



PENGELOMPOKAN MAHASISWA BERPOTENSI *DROP OUT* PADA PROGRAM STUDI TEKNIK INFORMATIKA MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Clustering Students With Potential To Drop Out In The Informatics Engineering Study Program Using The K-Means Clustering Method

Anggun Suzana Barnabas Lomi¹, Arini Aha Pekuwali², Reynaldi Thimotius Abineno³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Kristen Wira Wacana Sumba

Jl.R.Suprato No.35,Prailiu,Kec.Kota Waingapu,Kabupaten Sumba Timur,Nusa Tenggara Timur

Corresponding author: lomianggun@gmail.com¹, arini.pekuwali@unkriswina.ac.id²
reynaldi@unkriswina.ac.id³

ABSTRACT

In the Wira Wacana Christian University Sumba Informatics Study Program, the classes of 2016 and 2017 recorded dropout rates of 40.71% and 31.03%, respectively. This research aims to identify students with the potential to drop out using the K-Means Clustering method on data from the class of 2016. The variables used include Name, NIM, Grade Point Average (GPA), Gender, and Course Grades. Analysis using the RapidMiner tool resulted in two clusters: cluster 0 with 25 students and cluster 1 with 88 students. In cluster 0, 25 students showed low GPA due to missing or incomplete course grades, indicating a higher risk of dropout. In contrast, cluster 1 students had high grades in each course, showing no risk of dropout each semester. The research findings are expected to form a basis for developing strategies to reduce dropout rates and increase graduation rates in the Informatics Engineering Study Program. Knowing these clusters, the program needs to enhance academic monitoring and counseling, especially for students in cluster 0.

Keywords: *Information Technology, K-Means Clustering, Drop Out.*

ABSTRAK

Pada Program Studi Teknik Informatika di Universitas Kristen Wira Wacana Sumba, angkatan 2016 dan 2017 mencatat tingkat dropout masing-masing sebesar 40.71% dan 31.03%. Penelitian ini bertujuan untuk mengidentifikasi mahasiswa yang berpotensi *drop out* dengan menggunakan metode *K-Means Clustering* pada data mahasiswa angkatan 2016. Variabel yang digunakan meliputi Nama, NIM, Indeks Prestasi Kumulatif (IPK), Jenis Kelamin, dan Nilai Mata Kuliah. Berdasarkan hasil analisis hasil pengujian data menggunakan tools *RapidMiner* pada pengelompokan mahasiswa berpotensi *Drop Out* menggunakan metode *k-means clustering* melalui variabel yang telah ditentukan menghasilkan 2 *cluster* dimana *cluster 0* memiliki 25 mahasiswa dan *cluster 1* memiliki 88 mahasiswa. *Cluster 0* dapat disimpulkan bahwa dari 25 mahasiswa dimana pada *cluster* tersebut pada mata kuliah tertentu tidak memiliki nilai mata kuliah atau tidak pernah mengikuti perkuliahan tersebut sehingga berdampak pada IPK yang rendah dan akan terjadinya lama studi dan akan *Drop Out*. Sedangkan *cluster 1* dapat disimpulkan bahwa mahasiswa pada setiap mata kuliah memenuhi nilai tinggi sehingga tidak terjadi *drop out* disetiap semester. Hasil penelitian diharapkan dapat menjadi dasar untuk mengembangkan strategi yang sesuai guna mengurangi angka dropout dan meningkatkan tingkat kelulusan di Program Studi Teknik Informatika.

Kata kunci: *Teknologi Informasi, K-Means Clustering, Drop Out.*

PENDAHULUAN

Salah satu perguruan tinggi terbaik di Waingapu, Sumba Timur, Nusa Tenggara Timur, adalah Universitas Kristen Wira Wacana Sumba memiliki Program Studi Teknik Informatika yang berhasil meluluskan beberapa mahasiswa setiap semesternya. Akan tetapi, ada beberapa hal yang tidak merata yaitu jumlah mahasiswa yang masuk dan keluar di Prodi Teknik Informatika. Berdasarkan informasi yang diperoleh dari hasil wawancara, jumlah mahasiswa putus sekolah menengah atas setiap tahunnya cukup tinggi. Data Mahasiswa Angkatan tahun 2016–2019. dapat dilihat pada tabel berikut.

Tabel 1 Data Mahasiswa Angkatan 2016-2019

Angkatan	Masuk	Lulus	Aktif	Non Aktif	Drop Out
2016	113	67	-	-	46
2017	116	63	8	9	36
2018	144	71	26	14	33
2019	151	33	64	26	28

Dari data tersebut, terlihat bahwa tingkat *drop out* cukup tinggi terutama pada angkatan 2016 dan 2017. Angkatan 2016 menunjukkan jumlah dropout sebesar 46 mahasiswa dari 113 mahasiswa yang masuk, sementara angkatan 2017 memiliki 36 mahasiswa *drop out* dari 116 mahasiswa yang masuk. Angkatan 2018 dan 2019 juga masih tergolong tinggi meskipun mungkin lebih rendah dari angkatan sebelumnya. Berikut adalah tabel yang membandingkan jumlah mahasiswa yang masuk dengan jumlah mahasiswa yang *drop out* berdasarkan data yang ada.

Tingginya angka *drop out* memiliki berbagai dampak negatif pada program studi. Dampak ini meliputi penurunan reputasi akademik program studi, yang dapat mengurangi minat calon mahasiswa baru. Selain itu, angka *drop out* yang tinggi dapat mempengaruhi akreditasi program studi, yang pada gilirannya memengaruhi kepercayaan publik dan daya saing program tersebut. Secara finansial, hilangnya pendapatan dari biaya kuliah mahasiswa yang *drop out* juga dapat mengurangi anggaran untuk pengembangan fasilitas dan kegiatan akademis lainnya. Beban administratif meningkat karena pengelolaan proses *drop out* memerlukan waktu dan sumber daya tambahan. Tingginya angka *drop out* juga dapat menurunkan moral dan motivasi mahasiswa yang masih aktif, serta mengurangi jumlah alumni yang bisa memberikan kontribusi positif bagi program studi.

Berdasarkan penelitian terdahulu. Sebagai penduan dalam menyusun penelitian ini, maka peneliti menggunakan jurnal sebelumnya dengan judul Data Mining dengan Metode *Clustering K-Means* untuk Pengelompokan Mahasiswa Potensial Drop Out Pada Program Studi Teknik Informatika Universitas Siliwangi (adi Firmansyah et al., 2017). Persamaan dari penelitian sebelumnya adalah menggunakan Metode *Clustering K-means*. Sedangkan untuk membedakan penelitian sebelumnya dengan penelitian yang akan dilakukan terletak pada variabel atau atribut, studi kasus, yang digunakan. Seperti pada penelitian terdahulu variabel yang digunakan yaitu Nim, Nama, Indeks Prestasi Kumulatif (IPK), Satuan Kredit Semester (SKS), dan Indeks Prestasi Semester (IPS). Lalu studi kasus dari penelitian terdahulu terletak di Program Studi Teknik Informatika Universitas Siliwangi. Sedangkan dengan penelitian yang akan dilakukan variabel yang digunakan adalah Nim, Nama, Indeks Prestasi Kumulatif (IPK), Jenis Kelamin, dan Nilai Mata Kuliah, dan Keterlambatan. Kemudian untuk studi kasus penelitian yang akan dilakukan di Pogram Studi Teknik Informatika Universitas Kristen Wira Wacana Sumba. Ada juga persamaan penelitian perdahulu yang berjudul Penerapan Data Mining Terhadap Prediksi Mahasiswa Drop Out Pada Kampus STMIK Widuri Jakarta Dengan Metode Decision Tree C4.5 (Ramdan et al., 2023) Persamaan dari penelitian sebelumnya

adalah menggunakan data mining. Sedangkan untuk membedakan penelitian sebelumnya dengan penelitian yang akan dilakukan terletak pada metode yang di gunakan dengan studi kasus. Seperti terletak pada penelitian terdahulu metode menggunakan Decission Tree kemudian studi kasus di kampus SMTIK Widuri Jakarata. Kemudian penelitian yang disusun menggunakan metode K-Means Clustering dengan studi kasus Pogram Studi Teknik Informatika Universitas Kristen Wira Wacana Sumba. Persamaan penelitian terdahulu Penerapan Algoritma C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out Pada Universitas Budi Luhur (Debora et al., 2023). Persaman dari penelitian sebelumnya adalah menggunakan Metode *Clustering K-means*. Sedangkan untuk membedakan penelitian sebelumnya dengan penelitian yang akan dilakukan terletak pada studi kasus Universitas Budi Luhur. Kemudian penelitian yang akan dilakukan studi kasus Pogram Studi Teknik Informatika Universitas Kristen Wira Wacana Sumba.

Penelitian ini akan difokuskan pada mahasiswa angkatan 2016 di Program Studi Teknik Informatika di Universitas Kristen Wira Wacana Sumba. Penelitian ini bertujuan untuk menerapkan metode *K-Means Clustering* pada data mahasiswa angkatan 2016 guna mengidentifikasi kelompok mahasiswa yang berisiko untuk *drop out*. Dapat membantu pihak program studi dalam mengelompokkan mahasiswa dan mendukung pengambilan keputusan yang lebih tepat. Bahwa hasil dari penelitian ini dapat menjadi dasar atau landasan untuk mengembangkan strategi atau tindakan yang lebih efektif dan sesuai dengan kebutuhan, seperti bimbingan akademik, program mentoring, atau dukungan psikologis, guna membantu mahasiswa mengatasi tantangan yang dihadapi dan meningkatkan kemungkinan mereka untuk menyelesaikan studi tepat waktu. Penggunaan metode *K-Means Clustering* diharapkan dapat memberikan perubahan yang besar dalam meningkatkan tingkat kelulusan dan mengurangi angka *drop out* di Program Studi Teknik Informatika.

MATERI DAN METODE

Data Mining

Data mining adalah data proses menemukan pola dalam kumpulan data besar pola dalam set data besar data bisa ini disimpan di berbagai lokasi seperti database, gudang data, seperti basis data, gudang atau fasilitas penyimpanan informasi lainnya pertambangan sangat erat kaitannya data ke berbagai bidang pengetahuan, seperti sistem basis data, pergudangan data, berskala besar, statistik, seperti mesin penambahan informasi dan komputasi kinerja tinggi. Selain itu ke pertambangan data juga dipengaruhi oleh berbagai bidang ilmiah termasuk pola pengenalan analisis data spasial citra basis data dan pemrosesan sinyal itu (Meilani et al., 2014).

Clustering

Clustering, sebagai salah satu teknik dalam Data Mining, merujuk pada proses pengelompokan data atau objek ke dalam cluster sehingga setiap cluster berisi data yang serupa dan berbeda dari cluster lainnya. *Metode clustering* yang umum digunakan adalah *K-Means clustering*, yang memiliki kekurangan terkait sensitivitas terhadap pemilihan pusat cluster awal dan pencapaian solusi lokal untuk mencapai kondisi optimal. Analisis *Cluster*, sebagai teknik multivariat, bertujuan untuk mengelompokkan objek berdasarkan karakteristiknya, di mana objek yang memiliki kesamaan paling dekat akan dikelompokkan bersama dalam satu cluster (Handoko, 2016).

Metode K-Means

K-means adalah teknik pengelompokan yang beroperasi dengan metode *partisi clustering*, di mana pengelompokan dilakukan secara bertahap. Sementara itu, pengelompokan *hierarchical clustering* dilakukan secara bertahap, di mana pada setiap iterasi

hanya ada satu pemilihan penggabungan item. Di sisi lain, dalam pengelompokan *partisi clustering*, pengelompokan item dilakukan secara acak berdasarkan centroid. Selain itu, dalam setiap iterasi dari pengelompokan *partisi clustering*, bisa terjadi lebih dari satu pemilihan item yang akan digabungkan (Muhammad, 2017).

Davies-Bouldin Index

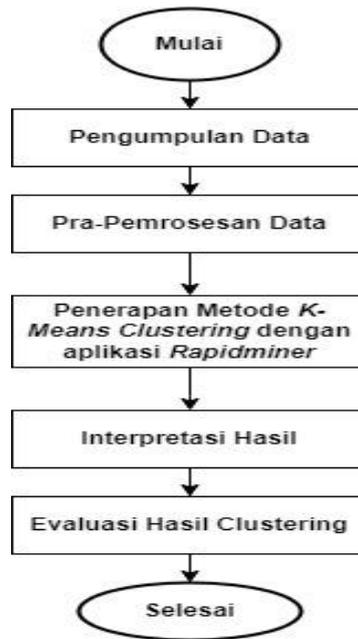
Abineno, (2022) menyatakan *Davies-Bouldin Index* merupakan alat evaluasi yang berguna dalam menilai kualitas klasterisasi data dengan mengukur tingkat keseragaman dan pemisahan antar *cluster*. Konsep dasarnya mencakup dua aspek utama, yaitu kekompakan (cohesion) dan pemisahan (separation). Kekompakan mencerminkan seberapa dekat setiap titik data dengan pusat klusternya sendiri, sementara pemisahan menggambarkan seberapa jauh pusat *cluster* satu dari yang lain. Pendekatan ini khususnya berguna dalam mengevaluasi efektivitas *K-Means*, memberikan pemahaman yang lebih dalam tentang seberapa baik data telah dikelompokkan dan seberapa jauh kluster tersebut terpisah satu sama lain. Metode ini digunakan untuk mengevaluasi performa dan mengukur akurasi dalam menggunakan metode *k-means*. Berikut merupakan langkah-langkah dalam menggunakan metode tersebut.

Pengertian Drop Out

Drop out atau keluar dari sekolah tanpa menyelesaikan pendidikan, seringkali disebabkan oleh berbagai faktor seperti kurangnya motivasi, rasa malu, atau tekanan sosial. Hal ini merupakan pemborosan investasi pendidikan yang telah dikeluarkan. Universitas juga dapat mencabut status mahasiswa jika mereka tidak memenuhi kewajiban administratif atau akademik, baik karena tidak mendaftarkan ulang atau gagal memenuhi standar penilaian. Terdapat dua jenis drop out: administratif, yang terjadi karena masalah administrasi seperti pendaftaran ulang atau pembayaran biaya, dan akademik, yang terjadi karena prestasi akademik yang tidak memadai. Ketentuan masa studi juga dapat menjadi faktor, di mana mahasiswa dapat dianggap drop out jika mereka tidak menyelesaikan studi dalam waktu yang ditentukan, misalnya tujuh tahun (Vhalla et al., 2018).

Rapidminer

RapidMiner merupakan sebuah alat atau *tool* yang digunakan dalam operasi data mining untuk menganalisis informasi. Penggunaan *RapidMiner* meliputi berbagai bidang seperti penelitian, pendidikan, *rapid prototyping*, pengembangan aplikasi, dan aplikasi dalam industri. Proyek *RapidMiner* dimulai pada tahun 2001 oleh Ralf Klinkenberg, Simon Fischer, dan Ingo Mierswa di Artificial Intelligence Group of Katharina Morik, yang berlokasi di Dortmund University of Technology. Aplikasi *RapidMiner* yang bersifat open source mencakup proses seperti pembersihan data, transformasi data, optimalisasi, validasi, dan visualisasi. Visualisasi dalam *RapidMiner* dapat berupa diagram batang, *scatter plot*, pie chart, dan lain-lain. *RapidMiner* memiliki sekitar 500 operator *data mining*, yang mencakup operator input, output, pra-pemrosesan data, dan visualisasi (Syarif, 2023).



Gambar 1 Alur Penelitian

HASIL DAN PEMBAHASAN

Mengumpulkan Data

Pengumpulan data telah dilakukan di Universitas Kristen Wira Wacana Sumba terkait data yang mencakup program studi Teknik Informatika. Data yang diperoleh berupa format *excel*. Penelitian menggunakan data mahasiswa angkatan tahun 2016 sebanyak 100 mahasiswa. Pengolahan data akan dilakukan dengan *dataset* yang akan disiapkan.

Pra Pemrosesan Data

Pra-pemrosesan data sangat penting karena setelah mengumpulkan data, perlu dilakukan pengecekan ulang terhadap data yang telah dikumpulkan, serta inisialisasi data. Berikut ini adalah data mentah yang telah disiapkan untuk diolah.

Tabel 2 Data Mentah

No	Nama	Nim	IPK	JK	DP	ADP	AK	SO	KJK	PW	RP L	AL	MD	SDP	KE T
1	Ni Putu F. C. A. Putri	2116001	3,55	1	85	80	75	80	75	71,3	75	75	70	75	0
2	Elmianti K. Here	2116002	2,61	1	29	32	70	65	0	72,05	51	65	70	65	0
3	Sisilia S. Lestari	2116003	3,33	1	55	50	75	80	75	77,6	75	80	70	80	0
4	Yesyurun R.H. Doko	2116004	3,11	0	49	65	75	80	75	77,75	75	80	75	80	0
5	William Christiansen	2116005	3,37	0	55	62	70	80	75	78,55	75	80	75	80	0
.....
99	Gilbertho R.S. R. Eda	2116108	1,7	0	50,75	34	70	10	67,25	66,4	49	65	10	23,3 3	0
100	Deriasti Tamu Ina	2116113	0,89	1	37,92	21	0	0	0	0	0	10	0	0	0

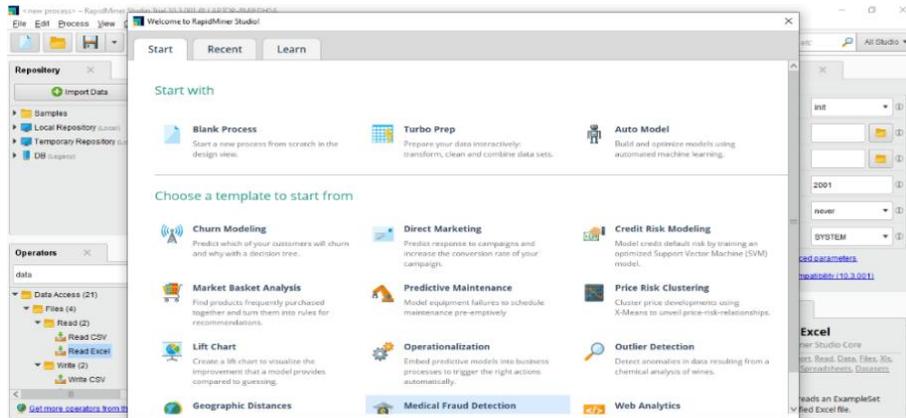
Berdasarkan data pada tabel 2 yang diperoleh dari pengumpulan data, data yang digunakan adalah data dengan atribut yang telah ditentukan. Jumlah data mahasiswa yang diperoleh adalah sebanyak 100 mahasiswa dari program studi Teknik Informatika angkatan 2016. Sebelum menerapkan metode *k-means*, dilakukan normalisasi terhadap data yang telah

dikumpulkan menggunakan rumus Min-Max. Normalisasi Min-Max mengubah nilai data ke dalam rentang [0, 1] atau rentang lain yang diinginkan.

Mengolah Data Cluster

Data yang diperoleh digunakan untuk sebagai input untuk membangun model algoritma *K-Means Clustering* menggunakan perangkat lunak *rapidminer*.

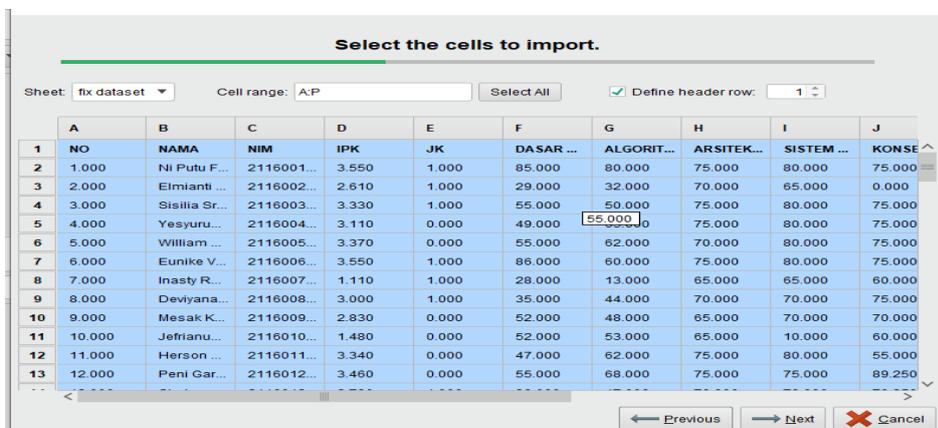
Tampilan Menu Utama *RapidMiner*



Gambar 2 Tampilan Awal Aplikasi *Rapidminer*

Tampilan awal menu utama *RapidMiner* mencakup berbagai komponen dan fitur. Salah satu fitur utamanya adalah "new process", yang merupakan langkah pertama dalam menggunakan aplikasi *RapidMiner*.

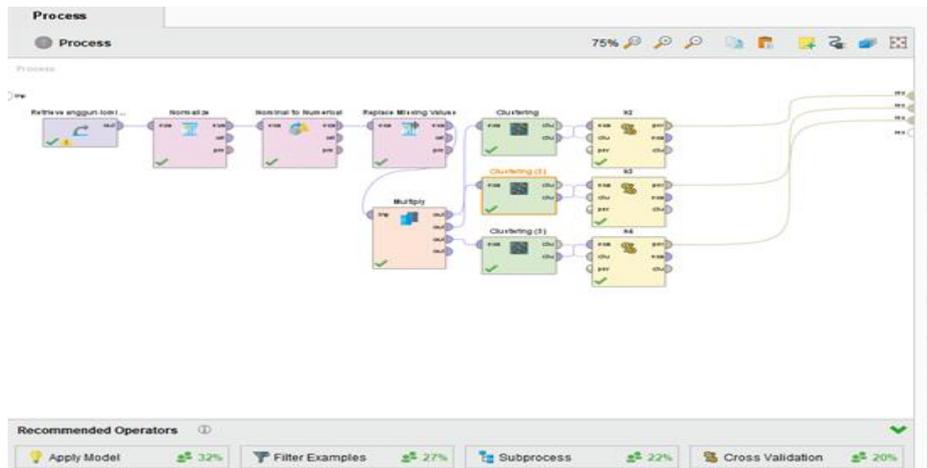
Tampilan *Select The Cells To Import*



Gambar 3 *Select The Cells To Import*

Sistem memberikan panduan tentang cara mengimpor data baru yang akan diproses lebih lanjut, dengan format data saat ini dalam bentuk *Excel*.

Tampilan *Clustering RapidMiner*



Gambar 4 Penerapan *K-Means*

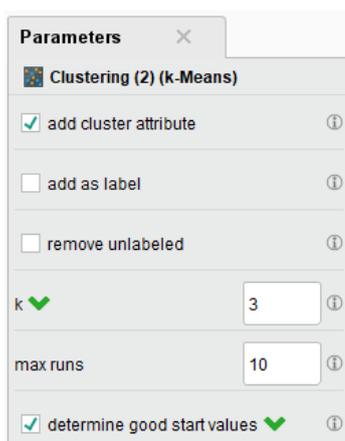
Algoritma *k-means* akan diterapkan dalam proses yang dirancang untuk melakukan klusterisasi atau pengelompokan data. Operator klusterisasi yang digunakan adalah *k-means*. Setelah model selesai dibuat, proses akan dijalankan untuk mendapatkan hasil klusterisasi.

Pada gambar 4.5 menunjukkan tampilan operator yang digunakan dalam pengklasteran. Berikut penjelasan dari setiap operator yang ditampilkan.

- 1) Operator *Retrieve* berfungsi untuk mengambil dataset dari *repository RapidMiner*. Dataset ini akan menjadi input untuk proses selanjutnya.
- 2) *Normalize* adalah operator yang digunakan untuk menormalkan data dalam sebuah *dataset*. Fungsi utama dari operator ini adalah untuk mengubah skala data agar fitur-fitur yang berbeda memiliki rentang nilai yang seragam.
- 3) *Nominal to Numerical* adalah proses konversi yang mengubah atribut atau variabel yang bersifat nominal (kategori) dalam dataset menjadi representasi numerik. Ini dilakukan untuk memfasilitasi analisis dan pemrosesan data lebih lanjut menggunakan algoritma *machine learning* atau statistik yang memerlukan data dalam format numerik.
- 4) *replace missing values* adalah proses dalam analisis data yang bertujuan untuk mengelola nilai-nilai yang tidak lengkap atau kosong dalam dataset. Fungsi utamanya adalah untuk memastikan keutuhan data sebelum dilakukan analisis lebih lanjut atau penggunaan dalam model *machine learning*. Dengan mengisi nilai-nilai yang hilang, kita dapat menghindari bias dan meningkatkan akurasi hasil analisis serta prediksi yang didasarkan pada data tersebut. Metode penggantian bisa beragam, tergantung pada karakteristik data dan tujuan analisis yang ingin dicapai.
- 5) *Multiply* adalah operator yang berfungsi untuk menggandakan *dataset* atau hasil proses. Operator ini berguna jika menggunakan *dataset* yang sama dalam beberapa proses yang berbeda.
- 6) *Clustering K-Means* adalah operator yang berfungsi untuk melakukan *clustering* menggunakan algoritma ini membagi dataset ke dalam sejumlah *cluster* berdasarkan kesamaan fitur.

Performarmace adalah operator untuk mengevaluasi hasil clustering. Operator ini menghitung berbagai metrik yang berguna untuk menilai kualitas dari hasil *clustering*

Penentu Jumlah Cluster



Gambar 5 Penentuan Jumlah Cluster

Berdasarkan Gambar 5 langkah pertama adalah memasukkan nilai K yang diinginkan. Selama iterasi sebanyak 10 kali, nilai K yang digunakan adalah 2, 3, dan 4. Tujuannya adalah untuk mencari *cluster* yang memiliki nilai *Davies-Bouldin Index* (DBI) mendekati atau yang terendah, menandakan klaster terbaik. Setelah iterasi selesai, proses klasterisasi dilakukan untuk menghasilkan output dari algoritma *K-Means* yang diimplementasikan dalam *RapidMiner*. Hasil iterasi menunjukkan bahwa nilai K terbaik adalah 2, dan hasil klasterisasi akhirnya diperoleh menggunakan algoritma *K-Means*.

Aggregat Value Clustering

Selama proses *clustering*, data diekstraksi dan di kelompokkan menjadi beberapa *cluster* berdasarkan kemiripan tertentu.

Tabel 3 Agregat Value

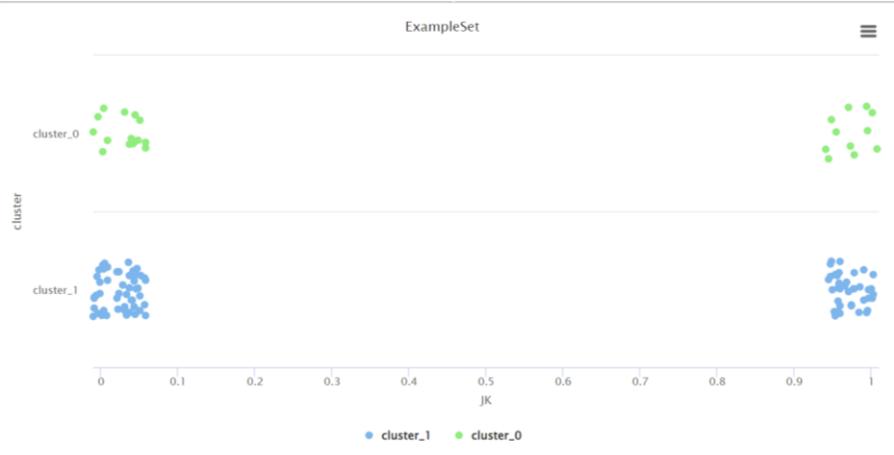
Cluster	Agregat Value	DP	ADP	AK	SO	KJK	PW	RPL	AL	MD	SDP
0	Min	12	0	0	0	0	0	0	0	0	0
	Max	75	86	70	65	0	74,1	70	75	10	53,76
	Rata-Rata	39,9	36,76	16,4	8	21,21	7,178	7,68	22,4	1,6	7,90
	Standar Deviasi	21,08	21,00	28,92	19,9	27,96	20,04	18,37	30,69	3,74	15,06
1	Min	28	21	60	10	0	0	0	65	0	0
	Max	97	95	80	80	97,25	92	90	80	80	1489
	Rata-Rata	57,7	50,2	70,1	71,4	64,4	67,17	65,17	69,5	66,59	57,28
	Standar Deviasi	16,1	17,46	3,86	11,6	14,56	19,11	18,78	4,37	14,07	155,1

Pada tabel ini menampilkan nilai-nilai agregat untuk *cluster* 0 dan cluster 1. Untuk nilai Dasar Pemograman minimum sebesar 12, sedangkan maksimum 75, nilai Algoritma dan Pemograman minimum sebesar 0. Sedangkan maksimum sebesar 86, nilai Arsitektur Komputer minimum sebesar 0, sedangkan maksimum 70, nilai Sistem Operasi minimum 0, sedangkan maksimum 65, nilai Konsep Jaringan Komputer minimum sebesar 0, sedangkan maksimum 75, nilai Pengembangan Web minimum sebesar 0, sedangkan maksimum 74,1, nilai minimum Matematika Diskrit sebesar 0, sedangkan maksimum 75, nilai minimum Aljabar Linear sebesar 0, sedangkan maksimum 10, nilai minimum Statistika dan Probalitas sebesar 0 sedangkan maksimum 53,76. Sedangkan nilai-nilai agregat untuk cluster 1, untuk minimum nilai Dasar Pemograman sebesar 28, sedangkan maksimum 97, nilai Algoritma dan Pemograman minimum sebesar 21. Sedangkan maksimum sebesar 95, nilai Arsitektur

Komputer minimum sebesar 60, sedangkan maksimum 80, nilai Sistem Operasi minimum sebesar 10, sedangkan maksimum 80, nilai Konsep Jaringan Komputer minimum sebesar 0, sedangkan maksimum 97,25, nilai Pengembangan Web minimum sebesar 0, sedangkan maksimum 92, nilai minimum Matematika Diskrit sebesar 65, sedangkan maksimum 80, nilai minimum Aljabar Linear sebesar 0, sedangkan maksimum 80, nilai minimum Statistika dan Probalitas sebesar 0 sedangkan maksimum 1489. Kita dapat melihat dapat melihat bahwa anggota claster 1 memiliki nilai rata-rata yang relatif. Analisis ini memberikan wawasan tentang bagaimana distribusi nilai di antara dua kelompok yang berbeda, di mana klaster 1 memiliki variasi yang lebih besar dalam kinerja, sementara klaster 0 lebih seragam.

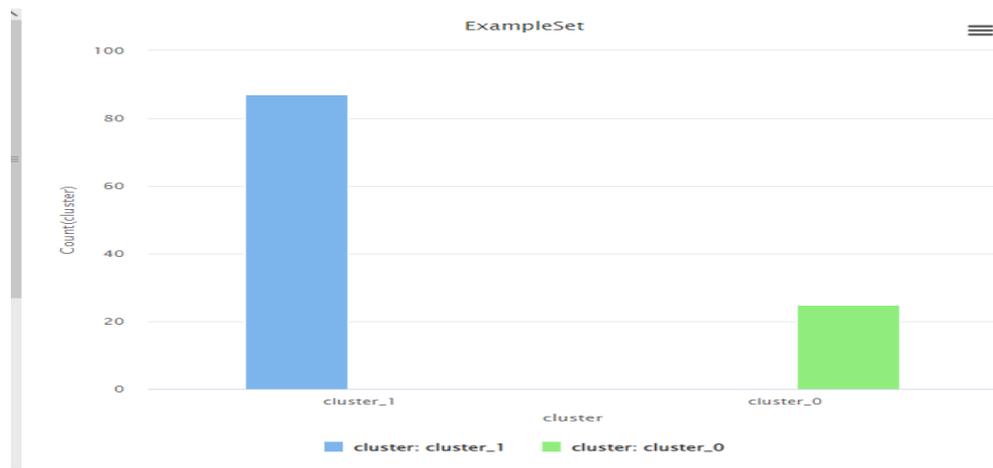
Visualisasi

Hasil pengujian data menggunakan perangkat lunak *RapidMiner* menghasilkan beberapa *output*. Dalam gambar visualisasi, terdapat garis yang menunjukkan jumlah kelompok *cluster* untuk kasus pengelompokan mahasiswa berpotensi *drop out* di program studi Teknik Informatika. Visualisasi tersebut menampilkan sumbu x, sumbu y, dan warna kustom (*cluster*) di sebelah bawah, yang semuanya terlihat dalam gambar di bawah ini. Dengan visualisasi ini, dapat melihat posisi klaster pada level 0 dan 1.



Gambar 6 Bar Plot

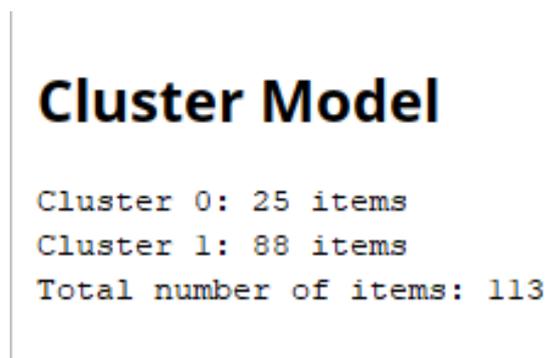
Gambar 6 menunjukkan tampilan gambar bar plot dari jenis kelamin yang sebagai variabel X dan claster sebagai variabel Y, sehingga dapat dilihat dimana wana biru diidentifikasi laki-laki dan warna hijau diidentifikasi sebagai perempuan. Kemudian pada cluster 0 merupakan mahasiswa yang berpotensi drop out bedasarkan jenis kelamin terdapat 11 mahasiswa dengan jenis kelamin laki laki dan 14 mahasiswa dengan jenis kelamin perempuan. Kemudian pada cluster 1 merupakan mahasiswa yang tidak drop out dilihat berdasarkan jenis kelamin terdapat 49 mahasiswa dengan jenis kelamin perempuan dan 39 mahasiswa dengan jenis kelamin laki-laki.



Gambar 7 Scutter Plot

Gambar 7 menunjukkan tampilan akhir dari *aplikasi K-Means* yang digunakan untuk melakukan pengelompokan data *cluster*. Berdasarkan hasil pengujian, dapat disimpulkan bahwa terdapat 25 mahasiswa dalam klaster 0 yang ditandai dengan warna hijau yang merupakan mahasiswa yang *drop out*, sementara *cluster 1* memiliki 88 mahasiswa yang ditandai dengan warna biru merupakan mahasiswa yang tidak *drop out*. Hasil dari *clustering* kemudian divisualisasikan menggunakan *Bar Plot* dan *Scatter Plot*, yang serupa dengan contoh yang terlihat pada gambar.

Cluster Model



Gambar 8 Cluster model

Dalam model (*clustering*), terdapat beberapa tampilan yang menampilkan hasil cluster, yang menunjukkan hasil pengelompokan berdasarkan *cluster* dan jumlah anggota dalam setiap cluster. *Cluster 0* memiliki 25 item, sedangkan *cluster 1* memiliki 88 item. Setelah dilakukan *clustering*, diketahui bahwa *cluster 0* berisi 25 mahasiswa yang *drop out*, sementara *cluster 1* berisi 88 mahasiswa yang tidak *drop out*. Berdasarkan data wawancara di Program Studi Teknik Informatika, terdapat 46 mahasiswa yang *drop out* sebelum dilakukan *clustering*. Hasil *clustering* yang menunjukkan 25 mahasiswa *drop out* dalam *cluster 0* karena disebabkan oleh penggunaan variabel akademik sebagai atribut dalam analisis. Namun, perlu diperhatikan bahwa ada kemungkinan mahasiswa yang *drop out* karena faktor non-akademik, seperti masalah pembayaran atau kendala lainnya yang tidak terkait dengan faktor akademik. Oleh karena itu, analisis lebih lanjut dengan mempertimbangkan faktor-faktor non-akademik juga diperlukan untuk mendapatkan gambaran yang lebih lengkap mengenai penyebab mahasiswa *drop out*.

Evaluasi Clustering

Davies-Bouldin Index (DBI) adalah sebuah ukuran yang digunakan untuk mengevaluasi kualitas *clustering*. DBI menghitung rasio antara jarak dalam *cluster* (*intra-cluster distance*) dan jarak antar-kluster (*inter-cluster distance*) dan memberikan nilai yang lebih rendah untuk *clustering* yang lebih baik

Berdasarkan Tabel 4.4 dibawah ini Melalui penggunaan metode *k-means clustering*, dan dibantu menggunakan perhitungan *Davies-bouldin index (DBI)* untuk penentuan *cluster* yang paling optimal. Didapat 3 kali percobaan iterasi untuk menentukan nilai *DBI* terbaik.

Tabel 3 *Davies Bouldan Index*

K	Avg. within centroid distance	DBI	Terpilih
2	0.585	0.873	K = 2
3	0.483	1.311	
4	0.314	1.092	



Gambar 9 Hasil Performa Klasterisasi *DBI*

Gambar 9 memberikan informasi tentang nilai *DBI* dan performa klasterisasi. Informasi ini diperoleh melalui pengujian dengan menggunakan perhitungan *Davies-Bouldin Index (DBI)* menggunakan *RapidMiner*. Hasil terbaik dari *DBI* ditemukan saat menggunakan nilai $K=2$, dengan nilai *DBI* sebesar 0.873 yang mendekati 0. Ini menunjukkan bahwa semakin kecil nilai *DBI* yang diperoleh (non-negatif ≥ 0), semakin baik klasterisasi yang terbentuk. Hasil ini diperoleh dari penggunaan metode *K-Means* dalam klasterisasi menggunakan *RapidMiner*.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, hasil pengujian data menggunakan tools *RapidMiner* pada pengelompokan mahasiswa berpotensi *Drop Out* menggunakan metode *k-means clustering* melalui variabel yang telah ditentukan menghasilkan 2 *cluster* dimana *cluster* 0 memiliki 25 mahasiswa dan *cluster* 1 memiliki 88 mahasiswa. *Cluster* 0 dapat disimpulkan bahwa dari 25 mahasiswa dimana pada *cluster* tersebut pada mata kuliah tertentu tidak memiliki nilai mata kuliah atau tidak pernah mengikuti perkuliahan tersebut sehingga berdampak pada *IPK* yang rendah dan akan terjadinya lama studi dan akan *Drop Out*. Sedangkan *cluster* 1 dapat disimpulkan bahwa mahasiswa pada setiap mata kuliah memenuhi nilainya sehingga bisa tidak terjadi *drop out* jika setiap mahasiswa terus meningkat nilainya pada setiap semesternya. Dengan mengetahui dari *cluster* tersebut program studi Teknik Informatika perlu meningkatkan sistem monitoring dan pembimbingan akademik, terutama bagi mahasiswa yang termasuk dalam *cluster* 0.

SARAN

Berdasarkan penelitian mengenai pengelompokan mahasiswa berpotensi drop out menggunakan metode *K-Means Clustering* dengan perangkat lunak *RapidMiner*, terdapat

beberapa saran untuk pengembangan lebih lanjut. Pertama, disarankan untuk mempertimbangkan faktor-faktor non-akademik, seperti masalah keuangan, kesehatan mental, atau masalah pribadi, yang mungkin berkontribusi pada potensi *drop out*. Pengumpulan data tambahan melalui wawancara, survei, atau sumber lain dapat memberikan wawasan yang lebih komprehensif. Selain itu, hasil penelitian ini dapat dijadikan dasar untuk studi selanjutnya, dengan memperluas analisis pada angkatan-angkatan berikutnya. Untuk meningkatkan kinerja klasterisasi, dapat dipertimbangkan modifikasi atau penggabungan algoritma *K-Means* dengan algoritma klasterisasi lainnya, guna memperoleh hasil yang lebih akurat dan efektif.

DAFTAR PUSTAKA

- Amin, F., Anggraeni, D. S., & Aini, Q. (2022). Penerapan Metode K-Means dalam Penjualan Produk Souq.Com. *Applied Information System and Management (AISM)*, 5(1), 7–14. <https://doi.org/10.15408/aism.v5i1.22534>
- Faozi, K. (2022). Optimasi Algoritma C4. 5 dengan Fuzzy Inference System Mamdani dalam Memprediksi Mahasiswa Berpotensi Dropout. *Scientia Sacra: Jurnal Sains, Teknologi Dan ...*, 2(3), 272–280. <http://pijarpemikiran.com/index.php/Scientia/article/view/238%0Ahttp://pijarpemikiran.com/index.php/Scientia/article/download/238/217>
- Fina Nasari, & Surya Darma, S. (2015). Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru. *Seminar Nasional Teknologi Informasi Dan Multimedia 2015*, 73–78.
- Handoko, K. (2016). Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan). *Jurnal Teknologi Dan Sistem Informasi*, 02(03), 31–40. <http://teknosi.fti.unand.id/index.php/teknosi/article/view/70>
- Irmayanti, D., Muhyidin, Y., & Nurjaman, D. A. (2021). Prediksi Mahasiswa Berpotensi Drop Out Dengan Metode Iteratif Dichotomiser 3 (ID3). *Jurnal Teknologi Informasi*, 5(2), 103–113. <https://doi.org/10.36294/jurti.v5i2.2054>
- Meilani, B. D., Susanti, N., Informatika, J. T., Informasi, F. T., Teknologi, I., & Tama, A. (2014). Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode NAÏVE BAYES. *Jurnal Sistem Informasi Universitas Suryadarma*, 3(2), 182–189. <https://doi.org/10.35968/jsi.v3i2.66>
- Mughnyanti, M., & Hafiz Nanda Ginting, S. (2023). Data Mining Manhattan Distance dan Euclidean Distance Pada Algoritma X-Means Dalam Klasifikasi Minat dan Bakat Siswa. *Remik*, 7(1), 835–842. <https://doi.org/10.33395/remik.v7i1.12162>
- Muhammad, Y. (2017). Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor Di PT Winstar Batam. *Media Informatika Budidarma*, 1(2), 28–31.
- Nugraha, R. P., Laxmi, G. F., & Riana, F. (2024). *PENERAPAN K-MEANS ++ UNTUK PENGELOMPOKAN MAHASISWA BERPOTENSI DROP OUT (STUDI KASUS : UNIVERSITAS IBN KHALDUN BOGOR)*. 8(3), 3493–3500.
- Ratniasih, N. L. (2019). Penerapan Algoritma K-Nearest Neighbour (K-Nn) Untuk Penentuan Mahasiswa Berpotensi Drop Out. *Jurnal Teknologi Informasi Dan Komputer*, 5(3), 314–318. <https://doi.org/10.36002/jutik.v5i3.804>