PENERAPAN METODE *CLUSTERING* UNTUK PENGELOMPOKAN MAHASISWA POTENSIAL *DROP OUT* MENGGUNAKAN ALGORITMA *K-MEANS* ++

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

Disusun Oleh

FAKHRI MOHAMMAD FALAHI

NIM: H02215004

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN AMPEL

SURABAYA

2019

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini.

Nama	: Fakhri Mohammad Falahi
NIM	: H02215004
Program Studi	: Matematika
Angkatan	: 2015

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "Penerapan Metode *Clustering* untuk Pengelompokan Mahasiswa Potensial *Drop Out* menggunakan Algoritma *K-Means* ++". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 8 Juli 2019

Yang menyatakan,

METERAI TEMPEL 4F5D5AFF896978613 6000 ENAM RIBU RUPIAH

(Fakhri Mohammad Falahi)
NIM. H02215004

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA : Fakhri Mohammad Falahi

NIM : H02215004

JUDUL : Penerapan Metode *Clustering* untuk Pengelompokan Mahasiswa Potensial *Drop Out* menggunakan Algoritma *K-Means* ++

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 8 Juli 2019

Dosen Pembimbing 1	Dosen Pembimbing 2
(Aris Fanani M. Kom)	(Nurissaidah Ulinnuha, M.Kom)
NIP: 198701272014031002	NIP 199901102201403200 4

LEMBAR PENGESAHAN

Skripsi oleh

NAMA : Fakhri Mohammad Falahi

NIM : H02215004

JUDUL : Penerapan Metode *Clustering* untuk Pengelompokan Mahasiswa Potensial *Drop Out* menggunakan Algoritma *K-Means ++*

Telah dipertahankan di depan tim penguji skripsi

Tanggal 11 Juli 2019

Mengesahkan,

Tim Penguji

Penguji I

(Aris Fanani M. Kom)
NIP: 198701272014031002

Penguji II

(Nurissaidah Ulinnuha, M.Kom)
NIP 199901102201403200 4

Penguji III

(Dian C. Rini Novitasari, M. Kom)
NIP: 198511242014032001

Penguji IV

(Putroue Keumala Intan, M. Si)
NIP: 198805282018012001

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

(Dr. Eni Purwati, M.Ag)
NIP: 196512211990022001

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : FAKHRI MOHAMMAD FALAHI
NIM : H0221 5009
Fakultas/Jurusan : SAINTEK / MATEMATIKA
E-mail address : fakhrimfalahi@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
☑ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :
PENERAPAN METODE CLUSTERING UNTUK PENGELOMPOKAN MAHASISWA POTENSIAL DROP OUT MENGGUNAKAN ALGORITMA K-MEANS ++

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 31 Juli 2019

Penulis

(Fakhri Mohammad Falahi)

ABSTRAK

PENERAPAN METODE *CLUSTERING* UNTUK PENGELOMPOKAN MAHASISWA POTENSIAL *DROP OUT* MENGGUNAKAN ALGORITMA *K-MEANS* ++

Pentingnya mengetahui sejak dini mahasiswa yang berpotensial *drop out* agar dapat meminimalisir mahasiswa yang terkena *drop out* yang tidak hanya berdampak pada mahasiswa saja tetapi juga berpengaruh dalam akreditasi universitas. Penelitian ini mengelompokkan mahasiswa potensial *drop out* di UIN Sunan Ampel Surabaya menggunakan *K-Means++*. *K-Means++* merupakan pengembangan dari algoritma *K-means*. Perbedaan dari algoritma *K-means* yaitu pada pemilihan nilai awal. *K-Means++* digunakan untuk meminimalisir dampak buruk dari algoritma *K-means* yang bergantung dari nilai awal. Proses pembentukan *cluster* dilakukan dengan 4 percobaan yaitu 2 *cluster*, 3 *cluster*, 4 *cluster* dan 5 *cluster*. Nilai terbesar yaitu tiga *cluster* dengan metode evaluasi *silhouette coefficient* memperoleh nilai 0,815 dan *purity* memperoleh nilai 1. Performa tersebut termasuk sangat baik karena nilai *silhouette coefficient* dan *purity* yang mendekati 1.

Kata kunci : *Kmeans++*, potensial *drop out*, *clustering*, *silhouette coefficient*, *purity*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

APPLICATION OF CLUSTERING METHOD FOR GROUPING OF POTENTIAL DROP OUT STUDENTS USING K-MEANS + ALGORITHM

The importance of knowing early on students who have the potential to drop out in order to minimize drop-out students not only affects students but also influences university accreditation. This study classified potential students to drop out at UIN Sunan Ampel Surabaya using K-Means ++. K-Means ++ is a development of the K-means algorithm. The difference from the K-means algorithm is the selection of initial values. K-Means ++ is used to minimize the adverse effects of the K-means algorithm which depends on the initial value. The cluster formation process is carried out with 4 experiments, obtained 2 clusters, 3 clusters, 4 clusters, and 5 clusters. The largest value, namely three clusters, has a Silhouette coefficient value of 0.815 and a purity value of 1. This performance was very good because the Silhouette coefficient value and purity were close to 1.

Keywords : *Kmeans++, potential to drop out, clustering, silhouette coefficient, purity*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR ISI

Halaman Judul i

Lembar Persetujuan Pembimbing ii

Lembar Pengesahan iii

Halaman Pernyataan Keaslian Karya Ilmiah iv

Halaman Motto v

Halaman Persembahan vi

Kata Pengantar vii

Abstrak ix

Abstract x

Daftar Isi xi

Daftar Tabel xiii

Daftar Gambar xv

Daftar Lampiran xvi

BAB I PENDAHULUAN 1

A. Latar Belakang 1

B. Rumusan Masalah 6

C. Tujuan Penelitian 6

D. Manfaat Penelitian 6

E. Batasan Masalah 7

BAB II KAJIAN PUSTAKA 8

A. *Drop Out* 8

B. Data Mining 9

C. *Clustering* 10

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

D. *K-Means++* 11

E. Silhouette coefficient 13

F. *Purity* 14

BAB III METODE PENELITIAN 16

A. Jenis Penelitian 16

B. Metode Pengumpulan Data 16

C. Teknik *Clustering* 16

D. Teknik Analisis Data 18

BAB IV HASIL PEMBAHASAN 19

A. Perhitungan Manual Algoritma K-Means ++ 19

B. Hasil Pembentukan *Cluster* 24

C. Perbandingan Hasil *Clustering* Berdasarkan Jumlah *Cluster* dengan Evaluasi 29

D. Interpretasi *Cluster* Terbaik 33

BAB V PENUTUP 48

A. Simpulan 48

B. Saran 49

DAFTAR PUSTAKA xviii

LAMPIRAN I

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR TABEL

Tabel 4.1 Sampel Data nilai Indeks Prestasi (IP), satuan kredit semester (SKS), dan Semester 19

Tabel 4.2 Perhitungan untuk mendapatkan *centroid* ke-2 20

Tabel 4.3 Perhitungan untuk mendapatkan *centroid* ke-3 20

Tabel 4.4 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-1 21

Tabel 4.5 Hasil Perhitungan *Centroid* Baru iterasi ke-1 21

Tabel 4.6 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-2 22

Tabel 4.7 Hasil Perhitungan *Centroid* Baru iterasi ke-2 22

Tabel 4.8 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-3 22

Tabel 4.9 Hasil Perhitungan *Centroid* Baru iterasi ke-3 23

Tabel 4.10 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-4 23

Tabel 4.11 Hasil Perhitungan *Centroid* Baru iterasi ke-4 23

Tabel 4.12 Hasil *centroid* 2 *cluster* 24

Tabel 4.13 Jumlah Data Hasil 2 *Cluster* 25

Tabel 4.14 Hasil *centroid* 3 *cluster* 25

Tabel 4.15 Jumlah Data Hasil 3 *Cluster* 26

Tabel 4.16 Hasil *centroid* 4 *cluster* 26

Tabel 4.17 Jumlah Data Hasil 4 *Cluster* 27

Tabel 4.18 Hasil *centroid* 5 *cluster* 28

Tabel 4.19 Jumlah Data Hasil 5 *Cluster* 28

Tabel 4.20 Hasil pengamatan Indeks *Silhouette coefficient*..... 30

Tabel 4.21 Hasil pengamatan *Purity* 32

Tabel 4.22 Hasil *Clustering* berdasarkan Fakultas di UINSA 33

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.23 Hasil *Clustering* berdasarkan Jalur Masuk UINSA 37

Tabel 4.24 Hasil *Clustering* berdasarkan Asal Sekolah Mahasiswa..... 40

Tabel 4.25 Hasil *Clustering* berdasarkan Penghasilan Orang tua..... 43

Tabel 4.26 Hasil Gabungan 4 Tinjauan mulai dari fakultas, jalur masuk, asal sekolah, dan penghasilan orang tua..... 46

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR GAMBAR

Gambar 2.1 Contoh *purity* 15

Gambar 3.1 Flowchart *K-Means++* 17

Gambar 4.1 Grafik Percobaan 2 *cluster* 31

Gambar 4.2 Grafik Percobaan 3 *cluster* 31

Gambar 4.3 Grafik Percobaan 4 *cluster* 31

Gambar 4.4 Grafik Percobaan 5 *cluster* 31

Gambar 4.6 Diagram Berdasarkan Fakultas di UINSA 36

Gambar 4.7 Persentase *Drop Out* bedasarkan Fakultas 37

Gambar 4.8 Diagram Berdasarkan Jalur Masuk Mahasiswa di UINSA 39

Gambar 4.9 Persentase *Drop Out* bedasarkan Jalur Masuk 40

Gambar 4.10 Diagram Berdasarkan Asal Sekolah Mahasiswa di UINSA 42

Gambar 4.11 Persentase *Drop Out* bedasarkan Asal Sekolah 43

Gambar 4.12 Diagram Berdasarkan Penghasilan Orangtua Mahasiswa di UINSA 45

Gambar 4.13 Persentase *Drop Out* bedasarkan Penghasilan Orangtua 46

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR LAMPIRAN

Lampiran 1 Sampel Data.....I-1

Lampiran 2 Source Code..... II-1

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB I

PENDAHULUAN

A. Latar Belakang

Pemanfaatan teknologi berbasis komputer selama bertahun-tahun, beberapa institusi memiliki volume data yang besar karena mengikuti perkembangan teknologi. Tujuan penyimpanan data dalam sistem komputer agar dapat dengan mudah dan akurat dalam mengoperasikannya maupun administrasinya. Perancangan data ini digunakan untuk pelaporan dan analisis data. Salah satu institusinya yaitu perguruan tinggi. Perguruan tinggi merupakan salah satu institusi yang jelas mempunyai data dengan volume yang besar. Hal ini dikarenakan adanya data mahasiswa, akademik, dan data administrasi yang disimpan di database perguruan tinggi. Data tersebut dapat menghasilkan informasi sehingga berguna untuk pengambilan keputusan.

Data yang dapat menghasilkan informasi, salah satunya yaitu pemahaman informasi yang berpotensi *drop out*. Informasi tersebut sangat dibutuhkan untuk mencegah kegagalan sehingga perlu di ketahui dan dipahami bersama khususnya bagi manajemen perguruan tinggi. Hal ini dikarenakan, informasi tersebut dapat membantu pihak perguruan tinggi untuk lebih mengenal kondisi mahasiswa sehingga dapat dijadikan dalam mengantisipasi mahasiswa *drop out*.

Minat belajar yang rendah menjadi penyebab mahasiswa *drop out*. Selain itu, kurangnya adaptasi secara akademik maupun sosial di semester awal juga menjadi

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

pemicu mahasiswa *drop out*. Sebagian mahasiswa merasakan kesulitan di semester awal ketika mereka tidak pernah mengenal mata kuliah yang sama saat menjadi siswa sehingga terjadi kegagalan pada mata kuliah dasar pada semester pertama. Tidak adanya kemandirian mahasiswa untuk belajar sendiri menyebabkan kegagalan tersebut. Kebanyakan dari mereka terbiasa mendapatkan bimbingan belajar dari lembaga-lembaga bimbingan belajar semasa menempuh pendidikan menengah atas.

Dampak dari *drop out* tidak hanya merugikan mahasiswa tetapi juga universitas. Salah satunya bagi universitas yaitu berpengaruh terhadap akreditasi universitas. Bagi mahasiswa yaitu orang tua mengeluarkan biaya dan tenaga banyak tapi hasilnya sia-sia, untuk mengurangi dampak dari *drop out* perlunya pihak universitas melakukan evaluasi dan mengetahui sejak dini yang berguna untuk meminimalisir *drop out* tersebut.

Setiap tahun universitas melakukan evaluasi terhadap mahasiswanya, dan universitas mempunyai mahasiswa yang terancam *drop out*. Dalam QS. Al-Hasyr ayat 18:

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُواْ ٱتَّقُواْ ٱللَّهَ وَلۡتَنظُرۡ نَفۡسٞ مَّا قَدَّمَتۡ لِغَدٖۖ وَٱتَّقُواْ ٱللَّهَۚ إِنَّ ٱللَّهَ خَبِيرُۢ بِمَا تَعۡمَلُونَ ١٨

Artinya: “Hai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat); dan bertakwalah kepada Allah, sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan.” (QS. Al-Hasyr : 18)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Ayat di atas menerangkan bahwa Allah memerintahkan orang-orang beriman hendaklah bertakwa kepada Allah dan hendaklah setiap manusia bermuhasabah diri dalam setiap perbuatannya seperti halnya dalam melakukan evaluasi diri dan berhati- hati dalam setiap mengambil keputusan. Universitas juga dapat melakukan evaluasi terhadap mahasiswanya salah satunya mengevaluasi dalam hal potensial *drop out.*

Keterlambatan dalam pengerjaan skripsi dan tidak memenuhi satuan kredit semester saat menempuh pendidikan menjadi salah satu penyebab mahasiswa *drop out*. Oleh karena itu, ketika melebihi 14 semester terdapat beberapa mahasiswa terpaksa terkena *drop out* karena sudah melebihi batas semester. Indeks Prestasi (IP) merupakan nilai mahasiswa yang telah ditempuh selama kuliah. Nilai dari setiap mata kuliah dijumlahkan dan dibagi dengan jumlah keseluruhan Satuan Kredit Semester (SKS) menjadi perhitungan dalam penentuan Indeks Prestasi (IP).

Berdasarkan buku pedoman akademik 2018 Universitas Islam Negeri Sunan Ampel (UINSA) Surabaya, mahasiswa strata satu menyelesaikan masa studi maksimal 14 semester. Universitas akan mengeluarkan mahasiswa dan tidak diperkenankan melanjutkan pendidikannya ketika telah mencapai batas akhir studinya. Selain itu apabila pada akhir semester keempat tidak mampu menyelesaikan minimal 40 SKS dengan IPK kurang dari 2.00, maka mahasiswa tersebut akan *drop out* oleh pihak kampus. Mahasiswa yang tidak melakukan herregistrasi satu semester dan tidak membuat surat izin cuti, maka dianggap mengundurkan diri dari kampus dan terkena sanksi *drop out* (Tim Penyusun, 2018).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Diketahui bahwa jumlah mahasiswa drop out di UINSA semakin menurun pada tahun 2012- 2016. Data terakhir pada tahun 2016, jumlah mahasiswa drop out sebanyak 426 (Mufarrohah, 2017). Jumlah tersebut membuktikan bahwa tidak sedikit mahasiswa yang drop out dari UINSA. Data mahasiswa *drop out* pada tahun 2012-2016 rata-rata mengalami penurunan. Namun pada tahun 2014-2015 terjadi kenaikan jumlah mahasiswa drop out sebesar 0,01%. Hal ini, diperlukan suatu upaya untuk menekan kenaikan tersebut.

Pihak universitas memantau perkembangan nilai prestasi akademik mahasiswa dan pengelompokan mahasiswa sesuai dengan prestasi pada kategori berbeda. Pengelompokan mahasiswa menggunakan perhitungan manual membutuhan waktu cukup banyak. Beberapa metode yang digunakan dalam pengelompokan dengan menggunakan bantuan Data Mining. Salah satu metodenya yaitu metode *clustering*. *Clustering* merupakan metode yang digunakan untuk mengelompokkan objek atau data ke dalam *cluster* (kelompok). Data dalam setiap *cluster* dibuat semirip mungkin dan jarak antar *cluster* dibuat sejauh mungkin. Algoritma *K-means* merupakan salah satu algoritma yang dapat digunakan untuk mengelompokkan data. *K-means* merupakan salah satu metode *clustering* non hirarki yang mempartisi data ke dalam satu atau lebih *cluster*. Data yang dipartisi tersebut akan memiliki kesamaan sifat dalam satu *cluster*.

Penelitian sebelumnya, (Firmansyah, Gufroni, & Rachman, 2017) melakukan penelitian tentang penggunaan *metode K-Means* pada suatu *clustering* data untuk studi kasus pengelompokan mahasiswa potensial drop out. Pada penelitian tersebut menggunakan parameter Indeks Prestasi (IP) dengan hasil 14 orang berpotesi *drop*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

out dengan Indeks Prestasi Keseluruhan (IPK). Penelitian lainnya (Ramadhani & Januarita, 2018) mengenai perbandingan antara *K-Means* dengan *K-Medoids* Memberikan hasil bahwa *K-Means* berhasil mengevaluasi lebih baik daripada *K-Medoids* dalam menangani dataset dengan ukuran kecil. Hal ini ditunjukkan dengan hasil evaluasi menggunakan Davies-Bouldin (DB) *Index* pada dataset Iris dengan menggunakan K-Means yaitu sebesar 0.662 dan pada dataset Wine menunjukkan hasil evaluasi sebesar 0.534.

Menurut (Sukarhat, Kurniati, & Shauflah, 2011) algoritma *K-Means* terdapat permasalahan dalam hal kecepatan dan akurasi. Algoritma *K-Means++* dapat mengatasi kelemahan yang terjadi pada algoritma *K-Means* yaitu menginisialisasi *centroid* dengan kondisi tertentu pada saat awal pembentukan *cluster*. Evaluasi dilakukan terhadap algoritma *K-Means++* melalui penghitungan akurasi menggunakan metode *silhouette coefficient*. Hasil yang didapat dari penelitian tersebut menunjukkan nilai akurasi algoritma *K-Means++* berkisar antara nilai medium dan strong pada data order kalibrasi. Nilai *silhouette coefficient* bervariasi pada tiap jumlah *cluster*. Nilai *silhouette coefficient* rata-rata tertinggi terdapat pada *cluster* yang berjumlah dua pada tiap-tiap jumlah dataset yang berbeda.

Berdasarkan uraian diatas, penulis terdorong untuk mengkaji dan mengambil judul "Penerapan Metode *Clustering* untuk Pengelompokan Mahasiswa Potensial *Drop Out* menggunakan Algoritma *K-Means++*".

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

B. Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah dari penelitian ini antara lain:

1. Bagaimana mengelompokkan mahasiswa yang berpotensi *drop out* menggunakan algoritma *K-Means* ++?
2. Bagaimana evaluasi pengelompokan mahasiswa yang berpotensi *drop out* menggunakan metode *silhouette coefficient* dan *purity*?

C. Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain:

1. Untuk mengelompokkan mahasiswa yang berpotensi *drop out* menggunakan algoritma *K-Means* ++
2. Untuk mengetahui evaluasi pengelompokan mahasiswa yang berpotensi *drop out* menggunakan metode *silhouette coefficient* dan *purity*.

D. Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Teoritis
   a. Untuk Prodi Matematika

      Hasil pembahasan dari penelitian ini dapat digunakan sebagai bahan perkuliahan khususnya dalam Data Mining.

   b. Untuk Peneliti

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Peneliti dapat menambah wawasan dalam penguasaan materi, serta pengalaman baru dalam melaksanakan penelitian dan menyusun karya ilmiah dalam bentuk skripsi.

2. Praktis

   a. Untuk Pihak Akademik Universitas

      Melalui penelitian ini dapat mendukung keputusan permasalahan – permasalahan terkait dengan *drop out*

E. Batasan Masalah

Dengan adanya keterbatasan waktu penulisan dan kemampuan, maka penulis membatasi permasalahan-permasalahan yang akan dibahas dalam penelitian ini. Adapun batasan-batasan tersebut antara lain:

1. Pengelompokkan mahasiswa *drop out* dalam penelitian ini merupakan mahasiswa di Universitas Islam Negeri Sunan Ampel Surabaya, angkatan 2012 - 2016.
2. Objek yang dikelompokkan kedalam *drop out* yaitu antara lain nilai Indeks Prestasi (IP) Mahasiswa, total satuan kredit semester (SKS), dan Semester yang ditempuh.
3. Algoritma yang digunakan dalam *clustering* adalah algoritma *K-Means++*.
4. Evaluasi yang digunakan yaitu metode *silhouette coefficient* dan *purity*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB II

KAJIAN PUSTAKA

A. *Drop Out*

Drop out atau dapat diartikan sebagai putus sekolah merupakan anak didik yang putus sekolah ditengah jalan atau keluar dan tidak melanjutkan sekolah lagi dikarenakan malas, malu, takut, atau hanya sekedar mengikuti temannya (Mufarrohah, 2017). Menurut (Mufarrohah, 2017) keluar dari sekolah sebelum waktunya juga dapat dikatakan sebagai *drop out*. *Drop out* demikian ini perlu dicegah, karena hal tersebut merupakan pemborosan biaya yang sudah dikeluarkan untuk sekolah.

Mencabut status kemahasiswaan yang dilakukan oleh universitas bersangkutan, disebabkan oleh beberapa hal yang telah disepakati juga merupakan pengertian dari *drop out. Drop out* sendiri terdiri atas dua bentuk, yaitu *drop out* administratif dan *drop out* akademik. *Drop out* administratif merupakan proses pencabutan status mahasiswa dikarenakan mahasiswa yang bersangkutan tidak melakukan pendaftaran ulang selama beberapa semester, atau tidak melunasi administrasi terkait dengan biaya semester. *Drop out* akademik merupakan proses pencabutan status mahasiswa dikarenakan tidak memenuhi standar nilai atau IP semester yang telah dditetapkan oleh pihak universitas dalam beberapa semester. *Drop out* juga diberlakukan bagi mahasiswa yang tidak memenuhi ketentuan masa studi, seperti tidak menyelesaikan masa studinya dalam kurun waktu tujuh tahun.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

(Guchi, 2013) berpendapat bahwa mahasiswa yang melakukan pelanggaran terkait dengan ketentuan hukum, etika, asusila, maupun terjerat kasus kriminal juga akan terkena imbas dari sistem *drop out.* Berdasarkan beberapa pengertian diatas, *drop out* merupakan keluarnya seseorang dari lembaga pendidikan sebelum dinyatakan lulus dikarenakan beberapa hal yang sudah menjadi kesepatakan bersama bagi lembaga pendidikan terkait.

B. Data Mining

Data mining atau penambangan data merupakan suatu proses yang digunakan untuk mengetahui yang berguna dari data yang bervolume besar. Data mining juga merupakan bagian dari proses KDD (Knowledge Discovery in Databases) (Santosa, 2007).

Penjabaran adanya pengetahuan didalam database atau proses penggunaan metode statistik, kecerdasan buatan, matematika dan machine learning yang digunakan untuk mengidentifikasi informasi yang bervolume besar disebut juga data mining.

Data mining merupakan salah satu bidang keilmuan untuk menangani permasalahan dalam pengambilan informasi dari database yang besar dengan menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi (Larose, 2005).

Berdasarkan beberapa definisi yang telah disampaikan, point penting terkait data mining adalah:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Jumlah data yang akan diproses bervolume cukup besar.
3. Data mining bertujuan untuk menghasilkan pola yang bermanfaat .

Data mining membutuhkan hubungan antara dua data atau lebih dalam satu dimensi. Seperti, keterkaitan pembelian satu produk dengan produk yang lain dapat digunakan untuk mengidentifikasi dimensi produk. Hubungan juga dapat diidentifikasi antara dua atau lebih objek serta dua atau lebih atribut (Ponniah, 2001).

Data mining juga memiliki *output* yang disebut penemuan pola. Misalnya, perusahaan mengumpulkan pola dari pelanggan dengan cara mengukur pelanggan yang memiliki potensi dan pelanggan yang tidak berpotensi ketika perusahaan ingin meningkatkan fasilitas kartu kredit dari pelanggan.

Definisi-definisi data mining yang telah dijelaskan di atas, beberapa diantaranya fokus terhadap proses yang dilakukan secara otomatis. Untuk mengetahui pola yang berarti dari data yang bervolume besar dibutuhkan proses analisis secara otomatis maupun semiotomatis. Proses tersebut dinamakan data mining.

C. *Clustering*

Clustering merupakan metode pengelompokan data yang dilihat dari kemiripan atau kedekatan. *Cluster* memiliki arti yang berbeda dengan kelompok.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Kelompok merupakan kondisi yang sejenis sedangkan *cluster* tidak harus sejenis tetapi pengelompokannya didasarkan pada kemiripan sampel yang ada. Misalnya, dengan menggunakan rumus jarak *euclidean.* Jarak *euclidean* merupakan jarak terdekat dalam sebuah data ke *centroid*. Guchi berpendapat dalam mengenali suatu masalah atau dalam penentuan suatu keputusan tidak selalu memberikan hasil yang sama tetapi cenderung memiliki kemiripan. Hal ini dikarenakan banyaknya pengaplikasian dalam *clustering* (Guchi, 2013).

Clustering bersifat unsupervised learning yaitu proses pembagian sekelompok data set ke dalam *cluster* berdasarkan kemiripan dari berbagai nilai atribut data set. Hal ini bertujuan untuk mempercepat waktu komputasi dengan perolehan *cluster* yang berkualitas. *Cluster* merupakan sekumpulan objek data dalam *cluster* yang sama yang memiliki kemiripan satu sama lain dan dipisahkan dengan objek-objek *cluster* lain (Guchi, 2013).

D. *K-Means++*

K-Means merupakan algoritma *clustering* yang berulang. Algoritma *K-Means* dimulai dengan pemilihan K yang dipilih secara acak, K merupakan jumlah *cluster* yang ingin dibuat yang nilainya ditentukan secara acak dan nilai tersebut dijadikan sebagai pusat dari *cluster* atau *centroid*. Selanjutnya digunakan rumus *Euclidean Distance* untuk menghitung jarak dari setiap data terhadap masing-masing *centroid* sampai setiap data tersebut ditemukan jarak yang paling dekat dengan *centroid*. Setiap data diklasifikasikan berdasarkan jarak kedekatan dengan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

centroid. Langkah-langkah tersebut dilakukan berulang sampai diperoleh kestabilan dari nilai *centroid* (Rismawan, 2008). Rumus *Euclidean Distance* yaitu:

$$D(x) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \tag{2.1}$$

$D(x) =$ nilai *euclidean distance*

$xi =$ objek 1

$yi =$ objek 2

Sedangkan *K-Means++* yaitu pengembangan dari algoritma *K-means.* Perbedaan dari algoritma *K-means* yaitu pada pemilihan nilai awal. *K-Means++* digunakan untuk meminimalisir dampak buruk dari algoritma *K-means* yang bergantung dari nilai awal (Sukarhat, Kurniati, & Shauflah, 2011). Rumus menentukan nilai awal pada *K-Means++* :

$$K = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \tag{2.2}$$

$$dengan\ x \in X$$

$D(x)^2 =$ Jarak *euclidean distance*

$\sum_{x \in X} D(x)^2$ = Jumlah jarak *euclidean distance*

Adapun langkah-langkah pada algoritma *K-Means++*adalah sebagai berikut (Fox, Emily, 2016) :

1. Pilih k pusat *cluster* pertama secara acak dari titik data.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Untuk setiap data yang diamati x, hitung jarak $D(x)$ ke pusat *cluster* terdekat.
3. Pilih *cluster* baru dari antara titik data. dengan probabilitas x dipilih proporsional dengan $D(x)^2$.
4. Ulangi langkah 2 dan 3 sampai k pusat *cluster* telah dipilih.

E. Silhouette coefficient

Silhouette coefficient merupakan metode yang menekankan pada validasi dan penafsiran set data yang dikembangkan oleh (Rousseeuw, 1987) untuk mengukur kualitas sebuah *cluster*. Selain itu, *silhouette coefficient* juga dapat digunakan untuk menandakan derajat kepemilikan dari masing-masing objek di dalam *cluster*. Penggabungan konsep *cohesion* dan *separation* digunakan untuk memvalidasi hasil *clustering*. Rumus *silhouette coefficient* yaitu:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{2.3}$$

$s(i) =$ nilai *silhouette coefficient*

$a(i) =$ rata-rata jarak suatu data i

$b(i) =$ rata rata jarak suatu data i dengan semua objek

Hasil perhitungan nilai *silhouette coefficient* bervariasi dengan rentang -1 sampai 1. Apabila nilai *silhouette coefficient* bernilai positif yaitu ($a_i < b_i$) dan a_i mendekati 0 maka nilai *clustering*-nya baik, sehingga nilai *silhouette coefficient* yang dihasilkan maksimal yaitu 1. Objek i dikatakan berada pada *cluster* yang tepat

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

apabila $SI = 1$. Sedangkan, apabila $SI = 0$ maka objek i berada diantara dua *cluster* artinya struktur yang dimiliki objek tersebut tidak jelas. Lebih lanjut, apabila $SI = -1$ maka objek tersebut memiliki nilai *overlapping* yang berakibat dimasukkan dalam *cluster* lain. SC adalah rata-rata dari SI setiap *cluster*.

Nilai rata-rata yang dimiliki oleh *silhouette coefficient* dari masing-masing data objek dalam suatu *cluster* menunjukan seberapa layak data tersebut dimasukkan dalam *cluster*. Berikut adalah nilai *silhouette coefficient* berdasarkan (Rousseeuw, 1987):

1. $0.7 < \mathrm{SC} \leq 1$ Struktur yang kuat
2. $0.5 < \mathrm{SC} \leq 0.7$ Struktur yang standar
3. $0.25 < \mathrm{SC} \leq 0.5$ Struktur yang lemah
4. $\mathrm{SC} \leq 0.25$ Tidak memiliki struktur

F. *Purity*

Purity merupakan metode yang digunakan untuk merepresentasikan anggota dari suatu *cluster* yang paling banyak sesuai disuatu kelas dengan tujuan untuk memperoleh perhitungan kemurnian dari suatu *cluster*. *Cluster* dikatakan semakin baik apabila nilai *purity* mendekati angka 1. Berikut rumus perhitungan nilai *purity* setiap *cluster* (Susanto, 2014):

$$Purity\ (\omega, C) = \frac{1}{N} \sum_{k} \max j \left| \omega_k \cap c_j \right| \qquad (2.4)$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dimana:

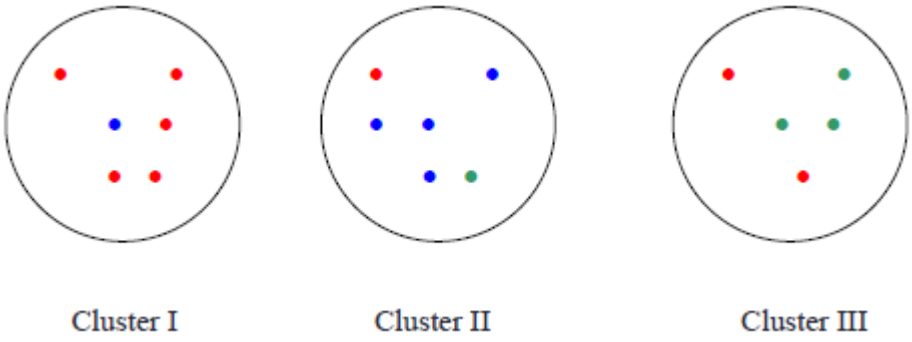
ω = himpunan *cluster*

ω_k = himpunan dokumen dalam ω_k

C = himpunan *class*

c_j = himpunan anggota dalam c_j

Contoh *purity* (Susanto, 2014):



Gambar 2.1 Contoh *purity*

Sumber: (Susanto, 2014)

Keterangan:

Titik warna merah : anggota *cluster* I

Titik warna biru : anggota *cluster* II

Titik warna hijau : anggota *cluster* III

Dapat dilihat pada Gambar 2.1 anggota pada *cluster* I sebanyak 5 titik, anggota pada *cluster* II sebanyak 4 titik, dan anggota pada *cluster* III sebanyak 3 titik. Jumlah keseluruhan data ada 17 titik.

$$purity = \frac{1}{17} \times (5 + 4 + 3) = \frac{12}{17} = 0{,}7058$$

Jadi nilai *purity* dari contoh diatas adalah 0,7058

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB III

METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini menggunakan jenis penelitian dengan pendekatan kuantitatif. Arikunto (1997) berpendapat bahwa penelitian yang dilakukan pada saat mengumpulkan data menggunakan instrumen penelitian serta dalam menafsirkan hasilnya berupa angka merupakan pengertian dari penelitian kuantitatif. Dikarenakan jenis data yang digunakan dalam skripsi ini berupa angka-angka, maka data tersebut temasuk dalam data kuantitatif.

B. Metode Pengumpulan Data

Data yang digunakan sebagai objek penelitian dalam penelitian ini merupakan data sekunder yang diperoleh dari bagian Akademik Rektorat UIN Sunan Ampel Surabaya. Data yang diambil angkatan 2012 - 2016 yaitu nilai Indeks Prestasi (IP) Mahasiswa, total Satuan Kredit Semester (SKS), dan semester yang ditempuh. Data faktor eksternalnya yaitu jalur masuk, asal sekolah dan penghasilan orang tua.

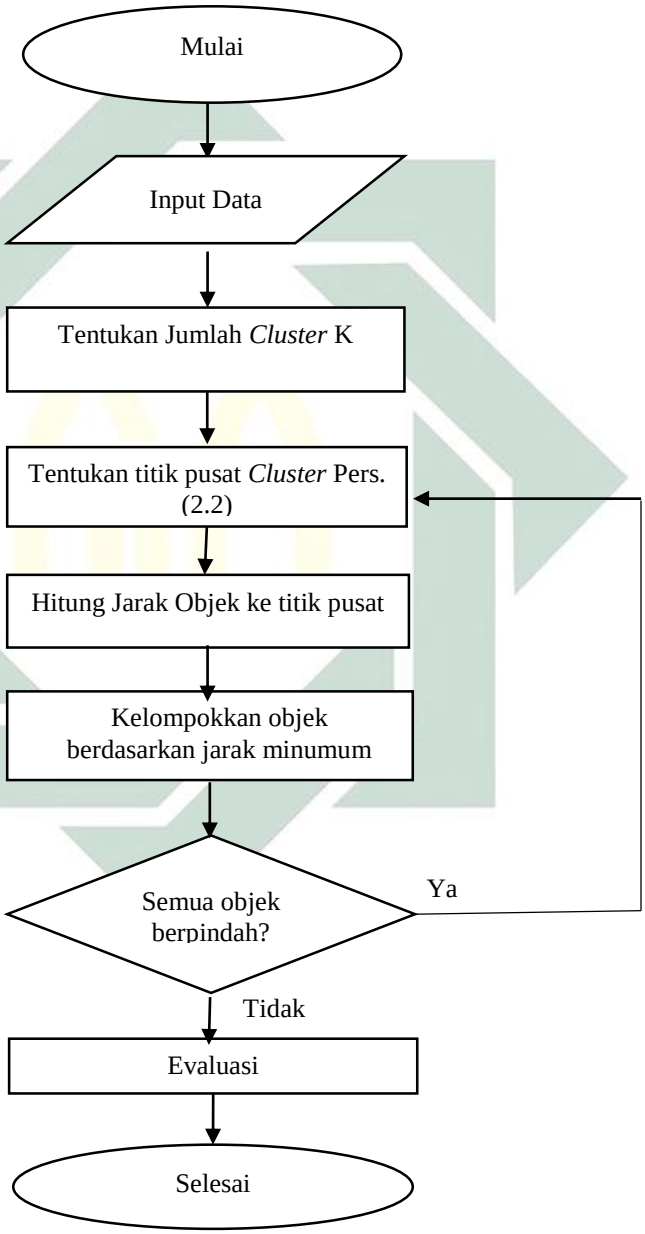
C. Teknik *Clustering*

Adapun langkah-langkah pada algoritma *K-Means++* adalah sebagai berikut:

1. Memilih K buah titik *centroid* dengan Persamaan (2.2)
2. Mengelompokkan data berdasarkan jarak minimum terhadap *centroid*.
3. Memperbarui nilai titik *centroid* dengan mencari rata-rata tiap *cluster*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4. Mengulangi langkah 2 dan 3 sampai semua objek berpindah
5. Menguji validasi menggunakan *silhouette coefficient index* di persamaan (2.3) dan *Purity* di Persamaan (2.4)



Gambar 3.1 Flowchart *K-Means++*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

D. Teknik Analisis Data

Terdapat dua teknik analisis data. Teknik pertama yaitu pemilihan model terbaik. Untuk mengetahui model *clustering* terbaik, dilakukan percobaan jumlah *cluster* (k) dimulai dari k = 2 hingga k = 5. Masing-masing model dihitung SC dan *Purity*. Model nilai SC dan nilai *purity* tertinggi merupakan model dengan performa terbaik.

Teknik kedua yaitu menginterpretasi hasil *cluster* dengan model terbaik, ditinjau dari fakultas, jalur masuk, sekolah dan penghasilan orang tua. Hal ini untuk mengetahui karakteristik tiap *cluster*. Selain itu juga untuk mengetahui faktor apa saja yang mempengaruhi tingginya tingkat DO, dan sebaliknya. Setiap variabel yang ditinjau dengan melihat persentase data setiap *cluster* kemudian dilakukan penarikan kesimpulan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB IV

HASIL PEMBAHASAN

A. Perhitungan Manual Algoritma K-Means ++

Langkah awal pada *K-Means++* adalah menentukan *centroid* pertama awal secara acak dari data lalu untuk mendapatkan *centroid* berikutnya akan menggunakan metode *K-Means++*. Berikut ini menggunakan sampel dari data yang diambil angkatan 2012 - 2016 yaitu nilai Indeks Prestasi (IP) Mahasiswa, total satuan kredit semester (SKS), dan Semester yang ditempuh. Pada contoh ini, data akan dikelompokkan menjadi tiga *cluster* sehingga dibutuhkan 3 *centroid*.

Tabel 4.1 Sampel Data nilai Indeks Prestasi (IP), satuan kredit semester (SKS), dan Semester

Data Ke-	IPK	SKS	SEMESTER
1	1.45	132	12
2	2.7	86	8
3	2.08	166	14
4	3.34	148	8
5	0	18	4
6	0.33	18	4
7	3.2	148	9
8	0	18	4
9	3.14	148	14
10	3.22	148	9

Langkah pertama, ditentukan *centroid* ke-1 dengan mengambil nilai acak dari 1 hingga 10. Didapatkan data ke-6 sebagai *centroid* ke-1.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.2 Perhitungan untuk mendapatkan *centroid* ke-2

Data Ke	D6	MIN	D(x)/∑D(x)	Cumulative
1	114.286	114.286	0.134	0.134
2	68.159	68.159	0.080	0.214
3	**148.348**	**148.348**	**0.174**	**0.388**
4	130.096	130.096	0.153	0.541
5	0.330	0.330	0.000	0.541
6	0.000	0.000	0.000	0.541
7	130.128	130.128	0.153	0.694
8	0.330	0.330	0.000	0.694
9	130.414	130.414	0.153	0.847
10	130.128	130.128	0.153	1.000

Setelah memperhitungkan jarak data dengan *centroid* 1 maka langkah yang dapat dilakukan selanjutnya yaitu menentukan nilai acak untuk memperoleh *centroid* ke 2. Didapatkan nilai acak sebesar 0,26. Diambil *centroid* baru dengan mengambil data dengan kumulatif lebih dari nilai acak yang dihasilkan 0,26. Dalam hal ini adalah data ke-3 sesuai yang ditunjukkan pada Tabel 4.3.

Tabel 4.3 Perhitungan untuk mendapatkan *centroid* ke-3

Data Ke	D6	D3	MIN	D(x)/∑D(x)	Cumulative
1	114.286	34.065	34.065	0.192	0.192
2	**68.159**	**80.227**	**68.159**	**0.384**	**0.576**
3	148.348	0.000	0.000	0.000	0.576
4	130.096	19.015	19.015	0.107	0.684
5	0.330	148.352	0.330	0.002	0.685
6	0.000	148.348	0.000	0.000	0.685
7	130.128	18.715	18.715	0.106	0.791
8	0.330	148.352	0.330	0.002	0.793
9	130.414	18.031	18.031	0.102	0.894
10	130.128	18.716	18.716	0.106	1.000

Setelah memperhitungkan jarak data dengan *centroid* 2 maka langkah yang dapat dilakukan selanjutnya yaitu menentukan nilai acak untuk memperoleh

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

centroid ke tiga. Nilai acak: 0,54. Maka, akan didapatkan nilai *centroid* ke 3 adalah Data ke-2.

Maka *centroid* yang terpilih yaitu data ke-6, data ke-3, dan data ke-2. Setelah memperoleh *centroid* 1, 2 dan 3 maka langkah yang dapat dilakukan yaitu memasukkan ke dalam kelompok yang berpotensi tinggi, sedang dan rendah.

Tabel 4.4 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-1

Data Ke-	C1	C2	C3	*Cluster*
1	13061.25	1160.40	2133.56	2
2	4645.62	6436.38	0.00	3
3	22007.06	0.00	6436.38	2
4	16925.06	361.59	3844.41	2
5	0.11	22008.33	4647.29	1
6	0.00	22007.06	4645.62	1
7	16933.24	350.25	3845.25	2
8	0.11	22008.33	4647.29	1
9	17007.90	325.12	3880.19	2
10	16933.35	350.30	3845.27	2

Setelah mengetahui letak *cluster* masing-masing data, dilakukan perhitungan rata-rata data tiap *cluster* untuk mengetahui posisi *centroid* yang baru. Hasil *centroid* baru dapat dilihat di Tabel 4.5.

Tabel 4.5 Hasil Perhitungan *Centroid* Baru iterasi ke-1

Cluster	IPK	SKS	SEM
C1	3.57	150.00	8.00
C2	2.74	148.33	11.00
C3	2.70	86.00	8.00

Maka hasil perhitungan *centroid* baru di atas yang akan digunakan ke iterasi berikutnya yaitu iterasi-2.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.6 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-2

Data Ke-	C1	C2	C3	*Cluster*
1	344.49	269.44	2133.56	2
2	4096.76	3894.45	0.00	3
3	294.22	321.54	6436.38	1
4	4.05	9.47	3844.41	1
5	17452.74	17043.28	4647.29	3
6	17450.50	17041.58	4645.62	3
7	5.14	4.32	3845.25	2
8	17452.74	17043.28	4647.29	3
9	40.18	9.27	3880.19	2
10	5.12	4.34	3845.27	2

Langkah selanjutnya menghasilkan *centroid* baru yang akan digunakan untuk iterasi berikutnya yang ditunjukkan Tabel 4.7.

Tabel 4.7 Hasil Perhitungan *Centroid* Baru iterasi ke-2

Cluster	IPK	SKS	SEM
C1	2.71	157.00	11.00
C2	2.75	144.00	11.00
C3	0.76	35.00	5.00

Hasil perhitungan *centroid* baru di atas yang akan digunakan ke iterasi berikutnya yaitu iterasi-3.

Tabel 4.8 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-3

Data Ke-	C1	C2	C3	*Cluster*
1	627.59	146.70	9458.48	2
2	5050.00	3373.00	2613.77	3
3	90.40	493.45	17243.75	1
4	90.40	25.35	12784.67	2
5	19377.34	15932.58	290.57	3
6	19375.66	15930.87	290.18	3
7	85.24	20.20	12790.97	2
8	19377.34	15932.58	290.57	3
9	90.18	25.15	12855.68	2

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

10	85.26	20.22	12791.06	2

Langkah selanjutnya menghasilkan *centroid* baru yang akan digunakan untuk iterasi berikutnya yang ditunjukkan Tabel 4.9.

Tabel 4.9 Hasil Perhitungan *Centroid* Baru iterasi ke-3

Cluster	IPK	SKS	SEM
C1	2.08	166.00	14.00
C2	2.87	144.80	10.40
C3	0.76	35.00	5.00

Hasil perhitungan *centroid* baru di atas yang akan digunakan ke iterasi berikutnya yaitu iterasi-4.

Tabel 4.10 Perhitungan *Centroid* Terdekat untuk Setiap Objek di iterasi ke-4

Data Ke-	C1	C2	C3	*Cluster*
1	1160.40	168.42	9458.48	2
2	6436.38	3463.23	2613.77	3
3	0.00	463.02	17243.75	1
4	361.59	16.22	12784.67	2
5	22008.33	16127.44	290.57	3
6	22007.06	16125.65	290.18	3
7	350.25	12.31	12790.97	2
8	22008.33	16127.44	290.57	3
9	325.12	23.27	12855.68	2
10	350.30	12.32	12791.06	2

Langkah selanjutnya menghasilkan *centroid* baru yang akan digunakan untuk iterasi berikutnya yang ditunjukkan Tabel 4.11.

Tabel 4.11 Hasil Perhitungan *Centroid* Baru iterasi ke-4

Cluster	IPK	SKS	SEM
C1	2.08	166.00	14.00

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

C2	2.87	144.80	10.40
C3	0.76	35.00	5.00

Karena tidak ada *cluster* yang berpindah sama seperti iterasi-4 maka iterasi berhenti pada iterasi ke-4 dengan hasil *centroid* di Tabel 4.10 dan Tabel 4.11. Dapat dilihat pada baris pertama yaitu C1 dengan nilai IPK 2.08, SKS yang ditempuh 166.00, dan Semester 14.00 termasuk berpotensi rendah terkena *Drop Out*. Pada baris kedua yaitu C2 dengan nilai IPK 2.87, SKS yang ditempuh 144.80, dan Semester 10.40 termasuk berpotensi sedang terkena *Drop Out*. Dan pada baris ketiga yaitu C3 dengan nilai IPK 0.76, SKS yang ditempuh 35.00, dan Semester 5.00 termasuk berpotensi tinggi terkena *Drop Out*.

B. Hasil Pembentukan *Cluster*

Pada penelitian ini terdapat lima kategori untuk menunjukkan kondisi *cluster* diantaranya yaitu Sangat Rendah (SR), Rendah (R), Sedang (S), Tinggi (T), dan Sangat Tinggi (T).

1. Hasil Pembentukan 2 *Cluster*

Tabel 4.12 Hasil *centroid* 2 *cluster*

Cluster Ke	IPK	SKS	Semester
C1 (T)	1.7939	40.1341	4.4340
C2 (R)	3.4243	140.1929	8.2890

Dapat dilihat pada Tabel 4.12 baris pertama yaitu C1 dengan nilai IPK 1.7939, SKS yang ditempuh 40.1341, dan Semester 4.4340. Pada baris kedua

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yaitu C2 dengan nilai IPK 3.4243, SKS yang ditempuh 140.1929, dan Semester 8.2890 .

Tabel 4.13 Jumlah Data Hasil 2 *Cluster*

Cluster	Jumlah
C1 (T)	2177
C2 (R)	14719
Total	16896

Jika digabungkan dengan Tabel 4.13 maka C1 dengan nilai IPK 1.7939, SKS yang ditempuh 40.1341, dan Semester 4.4340 dengan jumlah 2177 Mahasiswa berpotensi tinggi terkena *Drop Out* degan persentase 12,9%. Pada baris kedua yaitu C2 dengan nilai IPK 3.4243, SKS yang ditempuh 140.1929, dan Semester 8.289 dengan jumlah 14719 mahasiswa yang berpotensi rendah terkena *Drop Out* dengan persentase 87,1%.

2. Hasil Pembentukan 3 *Cluster*

Tabel 4.14 Hasil *centroid* 3 *cluster*

Cluster Ke	IPK	SKS	Semester
C1 (T)	1.7450	36.5908	4.2403
C2 (S)	3.4268	150.5517	9.1251
C3 (R)	3.3801	113.2610	6.2299

Dapat dilihat pada Tabel 4.14 baris pertama yaitu C1 dengan nilai IPK 1.7450, SKS yang ditempuh 36.5908, dan Semester 4.2403. Pada baris kedua yaitu C2 dengan nilai IPK 3.4268, SKS yang ditempuh 150.5517, dan Semester 9.1251. Pada baris ketiga yaitu C3 dengan nilai IPK 3.3801, SKS yang ditempuh 113.2610, dan Semester 6.2299.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.15 Jumlah Data Hasil 3 *Cluster*

Cluster	Jumlah
C1 (T)	2014
C2 (S)	10502
C3 (R)	4380
Total	16896

Jika digabungkan dengan Tabel 4.15 maka C1 dengan nilai IPK 1.7450, SKS yang ditempuh 36.5908, dan Semester 4.2403 dengan jumlah 2014 Mahasiswa berpotensi tinggi terkena *Drop Out* dengan persentase 11,9%. Pada baris kedua yaitu C2 dengan nilai IPK 3.4268, SKS yang ditempuh 150.5517, dan Semester 9.1251 dengan jumlah 10502 Mahasiswa yang berpotensi sedang terkena *Drop Out* dengan persentase 62,2%. Pada baris ketiga yaitu C3 dengan nilai IPK 3.3801, SKS yang ditempuh 113.2610, dan Semester 6.2299 dengan jumlah 4380 Mahasiswa berpotensi rendah terkena *Drop Out* dengan persentase 25,9%.

3. Hasil Pembentukan 4 *Cluster*

Tabel 4.16 Hasil *centroid* 4 *cluster*

Cluster Ke	IPK	SKS	Semester
C1 (ST)	0.948	21.863	3.559
C2 (T)	2.645	55.918	5.178
C3 (R)	3.428	150.578	9.126
C4 (SR)	3.399	113.934	6.228

Dapat dilihat pada Tabel 4.16 baris pertama yaitu C1 dengan nilai IPK 0.948, SKS yang ditempuh 21.863, dan Semester 3.559. Pada baris kedua yaitu C2 dengan nilai IPK 2.645, SKS yang ditempuh 55.918, dan Semester

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5.178. Pada baris ketiga yaitu C3 dengan nilai IPK 3.4268, SKS yang ditempuh 150.578, dan Semester 9.126. Pada baris keempat yaitu C4 dengan nilai IPK 3.399, SKS yang ditempuh 113.934, dan Semester 6.228.

Tabel 4.17 Jumlah Data Hasil 4 *Cluster*

Cluster	Jumlah
C1 (ST)	1087
C2 (T)	1006
C3(R)	10487
C4 (SR)	4316
Total	16896

Jika digabungkan dengan Tabel 4.17 maka C1 dengan nilai IPK 0.948, SKS yang ditempuh 21.863, dan Semester 3.559 dengan jumlah 1087 Mahasiswa yang berpotensi sangat tinggi terkena *Drop Out* dengan persentase 6,4%. Pada baris kedua yaitu C2 dengan nilai IPK 2.645, SKS yang ditempuh 55.918, dan Semester 5.178 dengan jumlah 1006 Mahasiswa yang berpotensi tinggi terkena *Drop Out* dengan persentase 6%. Pada baris ketiga yaitu C3 dengan nilai IPK 3.4268, SKS yang ditempuh 150.578, dan Semester 9.126 dengan jumlah 10487 Mahasiswa yang berpotensi rendah terkena *Drop Out* dengan persentase 62,1%. Pada baris keempat yaitu C4 dengan nilai IPK 3.399, SKS yang ditempuh 113.934, dan Semester 6.228 dengan jumlah 4316 Mahasiswa yang berpotensi sangat rendah terkena *Drop Out* dengan persentase 25,5%.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4. Hasil Pembentukan 5 *Cluster*

Tabel 4.18 Hasil *centroid* 5 *cluster*

Cluster Ke	IPK	SKS	Semester
C1 (ST)	0.948	21.863	3.559
C2 (T)	2.645	55.918	5.178
C3 (S)	3.425	147.938	9.110
C4 (R)	3.430	156.901	9.163
C5 (SR)	3.403	113.807	6.208

Dapat dilihat pada Tabel 4.18 baris pertama yaitu C1 dengan nilai IPK 0.948, SKS yang ditempuh 21.863, dan Semester 3.559. Pada baris kedua yaitu C2 dengan nilai IPK 2.645, SKS yang ditempuh 55.918, dan Semester 5.178. Pada baris ketiga yaitu C3 dengan nilai IPK 3.425, SKS yang ditempuh 147.938, dan Semester 9.110. Pada baris keempat yaitu C4 dengan nilai IPK 3.430, SKS yang ditempuh 156.901, dan Semester 9.163. Pada baris kelima yaitu C5 dengan nilai IPK 3.403, SKS yang ditempuh 113.807, dan Semester 6.208.

Tabel 4.19 Jumlah Data Hasil 5 *Cluster*

Cluster	Jumlah
C1 (ST)	1087
C2 (T)	1006
C3 (S)	7486
C4 (R)	3032
C5 (SR)	4285
Total	16896

Jika digabungkan dengan Tabel 4.19 maka C1 dengan nilai IPK 0.948, SKS yang ditempuh 21.863, dan Semester 3.559 dengan jumlah 1087 Mahasiswa yang berpotensi sangat tinggi terkena *Drop Out* dengan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

persentase 6,4%. Pada baris kedua yaitu C2 dengan nilai IPK 2.645, SKS yang ditempuh 55.918, dan Semester 5.178 dengan jumlah 1006 Mahasiswa yang berpotensi tinggi terkena *Drop Out* dengan persentase 6%. Pada baris ketiga yaitu C3 dengan nilai IPK 3.425, SKS yang ditempuh 147.938, dan Semester 9.110 dengan jumlah 7486 Mahasiswa yang berpotensi sedang terkena *Drop Out* dengan persentase 44,3%. Pada baris keempat yaitu C4 dengan nilai IPK 3.430, SKS yang ditempuh 156.901, dan Semester 9.163 dengan jumlah 3032 Mahasiswa yang berpotensi rendah terkena *Drop Out* dengan persentase 17,9%. Pada baris kelima yaitu C5 dengan nilai IPK 3.403, SKS yang ditempuh 113.807, dan Semester 6.208 dengan jumlah 4285 Mahasiswa yang berpotensi sangat rendah terkena *Drop Out* dengan persentase 25,4%.

C. Perbandingan Hasil *Clustering* Berdasarkan Jumlah *Cluster* dengan Evaluasi

Clustering hasil dari data dilakukan dengan menerapkan algoritma *K-Means++*. Jumlah *cluster* yang diinginkan dalam algoritma ini menjadi hal pertama yang ditentukan. Penelitian ini menggunakan empat percobaan *cluster*. Percobaan pertama dengan menggunakan 2 *cluster*, percobaan kedua menggunakan 3 *cluster*, percobaan ketiga menggunakan 4 *cluster*, dan percobaan keempat menggunakan 5 *cluster*. Untuk mengukur performa model, dihitung nilai *silhouette coefficient* dan *purity*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1. *Silhouette coefficient*

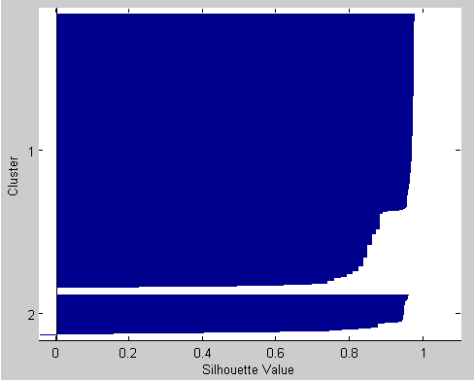
Silhouette coefficient digunakan untuk mengukur validitas hasil *clustering* dan mengindikasikan derajat kepemilikan setiap objek yang ada di dalam *cluster*. Hasil dari pengamatan *Silhouette coefficient* dapat dilihat pada Tabel 1.

Tabel 4.20 Hasil pengamatan Indeks *Silhouette coefficient*

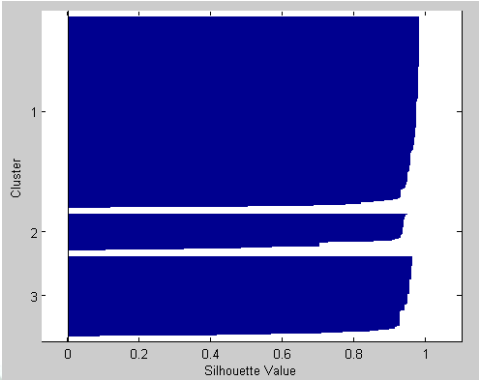
Banyaknya *Cluster*	*Indeks Silhouette coefficient*
2	0.790
3	0.815
4	0.812
5	0.631

Pada Tabel 4.20 *Indeks Silhouette coefficient* untuk percobaan *cluster* sebanyak 2 *cluster* memiliki indeks sebesar 0.790. Hal ini menunjukkan bahwa percobaan dengan 2 *cluster* tersebut memiliki rata-rata nilai *Indeks Silhouette coefficient* sebesar 0.790. Begitu juga percobaan dengan 3 *cluster* yang memiliki rata-rata nilai *Indeks Silhouette coefficient* sebesar 0.815, untuk percobaan dengan 4 *cluster* memiliki rata-rata nilai *Indeks Silhouette coefficient* sebesar 0.812, dan untuk percobaan dengan 5 *cluster* memiliki rata-rata nilai *Indeks Silhouette coefficient* sebesar 0.631. Sehingga, percobaan dengan 3 *cluster* yang paling besar berdasarkan nilainya yang memiliki rata-rata nilai *Indeks Silhouette coefficient* sebesar 0.815 dapat dilihat gambar grafik *Silhouette coefficient* dengan percobaan 2, 3, 4, dan 5.

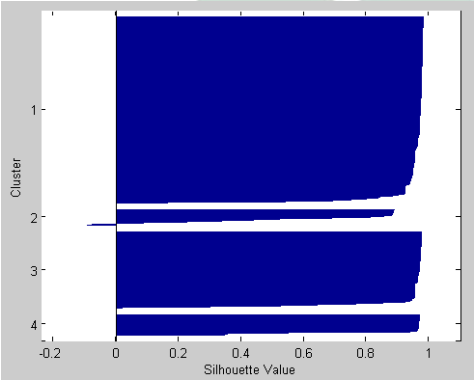
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



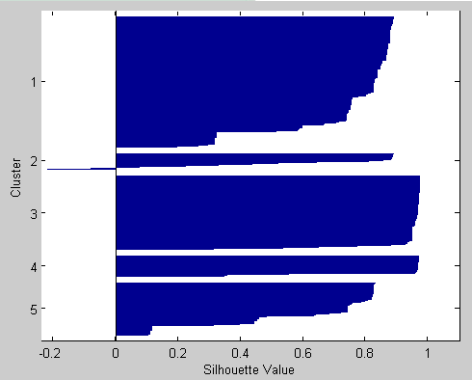
Gambar 4.1 Grafik Percobaan 2 *cluster*



Gambar 4.2 Grafik Percobaan 3 *cluster*



Gambar 4.3 Grafik Percobaan 4 *cluster*

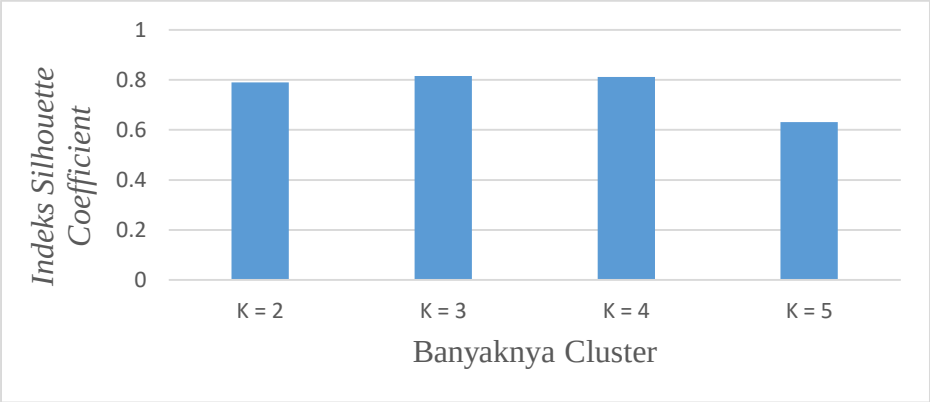


Gambar 4.4 Grafik Percobaan 5 *cluster*

Pada Gambar 4.1, 4.2, 4.3, dan 4.4 menunjukkan kualitas setiap *cluster*, sehingga dapat mengetahui *cluster* manakah yang memiliki struktur data yang kuat. Jika pada grafik terdapat hasil *cluster* yang terletak di sebelah kiri 0 maka nilai *Silhouette* kurang dari 0 atau minus. Hal tersebut membuat data *overlapping* atau tidak memiliki *cluster* yang tepat. Jadi tidak dapat disimpulkan bahwa memiliki rata-rata nilainya tinggi *cluster* tersebut baik sehingga harus dipertimbangkan dengan grafik *Silhouette* pada Gambar 4.1 hingga 4.4. Oleh karena itu dapat disimpulkan bahwa *cluster* terbaik adalah 3 *cluster*. Karena dilihat dari Gambar 4.2 hampir seluruh data mendekati 1 dan tidak ada data yang terletak disebelah kiri 0 atau minus, artinya struktur data tersebut kuat dan ditinjau

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dari rata-ratanya juga memiliki nilai tertinggi yaitu 0.815. Pada Tabel 4.20 tersebut, dapat juga digambarkan grafik menurut indeksnya.



Gambar 4.5 Grafik pengamatan *Indeks Silhouette coefficient*

Pada Gambar 4.5 grafik pengamatan *Indeks Silhouette coefficient* dan Tabel 4.21 dapat dilihat bahwa nilai *Indeks Silhouette coefficient* terbesar adalah 0.815 dengan percobaan hasil *clustering* sebanyak 3 *cluster*.

2. *Purity*

Metode evaluasi selanjutnya yaitu menggunakan *purity*. Mempresentasikan anggota *cluster* yang paling banyak sesuai (cocok) disuatu kelas untuk mengetahui kemurnian dari suatu *cluster* merupakan fungsi dari *Purity*. *Cluster* dikatakan semakin baik apabila memiliki nilai *purity* mendekati 1.

Tabel 4.21 Hasil pengamatan *Purity*

Banyaknya *Cluster*	Nilai *Purity*
2	0.7408
3	1.0000
4	0.9991
5	0.9950

Dapat dilihat pada Tabel 4.21 Dengan menggunakan *Purity* untuk percobaan *cluster* sebanyak 2 *cluster* memiliki nilai sebesar 0.7408. Begitu juga percobaan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

cluster sebanyak 3 *cluster* memiliki nilai sebesar 1, percobaan *cluster* sebanyak 4 *cluster* memiliki nilai sebesar 0.9991, percobaan *cluster* sebanyak 5 *cluster* memiliki nilai sebesar 0.9950. Berdasarkan Tabel 23. Nilai *Purity* terbesar adalah 1 dengan percobaan hasil *clustering* sebanyak 3 *cluster*.

D. Interpretasi *Cluster* Terbaik

Hasil *cluster* terbaik selanjutnya ditinjau dari fakultas, jalur masuk, sekolah dan penghasilan orang tua. Hal ini bertujuan untuk mengetahui karakteristik tiap *cluster*. Selan itu juga untuk mengetahui faktor apa saja yang mempengaruhi tingginya tingkat DO, dan sebaliknya.

1. Ditinjau dari Fakultas

   UIN Sunan Ampel Surabaya terdiri dari 9 fakultas yaitu Fakultas Adab dan Humaniora (FAHUM), Fakultas Dakwah dan Komunikasi (FDK), Fakultas Tarbiyah dan Keguruan (FTK), Fakultas Ushuludin dan Filsafat (FUF), Fakultas Syariah dan Hukum (FSH), Fakultas Sains dan Teknologi (FST), Fakultas Ekonomi dan Bisnis Islam (FEBI), Fakultas Psikologi dan Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik (FISIP)

Tabel 4.22 Hasil *Clustering* berdasarkan Fakultas di UINSA

Fakultas	*Cluster* 1 (Tinggi)	2 (Sedang)	3 (Rendah)	Total
FAHUM	291	1500	495	2286
FDK	233	1623	593	2449
FTK	451	2563	777	3791
FUF	414	1147	464	2025

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

FSH	360	2362	1005	3727
FST	79	318	428	825
FEBI	39	196	235	470
FPK	63	445	147	655
FISIP	84	348	236	668
Total	2014	10502	4380	16896

Dapat dilihat Tabel 4.22 Berdasarkan fakultas yang pertama yaitu Fakultas Adab dan Humaniora (FAHUM) dengan hasil *cluster* 1 sebanyak 291 mahasiswa, *cluster* 2 sebanyak 1500 mahasiswa, dan *cluster* 3 sebanyak 495 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Adab dan Humaniora (FAHUM) sebanyak 2286 mahasiswa. Fakultas yang kedua yaitu Fakultas Dakwah dan Komunikasi (FDK) dengan hasil *cluster* 1 sebanyak 233 mahasiswa, *cluster* 2 sebanyak 1623 mahasiswa, dan *cluster* 3 sebanyak 593 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Dakwah dan Komunikasi (FDK) sebanyak 2449 mahasiswa.

Fakultas ketiga yaitu Fakultas Tarbiyah dan Keguruan (FTK) dengan hasil *cluster* 1 sebanyak 451 mahasiswa, *cluster* 2 sebanyak 2563 mahasiswa, dan *cluster* 3 sebanyak 777 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Tarbiyah dan Keguruan (FTK) sebanyak 3791 mahasiswa. Fakultas keempat yaitu Fakultas Ushuludin dan Filsafat (FUF) dengan hasil *cluster* 1 sebanyak 414 mahasiswa, *cluster* 2 sebanyak 1147 mahasiswa, dan *cluster* 3 sebanyak 464 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Ushuludin dan Filsafat (FUF) sebanyak 2025 mahasiswa.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

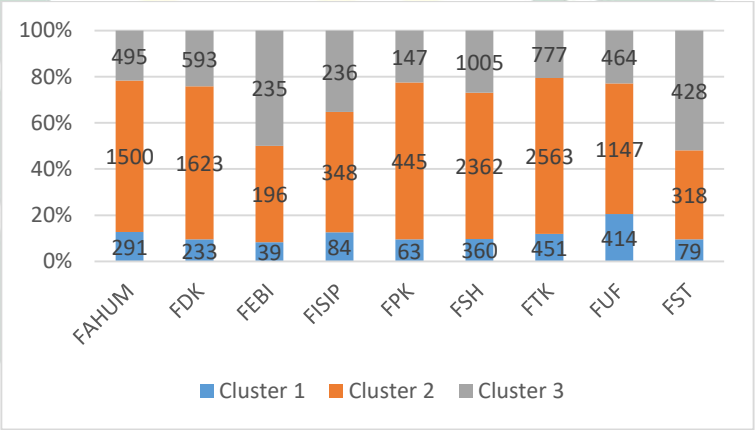
Fakultas kelima yaitu Fakultas Syariah dan Hukum (FSH) dengan hasil *cluster* 1 sebanyak 360 mahasiswa, *cluster* 2 sebanyak 2362 mahasiswa, dan *cluster* 3 sebanyak 1005 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Syariah dan Hukum (FSH) sebanyak 3727 mahasiswa. Fakultas keenam yaitu Fakultas Sains dan Teknologi (FST) dengan hasil *cluster* 1 sebanyak 79 mahasiswa, *cluster* 2 sebanyak 318 mahasiswa, dan *cluster* 3 sebanyak 428 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Sains dan Teknologi (FST) sebanyak 825 mahasiswa.

Fakultas ketujuh yaitu Fakultas Ekonomi dan Bisnis Islam (FEBI) dengan hasil *cluster* 1 sebanyak 39 mahasiswa, *cluster* 2 sebanyak 196 mahasiswa, dan *cluster* 3 sebanyak 235 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Ekonomi dan Bisnis Islam (FEBI) sebanyak 470 mahasiswa. Fakultas kedelapan yaitu Fakultas Psikologi dan Kesehatan (FPK) dengan hasil *cluster* 1 sebanyak 63 mahasiswa, *cluster* 2 sebanyak 445 mahasiswa, dan *cluster* 3 sebanyak 147 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Psikologi dan Kesehatan (FPK) sebanyak 655 mahasiswa.

Fakultas kesembilan yaitu Fakultas Ilmu Sosial dan Ilmu Politik (FISIP) dengan hasil *cluster* 1 sebanyak 84 mahasiswa, *cluster* 2 sebanyak 348 mahasiswa, dan *cluster* 3 sebanyak 236 mahasiswa. Jadi, jumlah total mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik (FISIP) sebanyak 825 mahasiswa.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Jumlah mahasiswa terbanyak adalah Fakultas Tarbiyah dan Keguruan (FTK). Meskipun demikian jika dilihat secara angka, jumlah mahasiswa terbesar yang berpotensi *drop out* rendah diraih oleh Fakultas Syariah dan Hukum (FSH). Fakultas Ekonomi dan Bisnis Islam (FEBI) sebagai fakultas baru dan fakultas dengan jumlah mahasiswa tersedikit turut menyumbang jumlah mahasiswa potensi tinggi *drop out* dengan jumlah yang paling kecil diantara fakultas lain. Untuk melihat secara proporsional, maka diperlukan persentase untuk melihat fakultas mana yang memiliki persentase tinggi, sedang dan rendah *drop out*.

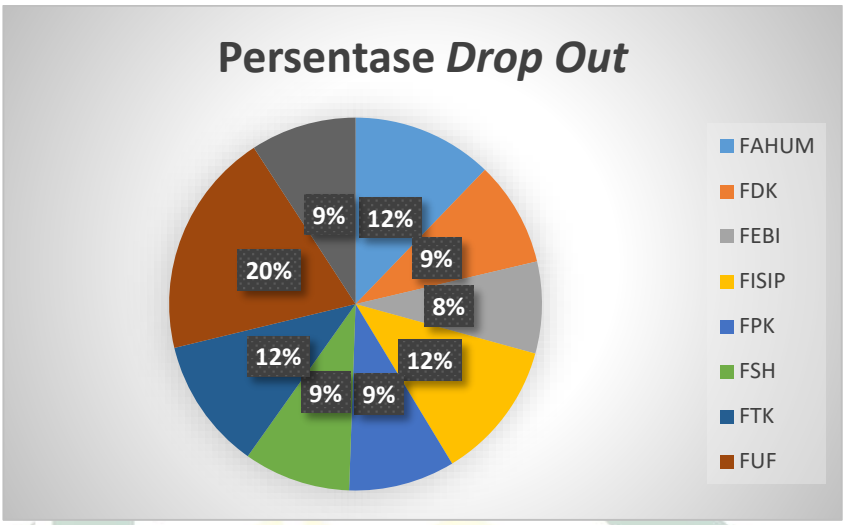


Gambar 4.6 Diagram Berdasarkan Fakultas di UINSA

Pada Gambar 4.6 dapat dilihat bahwa hasil *cluster* 1 terbesar yaitu Fakultas Ushuludin dan Filsafat (FUF) dengan persentase 20,4% mahasiswa yang berpotensi tinggi terkena *drop out*, hasil *cluster* 2 terbesar yaitu Fakultas Psikologi dan Kesehatan (FPK) dengan persentase 67,9% mahasiswa yang berpotensi sedang terkena *drop out*, dan hasil *cluster* 3 terbesar yaitu Fakultas Sains dan Teknologi (FST)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dengan persentase 51,9% mahasiswa yang berpotensi rendah terkena *drop out*.



Gambar 4.7 Persentase *Drop Out* bedasarkan Fakultas

2. Ditinjau dari Jalur Masuk

Jalur masuk merupakan jalur mahasiswa yang masuk di UIN Sunan Ampel Surabaya. Beberapa jalur masuk yang ada di UIN Sunan Ampel Surabaya yaitu Kemitraan, Mandiri, Pindahan, Prestasi, SBMPTN, UMPTAIN, Beasiswa, SNMPTN

Tabel 4.23 Hasil *Clustering* berdasarkan Jalur Masuk UINSA

Jalur Masuk	*Cluster* 1 (Tinggi)	2 (Sedang)	3 (Rendah)	Total
Kemitraan	5	79	27	111
Mandiri	718	4398	1591	6707
Pindahan	6	11	4	21
Prestasi	88	409	18	515
SBMPTN	279	1714	544	2537
UMPTAIN	481	2321	1614	4416
Beasiswa	211	197	36	444
SNMPTN	226	1373	546	2145
Total	2014	10502	4380	16896

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dilihat dari Tabel 4.23 Berdasarkan jalur masuk yang pertama yaitu Kemitraan dengan hasil *cluster* 1 sebanyak 5 mahasiswa, *cluster* 2 sebanyak 79 mahasiswa, dan *cluster* 3 sebanyak 27 mahasiswa. Jadi, jumlah total mahasiswa jalur Kemitraan sebanyak 111 mahasiswa. Jalur masuk yang kedua yaitu Mandiri dengan hasil *cluster* 1 sebanyak 718 mahasiswa, *cluster* 2 sebanyak 4398 mahasiswa, dan *cluster* 3 sebanyak 1591 mahasiswa. Jadi, jumlah total mahasiswa jalur Mandiri sebanyak 6707 mahasiswa.

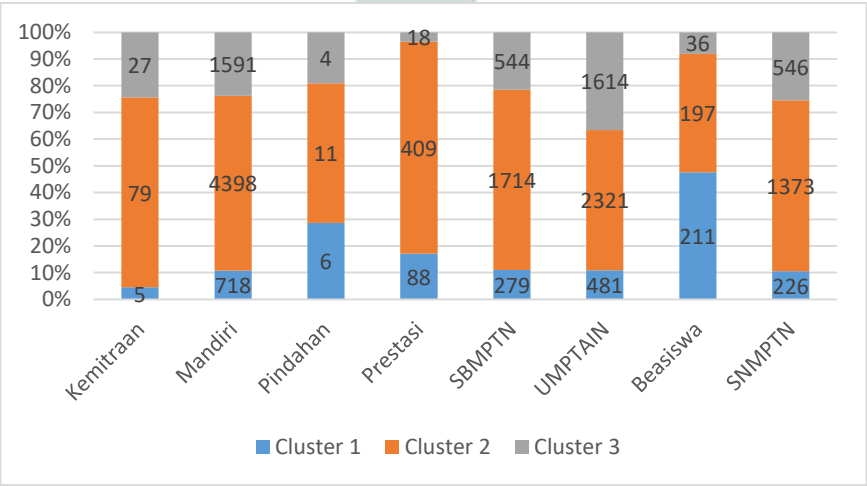
Jalur masuk ketiga yaitu Pindahan dengan hasil *cluster* 1 sebanyak 6 mahasiswa, *cluster* 2 sebanyak 11 mahasiswa, dan *cluster* 3 sebanyak 4 mahasiswa. Jadi, jumlah total mahasiswa jalur Pindahan sebanyak 21 mahasiswa. Jalur masuk keempat yaitu Prestasi dengan hasil *cluster* 1 sebanyak 88 mahasiswa, *cluster* 2 sebanyak 409 mahasiswa, dan *cluster* 3 sebanyak 18 mahasiswa. Jadi, jumlah total mahasiswa jalur Prestasi sebanyak 515 mahasiswa.

Jalur masuk kelima yaitu SBMPTN dengan hasil *cluster* 1 sebanyak 279 mahasiswa, *cluster* 2 sebanyak 1714 mahasiswa, dan *cluster* 3 sebanyak 544 mahasiswa. Jadi, jumlah total mahasiswa jalur SBMPTN 2537 mahasiswa. Jalur masuk keenam yaitu UMPTAIN dengan hasil *cluster* 1 sebanyak 481 mahasiswa, *cluster* 2 sebanyak 2321 mahasiswa, dan *cluster* 3 sebanyak 1614 mahasiswa. Jadi, jumlah total mahasiswa jalur UMPTAIN sebanyak 4416 mahasiswa.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Jalur masuk ketujuh yaitu Beasiswa dengan hasil *cluster* 1 sebanyak 211 mahasiswa, *cluster* 2 sebanyak 197 mahasiswa, dan *cluster* 3 sebanyak 36 mahasiswa. Jadi, jumlah total mahasiswa jalur Beasiswa sebanyak 444 mahasiswa. Jalur masuk kedelapan yaitu SNMPTN dengan hasil *cluster* 1 sebanyak 226 mahasiswa, *cluster* 2 sebanyak 1373 mahasiswa, dan *cluster* 3 sebanyak 546 mahasiswa. Jadi, jumlah total mahasiswa jalur SNMPTN sebanyak 2145 mahasiswa.

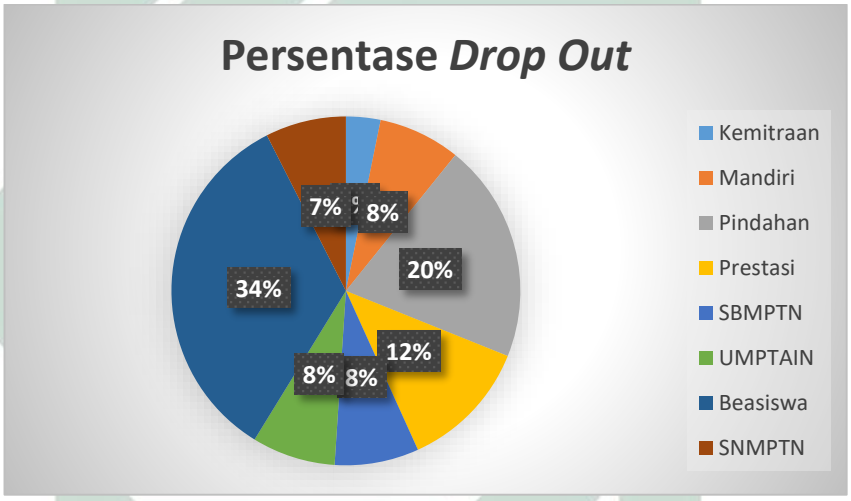
Jumlah mahasiswa terbanyak ditinjau dari jalur masuk adalah Mandiri. Jumlah mahasiswanya juga terbesar yang berpotensi *drop out* tinggi. Meskipun demikian jika dilihat secara angka, jumlah mahasiswa terbesar yang berpotensi *drop out* rendah ditinjau dari jalur masuk yaitu UMPTAIN. Untuk melihat secara proporsional, dapat diamati di Gambar 4.3



Gambar 4.8 Diagram Berdasarkan Jalur Masuk Mahasiswa di UINSA

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Pada Gambar 4.8 dapat dilihat bahwa hasil *cluster* 1 terbesar yaitu jalur masuk Beasiswa dengan persentase 47,5% mahasiswa yang berpotensi tinggi terkena *drop out*, hasil *cluster* 2 terbesar yaitu jalur masuk Prestasi dengan persentase 79,4% mahasiswa yang berpotensi sedang terkena *drop out*, dan hasil *cluster* 3 terbesar yaitu jalur masuk UMPTAIN dengan persentase 36,5% mahasiswa yang berpotensi rendah terkena *drop out*.



Gambar 4.9 Persentase *Drop Out* bedasarkan Jalur Masuk

3. Ditinjau dari Asal Sekolah

Asal sekolah merupakan asal sekolah mahasiswa sebelum masuk di UIN Sunan Ampel Surabaya diantaranya yaitu MA, SMA, SMK, Lain-lain. Lain-lain yaitu mahasiswa yang belum menginput data atau melengkapi informasi pada waktu registrasi.

Tabel 4.24 Hasil *Clustering* berdasarkan Asal Sekolah Mahasiswa

Sekolah	*Cluster*			Total
	1	2	3	

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	(Tinggi)	(Sedang)	(Rendah)	
MA	936	6067	2074	9077
SMA	876	3420	1406	5702
SMK	155	726	732	1613
Lain-Lain	47	289	168	504
Total	2014	10502	4380	16896

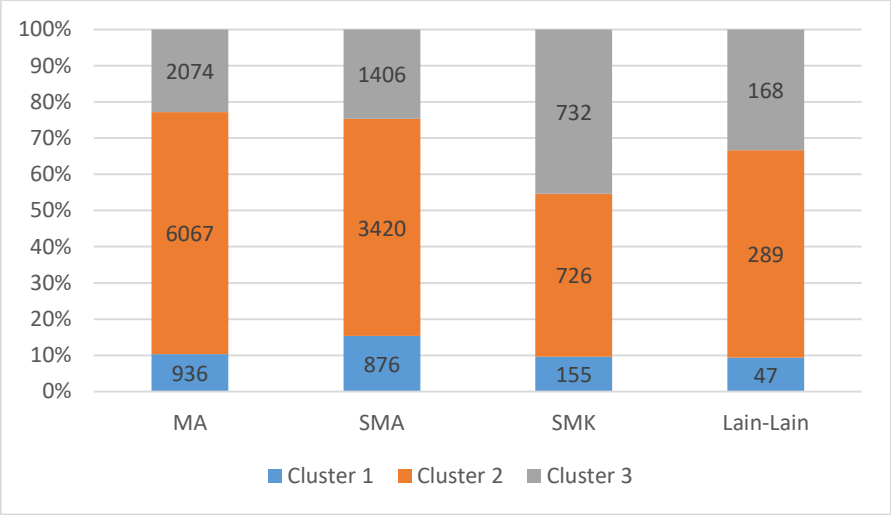
Dapat dilihat Tabel 4.24 Berdasarkan asal sekolah Mahasiswa yang pertama yaitu MA dengan hasil *cluster* 1 sebanyak 936 mahasiswa, *cluster* 2 sebanyak 6067 mahasiswa, dan *cluster* 3 sebanyak 2074 mahasiswa. Jadi, jumlah total mahasiswa asal sekolah MA sebanyak 9077 mahasiswa. Asal sekolah yang kedua yaitu SMA dengan hasil *cluster* 1 sebanyak 876 mahasiswa, *cluster* 2 sebanyak 3420 mahasiswa, dan *cluster* 3 sebanyak 1406 mahasiswa. Jadi, jumlah total mahasiswa asal sekolah SMA sebanyak 5702 mahasiswa.

Asal sekolah Mahasiswa yang ketiga yaitu SMK dengan hasil *cluster* 1 sebanyak 155 mahasiswa, *cluster* 2 sebanyak 726 mahasiswa, dan *cluster* 3 sebanyak 732 mahasiswa. Jadi, jumlah total mahasiswa asal sekolah SMK sebanyak 1613 mahasiswa. Asal sekolah yang keempat yaitu Lain-lain dengan hasil *cluster* 1 sebanyak 47 mahasiswa, *cluster* 2 sebanyak 289 mahasiswa, dan *cluster* 3 sebanyak 168 mahasiswa. Jadi, jumlah total mahasiswa asal sekolah Lain-lain sebanyak 504 mahasiswa.

Jumlah mahasiswa terbanyak ditinjau dari asal sekolah adalah MA. MA mempunyai jumlah mahasiswa terbesar di semua *cluster* yang berpotensi *drop out* tinggi, sedang dan rendah. Untuk melihat secara

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

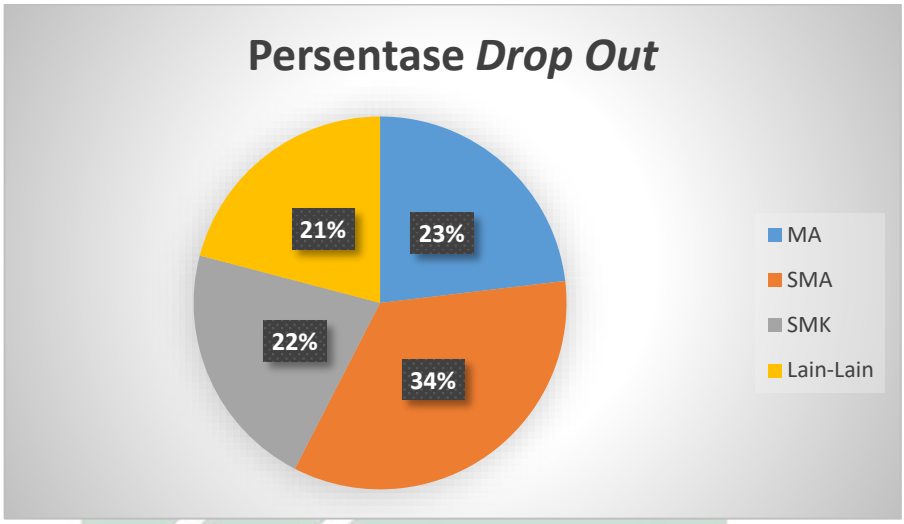
proporsional, maka diperlukan persentase untuk melihat asal sekolah manakah yang memiliki persentase tinggi, sedang, dan rendah *drop out*.



Gambar 4.10 Diagram Berdasarkan Asal Sekolah Mahasiswa di UINSA

Pada Gambar 4.10 dapat dilihat bahwa hasil *cluster* 1 terbesar yaitu asal sekolah SMA dengan persentase 15,4% mahasiswa yang berpotensi tinggi terkena *drop out*, hasil *cluster* 2 terbesar yaitu asal sekolah MA dengan persentase 66,8% mahasiswa yang berpotensi sedang terkena *drop out*, dan hasil *cluster* 3 terbesar yaitu asal sekolah SMK dengan persentase 45,4% mahasiswa yang berpotensi rendah terkena *drop out*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.11 Persentase *Drop Out* bedasarkan Asal Sekolah

4. Ditinjau dari Penghasilan Orang tua

Tabel 4.25 Hasil *Clustering* berdasarkan Penghasilan Orang tua

Penghasilan Orang Tua (dalam satuan juta)	*Cluster* 1 (Tinggi)	2 (Sedang)	3 (Rendah)	Total
< 1	741	4491	1252	6484
1-2	405	3043	1402	4850
2-3	206	1659	915	2780
3-5	201	1151	606	1958
5-7	18	100	122	240
7-10	13	28	53	94
> 10	5	30	30	65
Belum terinput	425			425
Total	2014	10502	4380	16896

Dapat dilihat Tabel 4.25 Berdasarkan penghasilan orang tua yang pertama yaitu penghasilan (< 1.000.000) dengan hasil *cluster* 1 sebanyak 741 mahasiswa, *cluster* 2 sebanyak 4491 mahasiswa, dan *cluster* 3 sebanyak 1252 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (< 1.000.000) sebanyak 6484 mahasiswa. Penghasilan orang tua yang kedua yaitu penghasilan (1.000.000 s.d. 2.000.000) dengan hasil *cluster* 1

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sebanyak 405 mahasiswa, *cluster* 2 sebanyak 3043 mahasiswa, dan *cluster* 3 sebanyak 1402 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (1.000.000 s.d. 2.000.000) sebanyak 4850 mahasiswa.

Penghasilan orang tua yang ketiga yaitu penghasilan (2.000.000 s.d. 3.000.000) dengan hasil *cluster* 1 sebanyak 206 mahasiswa, *cluster* 2 sebanyak 1659 mahasiswa, dan *cluster* 3 sebanyak 915 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (2.000.000 s.d. 3.000.000) sebanyak 2780 mahasiswa. Penghasilan orang tua yang keempat yaitu penghasilan (3.000.000 s.d. 5.000.000) dengan hasil *cluster* 1 sebanyak 201 mahasiswa, *cluster* 2 sebanyak 1151 mahasiswa, dan *cluster* 3 sebanyak 606 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (3.000.000 s.d. 5.000.000) sebanyak 1958 mahasiswa.

Penghasilan orang tua yang kelima yaitu penghasilan (5.000.000 s.d. 7.000.000) dengan hasil *cluster* 1 sebanyak 18 mahasiswa, *cluster* 2 sebanyak 100 mahasiswa, dan *cluster* 3 sebanyak 122 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (5.000.000 s.d. 7.000.000) sebanyak 240 mahasiswa. Penghasilan orang tua yang keenam yaitu penghasilan (7.000.000 s.d. 10.000.000) dengan hasil *cluster* 1 sebanyak 13 mahasiswa, *cluster* 2 sebanyak 28 mahasiswa, dan *cluster* 3 sebanyak 53 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (7.000.000 s.d. 10.000.000) sebanyak 94 mahasiswa.

Penghasilan orang tua yang ketujuh yaitu penghasilan (> 10.000.000) dengan hasil *cluster* 1 sebanyak 5 mahasiswa, *cluster* 2 sebanyak 30

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

mahasiswa, dan *cluster* 3 sebanyak 30 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua (> 10.000.000) sebanyak 65 mahasiswa. Penghasilan orang tua yang kedelapan yaitu hanya penghasilan belum terinput dengan hasil *cluster* 1 sebanyak 425 mahasiswa. Jadi, jumlah total mahasiswa yang penghasilan orang tua belum terinput sebanyak 425 mahasiswa. Penghasilan berlum terinput merupakan penghasilan orang tua mahasiswa yang belum terinput pada waktu registrasi.

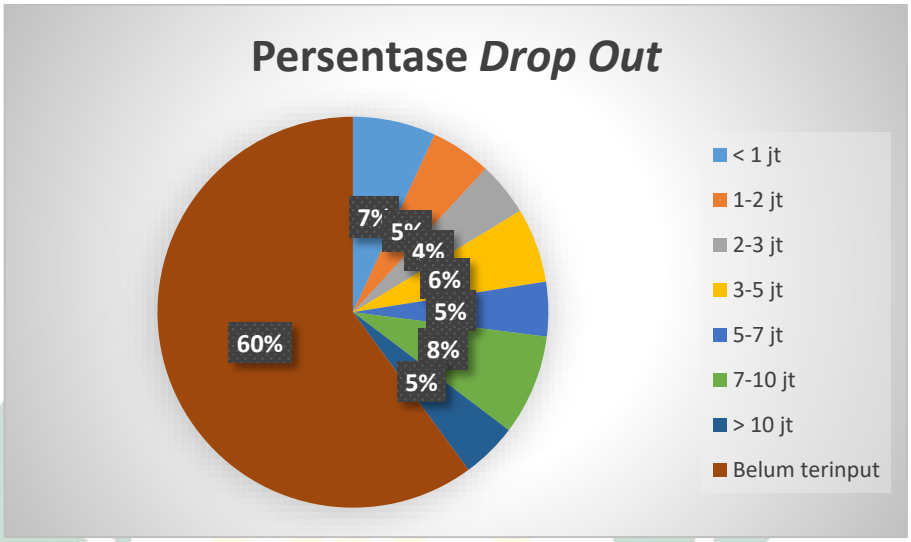
Jumlah mahasiswa terbanyak ditinjau dari penghasilan orang tua adalah penghasilan (< 1.000.000) dan berpotensi *drop out* tinggi. Jumlah mahasiswa terbesar yang berpotensi *drop out* rendah ditinjau dari penghasilan orang tua yaitu penghasilan (1.000.000 s.d. 2.000.000). Untuk melihat secara proporsional, dapat diamati di Gambar 4.7

Gambar 4.12 Diagram Berdasarkan Penghasilan Orangtua Mahasiswa di UINSA

Pada Gambar 4.12 dapat dilihat bahwa hasil *cluster* 1 terbesar yaitu penghasilan belum terinput dengan persentase 100% mahasiswa yang berpotensi tinggi terkena *drop out*, hasil *cluster* 2 terbesar yaitu penghasilan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

(< 1.000.000) dengan persentase 69,3% mahasiswa yang berpotensi sedang terkena *drop out*, dan hasil *cluster* 3 terbesar yaitu penghasilan (7.000.000 s.d. 10.000.000) dengan persentase 56,4% mahasiswa yang berpotensi rendah terkena *drop out*.



Gambar 4.13 Persentase *Drop Out* bedasarkan Penghasilan Orangtua

Untuk mengetahui karakteristiknya baik yang berpotensi tinggi maupun rendah *drop out*, akan dianalisis empat tinjauan mulai dari fakultas, jalur masuk, asal sekolah, dan penghasilan orang tua

Tabel 4.26 Hasil Gabungan 4 Tinjauan mulai dari fakultas, jalur masuk, asal sekolah, dan penghasilan orang tua

Tinjauan	Berpotensi *drop out*	
	Tinggi	Rendah
Fakultas	FUF	FST
Jalur Masuk	Beasiswa	UMPTAIN
Asal Sekolah	SMA	SMK
Penghasilan Orangtua	Belum terinput	2-3 jt

Dapat dilihat pada Tabel 4.26 maka perlunya diperbaiki yang termasuk di dalam *cluster* mahasiswa berpotensi tinggi *drop out* dalam hal ini,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

perbaikan dapat dilakukan dengan memperhatikan masing-masing tinjauan pada *cluster* yang berpotensi tinggi. Misalnya, pada tinjauan jalur masuk, seleksi beasiswa perlu diperketat untuk meminimalisir mahasiswa yang berpotensi tinggi *drop out.* Perlu dipertahankan pada masing-masing tinjauan di dalam *cluster* mahasiswa yang berpotensi rendah *drop out.* Misalnya, dilihat dari tinjauan fakultas yaitu Fakultas Sains dan Teknologi (FST) dikarenakan FST merupakan fakultas baru, perlu dilakukan pengawasan lebih intens terhadap mahasiswanya untuk mempertahankan persentase mahasiswa berpotensi rendah *drop out.*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB V

PENUTUP

A. Simpulan

Berdasarkan hasil dan pembahasan tersebut dapat disimpulkan sebagai berikut :

1. Langkah awal pada *K-Means++* adalah menentukan *centroid* pertama awal secara acak dari data lalu untuk mendapatkan *centroid* berikutnya. Setelah itu menghitung jarak terdekat dari *centroid* yang akan menghasilkan *centroid* baru yang digunakan di iterasi berikutnya. Pengelompokkan dilakukan dengan empat percobaan yakni membentuk 2, 3, 4 dan 5 *cluster*. Percobaan pertama yaitu 2 *cluster* dengan persentase 87,1% potensi rendah *drop out* dan 12,9% potensi tinggi *drop out*. Percobaan kedua yaitu 3 *cluster* dengan persentase 25,9% potensi rendah *drop out,* 62,2% potensi sedang *drop out,* dan 12,9% potensi tinggi *drop out*. Percobaan ketiga yaitu 4 *cluster* dengan persentase 25,5% potensi sangat rendah *drop out,* 62,1% potensi rendah *drop out,* 6% potensi tinggi *drop out,* dan 6,4% potensi sangat tinggi *drop out*. Percobaan 5 *cluster* dengan persentase 25,4% potensi sangat rendah *drop out,* 17,9% potensi rendah *drop out,* 44,3% potensi sedang *drop out,* 6% potensi tinggi *drop out,* dan 6,4% potensi sangat tinggi *drop out*.
2. Model terbaik dihasilkan ketika pembagian 3 *cluster* dengan nilai SC dan *purity* masing-masing sebesar 0.815 dan 1. Performa tersebut termasuk

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sangat baik karena nilai *Silhouette coefficient* dan *purity* yang mendekati 1.

B. Saran

Peneliti menyarankan pengembangan penelitian lebih lanjut terhadap "Penerapan Metode *Clustering* untuk Pengelompokan Mahasiswa Potensial *Drop Out* menggunakan Algoritma *K-Means++*" pada program studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya sebagai berikut:

1. Mengembangkan penelitian serupa dengan menggunakan metode pengelompokkan lain dari algoritma seperti *hierarchical clustering*, *partitional clustering*, *Fuzzy C-Means*, K-Modes, K-Medoids dan lain-lain agar memperoleh hasil yang lebih variatif.
2. Melakukan penelitian serupa dengan menambahkan variabel lain seperti kehadiran kuliah, kesesuaian jurusan yang diminati, dan lain-lain yang terkait dengan potensial *drop out.*
3. Saran dari hasil penelitian untuk Universitas sebaiknya dapat mengambil kebijakan terhadap mahasiswa yang berpotensi tinggi terkena *drop out*. Seperti memberikan motivasi dalam sebuah forum, memberikan solusi jika masih ada kesempatan untuk melanjutkan, dan lain-lain.
4. Untuk penelitian selanjutnya dalam evaluasi *silhouette coefficient* sebaiknya tidak hanya menggunakan nilai *mean* saja tetapi perlu melihat *varian* datanya.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR PUSTAKA

Firmansyah, A., Gufroni, A. I., & Rachman, A. N. (2017). Data Mining dengan Metode *Cluster*ing K-mean untuk Pengelompokan Mahasiswa Potensial Drop Out pada Program Studi Teknik Informatika Universitas Siliwangi. *Teknik Informatika Universitas Siliwangi Tasikmalaya*.

Guchi, N. M. (2013). Pengelompokan Mahasiswa Potensial Drop Out menggunakan Metode *Cluster*ing pada Program Studi Strata 1 Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara. *Universitas Sumatera Utara*.

K. Prilianti , & H. Wijaya. (2014). Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means *Cluster*ing. *Jurnal Cybermatika*.

Larose, D. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining : John Willey and Sons.* Inc.

Mufarrohah. (2017). Resiliensi pada mahasiswa yang terancam Drop Out dari UIN Sunan Ampel Surabaya. *UIN Sunan Ampel Surabaya*.

Ramadhani, R. D., & Januarita, D. (2018). Evaluasi K-Means dan K-Medoids pada Dataset Kecil. *Sekolah Tinggi Teknologi Telematika Telkom*.

Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.* Yogyakarta: Graha Ilmu.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Sukarhat, G. M., Kurniati, A. P., & Shauflah. (2011). Analisis dan Implementasi Algoritma *K-Means++* pada Klustering. *Telkom University*.

Susanto, B. (2014). Text and Web Mining. *FTI UKDW*.

Tim Penyusun. (2018). *Pedoman Akademik Tahun 2018.* Surabaya: UIN Sunan Ampel Surabaya.

Tuban , E., Aronson, J., & Liang, T. (2005). *Decision Support System and Intelligent Systems. Terjemahan Dwi Prabantini. Edisi 7. Jilid 1.* Yogyakarta: Andi.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id