultas Teknik Universitas Muhammadiyah Yogyakarta dari angkatan 2012 hingga 2016 yang didapatkan dari BSI (Biro Sarana Informasi) Universitas Muhammadiyah Yogyakarta.

3.3 Metode/Disain Algoritma

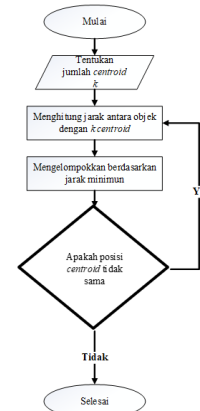
Metode yang digunakan dalam penelitian ini merupakan metode-metode yang paling banyak digunakan dalam penelitian data mining. Maka dari itu penelitian ini dimaksudkan untuk membandingkan metode mana yang lebih efisien penggunaannya untuk memenuhi tujuan dalam penelitian ini. Algoritma C4.5 adalah algoritma yang paling efektif dan fleksibel dalam proses pengklasifikasian. Sedangkan algoritma *K-Means* merupakan algoritma yang menggunakan pengelompokan data.

3.3.1 Flowchart C4.5



Gambar 3.2 Diagram alur algoritma C4.5

3.3.2 Flowchart K-Means



Gambar 3.3 Diagram alur algoritma K-Means

3.4 Prosedur Validasi

Penelitian ini menggunakan proses validasi dimana data dari atribut-atribut yang digunakan pada penelitian sesuai

dengan data sebenarnya. Pengujian metode ini dilakukan dengan mengimplementasikan rumus-rumus pada masing-masing metode menggunakan mesin pendukung.

4 HASIL DAN PEMBAHASAN

4.1 Pengambilan Data

Sumber data utama yang digunakan pada penelitian ini adalah *record* data akademik mahasiswa di Fakultas Teknik tahun 2012, 2013, 2014, 2015, dan 2016. Data yang diterima dalam bentuk tabel dapat diakses menggunakan *SQL Server Management Studio 2014*. Agar dapat mengakses data tabel peneliti harus memiliki hak akses ke *database* server kampus Universitas Muhammadiyah Yogyakarta.

4.2 Penyeleksian Data

Data akademik yang akan digunakan disaring dan diambil beberapa atribut untuk dianalisis. Pengambilan data dari *SQL Server Management Studio* diolah menggunakan *Pentaho* untuk mendapatkan *dataset* yang baru. Jumlah data mahasiswa Fakultas Teknik dari tahun 2012 sampai 2016 sebanyak 20985 *record*. Atribut-atribut yang digunakan antara lain Id mahasiswa, tahun ajaran, jurusan, fakultas, jumlah sks, Indeks Prestasi Semester, angkatan, dan NIM. Atribut jumlah sks dan Indeks Prestasi Semester (IPS) merupakan hasil dari penambahan *query* pada proses ini.

```
SELECT dbo.Nds_Transcript.Student_Id, dbo.Nds_Departement.Faculty_Id,
dbo.Nds_Student.Department_Id, dbo.Nds_Student.Entry_Year_Id,
dbo.Nds_Transcript.Term_Year_Id, dbo.Nds_Transcript.IPS,
dbo.Nds_Transcript.jumlahsks
FROM dbo.Nds_Student INNERJOIN dbo.Nds_Transcript ON
dbo.Nds_Student.Register_Id = dbo.Nds_Transcript.Student_Id
INNERJOIN dbo.Nds_Departement ON dbo.Nds_Student.Department_Id =
dbo.Nds_Departement.Department_Id
WHERE (dbo.Nds_Departement.Faculty_Id = 1)
```

Listing program 1 *Query view* pada *database*

4.3 Preprocessing Data

Setelah data diseleksi sesuai dengan atribut yang akan digunakan dilakukan *preprocessing* data untuk memisahkan data agar tidak ada missing value, duplikat data, dan kesalahan-kesalahan *dataset* dalam format *excel*. *Dataset* tersebut disimpan dalam format *.csv* agar dapat dibaca dan dibuka dalam *WEKA*.

Pada *record* data tersebut didapatkan sebanyak 384 *record* data dengan jumlah SKS lebih dari 24 SKS per semester. *Record* data tersebut tidak digunakan dalam proses analisis karena tidak memenuhi kriteria yang digunakan.

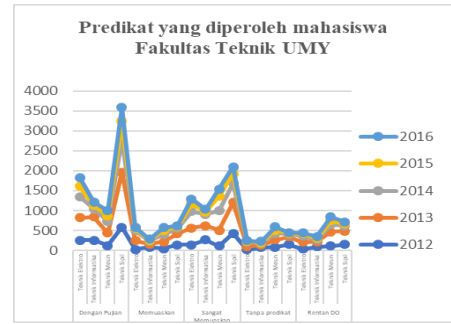
4.4 Pengolahan dan Pengujian Data

Data mahasiswa di Fakultas Teknik memiliki 20985 *record* data mahasiswa yang dipisahkan pada setiap angkatan. Ditambahkan pengelompokkan *record* dalam kolom baru yaitu kolom kategori berdasarkan IPS nya yang akan digunakan sebagai validator prediksi pada algoritma C4.5. Penambahan kolom kategori dilakukan pada semua angkatan dari angkatan 2012 hingga 2016. Sedangkan untuk pengujian menggunakan algoritma *K-Means* kolom tersebut di *remove*.

4.4.1 Analisa algoritma C4.5 menggunakan WEKA

Tahapan analisa dan pengujian menggunakan software *WEKA* dilakukan pada semua angkatan. Berdasarkan hasil

pengujian menggunakan *record-record* yang masuk kedalam kategori rentan *drop out* sesuai dengan pengimplementasian menggunakan rumus *Entropy* dan *Gain*.



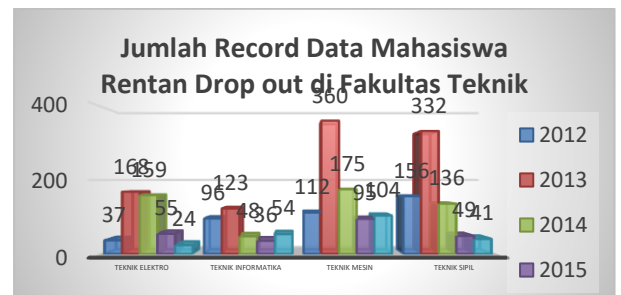
Gambar 4.1 Grafik hasil analisis seluruh *record* data mahasiswa

Detail jumlah *record* data mahasiswa dalam semua kategori dan jurusan dapat dilihat dalam tabel dibawah ini.

Tabel 4.1 Jumlah seluruh *record* data pada semua kategori

Row Labels	2012	2013	2014	2015	2016	Grand Total
Dengan Pujian	1219	2843	1833	951	786	7632
Teknik Elektro	253	566	528	265	219	1831
Teknik Informatika	259	579	206	81	93	1218
Teknik Mesin	125	313	294	143	125	1000
Teknik Sipil	582	1385	805	462	349	3583
Memuaskan	295	757	641	208	152	2053
Teknik Elektro	28	224	243	59	29	583
Teknik Informatika	82	75	81	15	30	283
Teknik Mesin	44	175	184	99	67	569
Teknik Sipil	141	283	133	35	26	618
Sangat Memuaskan	941	1956	1674	848	547	5966
Teknik Elektro	141	420	419	188	127	1295
Teknik Informatika	263	363	284	55	69	1034
Teknik Mesin	111	391	490	366	186	1544
Teknik Sipil	426	782	481	239	165	2093
Tanpa predikat	329	513	383	151	140	1516
Teknik Elektro	18	85	105	29	11	248
Teknik Informatika	72	57	50	21	25	225
Teknik Mesin	84	187	157	77	87	592
Teknik Sipil	155	184	71	24	17	451
Rentan DO	401	983	518	235	223	2360
Teknik Elektro	37	168	159	55	24	443
Teknik Informatika	96	123	48	36	54	357
Teknik Mesin	112	360	175	95	104	846
Teknik Sipil	156	332	136	49	41	714
Grand Total	3185	7052	5049	2393	1848	19527

Dari grafik dan tabel diatas didapatkan hasil bahwa jumlah *record* data mahasiswa yang masuk kategori rentan *drop out* lebih banyak dibandingkan dengan jumlah *record* data mahasiswa pada kategori memuaskan dan tanpa predikat. *Record* data yang masuk dalam kategori rentan *drop out* sebanyak 2360 *record*. Untuk melihat *record* data lebih detail dapat dilihat dari grafik dan tabel dibawah ini.



Gambar 4.2 *Record* data mahasiswa *drop out*

Pada grafik tersebut didapatkan hasil sebanyak 2360 *record* data mahasiswa yang harus diambil tindakan berupa peringatan ataupun *drop out*. Dengan jumlah 714 (30,2%) *record* data mahasiswa dari prodi Teknik Sipil, 443 (18,8%) *record* data mahasiswa dari prodi Teknik Elektro, 846 (35,8%) *record* data mahasiswa dari prodi Teknik Mesin, dan 357 (15,1%) *record* data mahasiswa dari prodi Teknik Informatika. Detail jumlah *record* data mahasiswa rentan *drop out* di masing-masing prodi dapat dilihat dalam tabel dibawah ini.

Tabel 4.2 Jumlah *record* data mahasiswa rentan *drop out* di prodi Teknik Sipil

No.	Angkatan	Jumlah
1	2012	156
2	2013	332
3	2014	136
4	2015	49
5	2016	41
Total		714

Tabel 4.3 Jumlah *record* data mahasiswa rentan *drop out* di prodi Teknik Elektro

No.	Angkatan	Jumlah
1	2012	37
2	2013	168
3	2014	159
4	2015	55
5	2016	24
Total		443

Tabel 4.4 Jumlah *record* data mahasiswa rentan *drop out* di prodi Teknik Mesin

No.	Angkatan	Jumlah
1	2012	112
2	2013	360
3	2014	175
4	2015	95
5	2016	104
Total		846

Tabel 4.5 Jumlah *record* data mahasiswa rentan *drop out* di prodi Teknik Informatika

No.	Angkatan	Jumlah
1	2012	96
2	2013	123
3	2014	48
4	2015	36
5	2016	54
Total		357

Dari grafik dan tabel *record* data mahasiswa bermasalah diatas didapatkan hasil dengan rata-rata keakuratan atau kebenaran pengujian sebesar 99,99%. Jumlah *record* data terbanyak ada pada prodi Teknik Mesin dengan persentase sebesar 35,8% sedangkan *record* data paling sedikit terdapat pada prodi Teknik Informatika dengan persentase sebesar 15,1% dari seluruh jumlah *record* data mahasiswa rentan *drop out*. Hasil yang diperoleh sebanyak 12% dari jumlah *record* seluruh data. Penggunaan Algoritma C4.5 dinilai sangat efektif untuk pemilihan data sehingga didapatkan hasil yang sangat akurat.

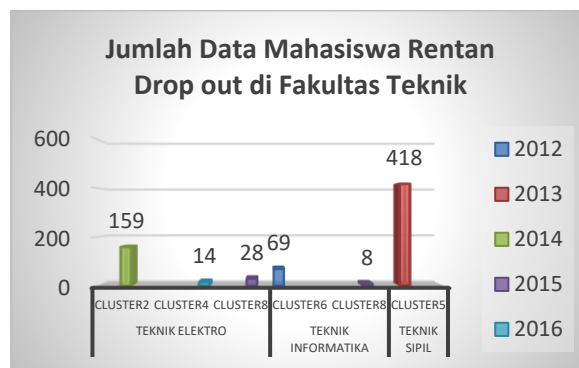
4.4.2 Analisa algoritma K-Means menggunakan WEKA

Tahapan analisa dan pengujian menggunakan software WEKA dilakukan pada semua angkatan dengan menggunakan 9 kluster pada angkatan 2012,2013,2015, 5 kluster pada angkatan 2014 dan 11 kluster pada angkatan 2016. Berdasarkan hasil pengujian menggunakan beberapa kluster sesuai dengan pengimplementasian menggunakan rumus *Euclidean* didapatkan hasil cluster instances seperti tabel berikut.

Tabel 4.6 Cluster Instances semua angkatan di Fakultas Teknik

Cluster Instances	2012	2013	2014	2015	2016
Cluster 0	337 (11%)	1025 (15%)	1300 (26%)	158 (7%)	296 (16%)
Cluster 1	1123 (35%)	300 (4%)	1295 (26%)	456 (19%)	219 (12%)
Cluster 2	476 (15%)	1426 (20%)	159 (3%)	111 (5%)	47 (3%)
Cluster 3	218 (7%)	1076 (15%)	669 (13%)	176 (7%)	97 (5%)
Cluster 4	477 (15%)	438 (6%)	1626 (32%)	123 (5%)	14 (1%)
Cluster 5	151 (5%)	209 (3%)		524 (22%)	107 (6%)
Cluster 6	69 (2%)	410 (6%)		424 (18%)	23 (1%)
Cluster 7	231 (7%)	971 (14%)		385 (16%)	410 (22%)
Cluster 8	103 (3%)	1197 (17%)		36 (2%)	176 (10%)
Cluster 9					188 (10%)
Cluster 10					271 (15%)
Σ data	Σ 3185	Σ 7052	Σ 5049	Σ 2393	Σ 1848

Dari tabel cluster instances diatas dapat dilihat kelompok cluster ekstrim pada grafik dibawah ini.



Gambar 4.3 Grafik *record* data mahasiswa

Dari grafik diatas dapat dilihat hasil yang kontras terjadi di Fakultas Teknik. Jumlah total *record* data yang masuk ke dalam kelompok kluster terendah sebanyak 696 *record* data mahasiswa. Adapun detail jumlah *record-record* yang masuk kedalam kelompok kluster terendah di Fakultas Teknik adalah sebagai berikut.

Tabel 4.7 Detail kelompok kluster terendah

Angkatan	TE	TI	TS	Cluster
2012	0	69	0	cluster 6
2013	0	0	418	cluster 5
2014	159	0	0	cluster 2
2015	28	8	0	cluster 8
2016	8	0	0	cluster 4
Total	195	77	418	Σ 696

Berdasarkan hasil grafik dan tabel diatas, diperoleh hasil yang cukup signifikan dari seluruh *record* data yang ada. Hasil yang ditemukan yaitu sebesar 3,6% dari seluruh *record* data. Hasil yang didapatkan lebih kecil dari jumlah *record* data yang dihasilkan menggunakan algoritma C4.5.

Dari hasil analisis menggunakan kedua algoritma diatas, Algoritma C4.5 lebih cocok digunakan pada *record*

data yang menggunakan IPK sedangkan pada *record* data IPS penggunaannya kurang maksimal dan membutuhkan penanganan ekstra. Namun pada algoritma *K-Means* dapat digunakan untuk *record* dengan penggunaan jangka pendek atau keperluan mendesak.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah metode klasifikasi dan klustering diimplementasikan menggunakan Algoritma C4.5 dan *K-Means* dapat disimpulkan bahwa:

1. Jumlah *record* data mahasiswa yang rentan *drop out* menggunakan algoritma C4.5 sebanyak 12% atau 2360 *record* data mahasiswa sedangkan menggunakan algoritma *K-Means* sebanyak 3.6% atau 696 *record* data mahasiswa dari jumlah total 19527 *record* data mahasiswa.
2. Jumlah *record* data *drop out* terbanyak terdapat pada data mahasiswa angkatan 2013 sebanyak 983 *record*.
3. Algoritma C4.5 lebih cocok untuk diimplementasikan di Fakultas Teknik Universitas Muhammadiyah Yogyakarta dengan kekurangan memerlukan penanganan ekstra sedangkan Algoritma *K-Means* yang dapat diimplementasikan untuk keperluan mendesak. Kedua algoritma tersebut dapat digunakan sebagai acuan pihak Universitas Muhammadiyah Yogyakarta dalam peningkatan kualitas pendidikan.
4. Perkembangan studi mahasiswa dari semester 1 sampai semester 3 merupakan faktor paling berpengaruh, Analisis ini dapat menjadi acuan untuk strategi penentuan kebijakan universitas.

5.2 Saran

Adapun saran yang dapat diberikan sebagai bahan perbaikan dimasa yang akan datang adalah sebagai berikut.

1. Data mahasiswa baru pada database harus dengan atribut yang lengkap, seperti kota asal, asal sekolah, nilai sekolah, jalur masuk, dll.
2. Konversi data apabila dilakukan equivalensi diharapkan sesuai dan sama dengan yang lainnya untuk efisiensi waktu pada penelitian selanjutnya.
3. Pihak universitas harusnya mengambil tindakan dan melakukan monitoring intensif terhadap mahasiswa setelah semester 3 agar tidak mengurangi kualitas pendidikan universitas.
4. Monitoring intensif oleh dosen pembimbing akademik diharapkan lebih optimal agar perkembangan studi mahasiswa dan lama studi mahasiswa dapat terkontrol dengan baik oleh universitas.
5. Untuk penyempurnaan analisis ini perlu dilakukan penelitian selanjutnya dengan menggunakan aturan universitas menggunakan SK rektor nomor : 015/SK-UMY/IX/2015

6. DAFTAR PUSTAKA

- [1] Artaye, K. (2015). Implementation Of Naïve Bayes Classification Method To Predict Graduation Time Of Ibi Darmajaya Scholar. In *Prosiding International*

conference on Information Technology and Business (ICITB) (pp. 284–290).

- [2] Asril, E., Wiza, F., & Yuneфри, Y. (2015). Analisis Data Lulusan Dengan Data Mining Untuk Mendukung Strategi Promosi Universitas Lancang Kuning. *Digital Zone*, 6(2), 24–32.
- [3] Asroni, A., & Adrian, R. (2016). Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *WEKA* Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. *Semesta Teknika*, 18(1), 76–82.
- [4] Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhathu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- [5] Handhayani, T., & Hiryanto, L. (2017). Predicting And Analyzing The Length Of Study-Time Using Support Vector Machine (Case Study: Computer Science Students).
- [6] Jananto, A. (2013). Algoritma *Naive bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa, 18.
- [7] Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa, 6(1).
- [8] Kaur, H., & Kaur, M. (2016). Review of K Mean and Fuzzy Clustering Algorithm. *International Journal of Advanced Engineering Research and Science*, 3(6). Retrieved from <https://www.neliti.com/id/publications/236736/review-of-k-mean-and-fuzzy-clustering-algorithm>
- [9] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining* (1st ed.). ANDI Yogyakarta.
- [10] Nugroho, Y. (2015). Klasifikasi Dan Klustering Mahasiswa Informatika Universitas Muhammadiyah Surakarta.
- [11] Nugroho, Y. S. (2014). Data Mining menggunakan Algoritma *Naive bayes* untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. 2014.
- [12] Setiawan, A., & Margono, K. S. L. (2015). Predicting Success Study Using Students GPA Category. *The Asian Journal of Technology Management (AJTM)*, 8(1), 68–74. <https://doi.org/10.12695/ajtm.2015.8.1.7>
- [13] Turban, E., Aronson, J. E., & Liang, T.-P. (2005). *Decision Support System and Intelligent System - 7th Ed. Jilid 1 (Sistem Pendukung Keputusan dan Sistem Cerdas)*. ANDI Yogyakarta.