

PENERAPAN TEKNIK CLUSTERING UNTUK PENGELOMPOKAN KONSETRASI MAHASISWA DENGAN METODE K-MEANS

Ni Luh Gede Pivin Suwirmayanti¹, I Gusti Ayu Desi Saryanti²

¹Program Studi Sistem Komputer, ²Program Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
Email: pivin@stikom-bali.ac.id¹, desi@stikom-bali.ac.id²

ABSTRACT

Determination of concentration selection is not an easy thing and will greatly affect the academic activities of students. The decision to determine concentration must be done with careful consideration so that students are not wrong in choosing the desired concentration. To determine the concentration of the appropriate majors for student students can be influenced by several factors, including academic values indicated by the value of student courses. With the selection of concentration students can find out the academic abilities they have and students more easily in taking themes for the thesis title that will be taken in the final semester. Based on these problems, then in this study using the Clustering K-Means method. Each of these algorithms has its own characteristics, with this clustering technique can classify student value data according to the similarity possessed according to the weight of the course values used, namely Computer Networking, Object Oriented Programming, Computer Organization and Physics. The value of each method used can be used as a reference for group determinants of concentration for students of Computer Systems study programs.

Keywords: *Concentration, Clustering, K-Means*

ABSTRAK

Penentuan pemilihan konsentrasi bukanlah hal yang mudah dan akan sangat berpengaruh terhadap kegiatan akademik mahasiswa. Keputusan penentuan konsentrasi harus dilakukan dengan pertimbangan yang matang agar mahasiswa tidak salah dalam memilih konsentrasi yang diinginkan. Untuk menentukan konsentrasi jurusan yang tepat bagi mahasiswa mahasiswa dapat dipengaruhi oleh beberapa faktor, antara lain nilai akademik yang ditunjukkan dengan nilai matakuliah mahasiswa. Dengan pemilihan konsentrasi mahasiswa dapat mengetahui kemampuan akademik yang dimiliki serta mahasiswa lebih mudah dalam pengambilan tema untuk judul skripsi yang akan diambil pada semester akhir. Berdasarkan dari permasalahan tersebut, maka dalam penelitian ini menggunakan metode Clustering K-Means. Setiap algoritma tersebut memiliki karakteristik masing-masing, dengan teknik clustering ini dapat mengelompokkan data nilai mahasiswa menurut kemiripan yang dimiliki menurut bobot nilai matakuliah yang digunakan yaitu matakuliah Jaringan Komputer, Pemrograman Berorientasi Obyek, Organisasi Komputer dan Fisika. Nilai masing-masing metode yang digunakan dapat dijadikan sebagai acuan informasi penentu kelompok konsentrasi bagi mahasiswa prodi Sistem Komputer.

Kata kunci: Konsetrasi, Clustering, K-Means

1. Pendahuluan

Sekolah Tinggi Manajemen Informatika dan Teknik Komputer STIKOM Bali, yang lebih dikenal dengan nama STIKOM Bali, dimana STIKOM Bali merupakan salah satu kampus IT yang ada di Bali. STIKOM Bali memiliki 3 program studi, salah satunya adalah Program Studi Sistem Komputer. Program studi Sistem Komputer STMIK STIKOM Bali pada dasarnya memiliki tujuan mencetak individu atau lulusan yang kompeten dan mampu dalam hal analisis, merancang, membangun serta mengoperasikan perangkat

keras/*hardware*, maupun perangkat lunak / *software*) ditambah lagi pengelolaan sistem komputer yang baik digunakan untuk riset maupun bisnis.

Penentuan Konsentrasi merupakan salah satu hal yang sangat vital dalam masa studi. Bagi mahasiswa penentuan pemilihan konsentrasi bukanlah hal yang mudah dan akan berpengaruh terhadap kegiatan akademik mahasiswa selanjutnya. Mahasiswa harus mempunyai pertimbangan yang matang sesuai bakat yang dimiliki dalam memutuskan penentuan konsentrasi, agar mahasiswa tidak salah dalam memilih konsentrasi yang diinginkan. Tidak sedikit mahasiswa yang belum mengetahui bakat, minat dan kemampuan yang dimilikinya. Kesalahan dalam memilih konsentrasi akan merugikan mahasiswa baik dalam hal waktu, tenaga, biaya, menurunnya kualitas prestasi mahasiswa, serta mempengaruhi pengerjaan tugas akhir mahasiswa

Untuk menentukan konsentrasi yang tepat bagi mahasiswa mahasiswa dapat dipengaruhi oleh beberapa faktor, antara lain nilai akademik yang ditunjukkan dengan nilai matakuliah mahasiswa. Dengan pemilihan konsentrasi mahasiswa dapat mengetahui kemampuan akademik yang dimiliki serta mahasiswa lebih mudah dalam pengambilan tema untuk judul skripsi yang akan diambil pada semester akhir. Dalam penelitian ini konsentrasi pilihan pada Prodi Sistem Komputer dibagi menjadi dua konsentrasi, dengan adanya konsentrasi tersebut mahasiswa diwajibkan untuk memilih salah satu dari konsentrasi yang ada. Penentuan pengelompokan konsentrasi ini akan dilaksanakan pada akhir semester 4 atau tepatnya masa perkuliahan untuk semester 5.

Teknik clustering merupakan salah satu metode pengelompokan data yang cukup banyak digunakan hingga saat ini. Analisis pengelompokan data ini telah banyak diterapkan dalam berbagai bidang seperti untuk analisis data medis, analisis pasar, analisis gambar maupun video, dan bidang-bidang lainnya. Clustering merupakan pengelompokan data yang bersifat *unsupervised* yaitu analisis pada data masukan tanpa label sehingga dapat mengetahui suatu pola tersembunyi dari data tersebut.

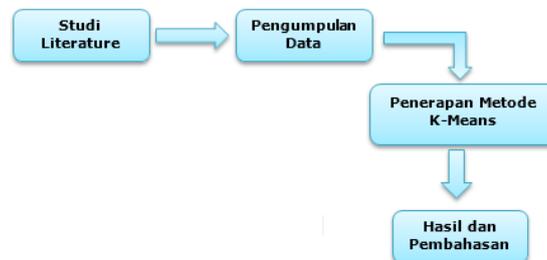
Beberapa penelitian yang telah dilakukan sebelumnya terkait dengan judul yang diangkat diantaranya Penelitian terkait dengan judul "Implementasi Data Mining Clustering Data Mahasiswa Teknik Informatika Menggunakan Algoritma K-Means" tahun 2018 oleh Turkhamun Adi K, Bosar Panjaitan, Rama Guntara dalam penelitiannya dibahas mengenai penentuan sebuah konsentrasi jurusan pada perguruan tinggi merupakan hal yang sangat vital dalam mendukung terselesaikannya proses study tepat waktu dan masa depan mahasiswa. Untuk menentukan konsentrasi jurusan yang tepat bagi mahasiswa mahasiswa dapat dipengaruhi oleh beberapa faktor, antara lain nilai akademik yang ditunjukkan dengan nilai rapor, tes kemampuan yang ditunjukkan oleh tes psikologi, dan prestasi. Dengan melakukan clustering data mining terhadap data mahasiswa menggunakan algoritma K-Means diharapkan dapat menjadi alternative solusi bagi manajemen perguruan tinggi dalam membuat strategi kebijakan.

Penelitian terkait lainnya yaitu "Pengelompokan Minat Belajar Mahasiswa Menggunakan Teknik Data Mining Dengan Metode Clustering" tahun 2015 oleh Andri dan Marlindawati dari Universitas Bina Darma, Dalam penelitiannya disebutkan bahwa data mining yang berhubungan dengan akademik mahasiswa untuk dapat mengetahui minat belajar dari mahasiswa yang ada dalam perguruan tinggi. Penelitian ini akan menghasilkan sebuah model yang dapat digunakan untuk melakukan pengelompokan mahasiswa berdasarkan minat belajarnya. Teknik Data mining yang digunakan dalam penelitian ini adalah teknik clustering.

Berdasarkan hal tersebut, maka dalam penelitian ini menggunakan metode Clustering K-Means dalam penentuan konsentrasi mahasiswa prodi Sistem Komputer. Dengan teknik clustering ini dapat mengelompokan data nilai mahasiswa menurut kemiripan yang dimiliki menurut bobot nilai matakuliah tertentu seperti matakuliah Jaringan Komputer, Pemrograman Berorientasi Obyek, Organisasi Komputer dan Fisika. Diharapkan nilai masing-masing metode ini dapat digunakan sebagai acuan informasi penentu kelompok konsentrasi bagi mahasiswa prodi Sistem Komputer.

2. Metode

Penelitian merupakan investigasi yang sistematis, terkontrol, empiris dan kritis dari suatu proposal hipotesis mengenai hubungan tertentu antar fenomena. Penelitian disini bertujuan untuk memberikan kontribusi terhadap ilmu pengetahuan dalam memecahkan masalah dengan menggunakan metode-metode yang sesuai. Tahapan-tahapan dari metode penelitian yang akan dilakukan penulis untuk pembuatan penelitian ini, diantaranya sebagai berikut :



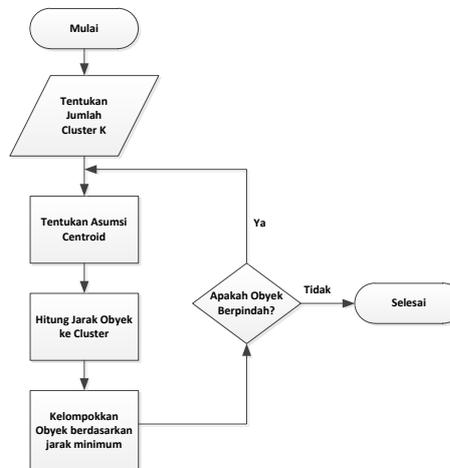
Gambar 1. Alur Analisis

1. Studi pustaka merupakan pengumpulan data ataupun pengumpulan referensi mengenai teori yang nantinya dijadikan sebagai pedoman-pedoman sesuai dengan penelitian yang diangkat. Manfaat dari studi literatur adalah dapat memberikan gambaran menyeluruh mengenai sejauh mana perkembangan penelitian-penelitian yang terkait dengan penelitian yang akan diambil. Dalam penelitian ini referensi diperoleh dari jurnal, artikel laporan penelitian, dan buku yang berhubungan mengenai penelitian yang terkait dengan teknik clustering dengan Metode K-Means.
2. Pengumpulan Data
Berdasarkan sumbernya, data penelitian dapat dikelompokkan dalam dua jenis yaitu data primer dan data sekunder. Pemahaman terhadap kedua jenis data di atas diperlukan sebagai landasan dalam menentukan teknik serta langkah-langkah pengumpulan data penelitian. disini menggunakan Data Kuantitatif Yang merupakan data yang berbentuk angka atau bilangan. Sesuai dengan bentuknya, data kuantitatif dapat diolah atau dianalisis menggunakan teknik perhitungan matematika atau statistika.
3. Clustering dengan Metode K-Means
Metode yang diusulkan dalam penelitian ini adalah metode Metode K-Means. Dimana dalam penelitian ini akan mencoba melakukan pengelompokan data konsentrasi mahasiswa pada prodi sistem komputer . Penentuan pengelompokan konsentrasi ini akan dilaksanakan pada akhir semester 4 atau tepatnya masa perkuliahan untuk semester 5 dengan teknik clustering ini dapat mengelompokan data nilai mahasiswa menurut kemiripan yang dimiliki menurut bobot nilai matakuliah Jaringan Komputer, Pemrograman Berorientasi Obyek, Organisasi Komputer dan Fisika. Hasilnya nantinya dapat menghasilkan kelompok konsentrasi bagi mahasiswa prodi Sistem Komputer. Algoritma K-Means clustering merupakan algoritma iteratif dengan meminimalkan jumlah kuadrat error antara vektor objek dengan pusat kluster terdekatnya, yaitu:

$$\sum_{j=1}^k \sum_{x \in \pi_j} \|x - m_j\|^2$$

m_j adalah pusat cluster (mean vector) dalam kluster ke j . Proses dimulai dengan memilih secara random k buah data sebagai pusat cluster awal. Pada tahap awal, algoritma K-Means memilih secara acak k buah data sebagai centroid. Centroid baru akan ditentukan bila semua data telah ditempatkan dalam cluster terdekat.

Proses penentuan centroid dan penempatan data dalam cluster diulangi sampai nilai centroid konvergen (centroid dari semua cluster tidak berubah lagi). Data clustering menggunakan metode K-Means ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

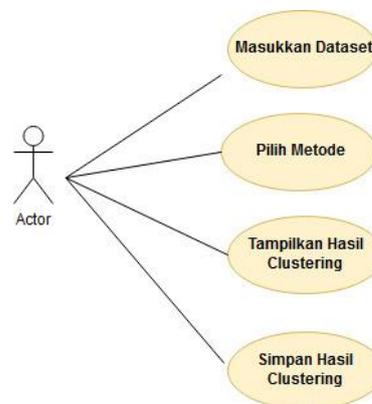


Gambar 2. Alur Algoritma K-Means

3. Hasil dan Pembahasan

Desain Sistem

Desain sistem menggunakan Use case diagram untuk menggambarkan fungsi sistem yang ada pada sistem pengelompokan data konsentrasi mahasiswa prodi sistem komputer dengan menerapkan metode K-Means Clustering, dapat dilihat pada gambar berikut:



Gambar 3. Use Case Diagram Sistem

Penerapan Metode Clustering K-Means

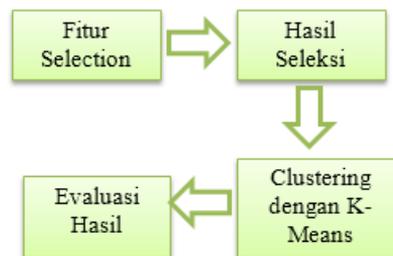
Algoritma menerima jumlah cluster untuk mengelompokkan data ke dalamnya, dan dataset yang akan dicluster sebagai nilai input. Kemudian menentukan jumlah cluster. Setiap cluster yang terbentuk akan dihitung nilai rata-ratanya. Rata-rata dari suatu cluster adalah rata-rata dari semua record yang terdapat di dalam cluster tersebut. Setelah itu dialokasikan data ke dalam cluster secara random. Selanjutnya hitung centroid /rata-rata dari data yang ada di masing-masing cluster. Alokasikan masing-masing data ke centroid /rata-rata terdekat. Kemudian Uangi langkah-langkah sebelumnya sampai terbentuk cluster-cluster stabil dan prosedur K-Means selesai. Cluster stabil terbentuk ketika iterasi atau perulangan dari K-Means tidak membuat cluster baru sebagai pusat cluster atau nilai rata-rata aritmatika dari semua cluster baru sama dengan cluster lama. Terdapat beberapa teknik untuk menentukan kapan suatu cluster stabil terbentuk atau kapan algoritma K-Means berakhir.

A	B	C	D	E	F	G	H	I
NIM	NAMA	JARKOM #1	PBO #1	ORKOM #1	JARKOM #2	PBO #2	FISIKA# 2	
160010123	I KOMANG YUDI DARMAI	A	B	A	A	B	B	
160010264	I KADEK YOGI PRAYOGA	A	A	B	A	A	A	
160010324	BOBBY VALENTINO HERMAWAN	A	B	BC	A	B	AB	
160010026	I MADE GEO ARYASA PUTRA	A	B	A	A	B	B	
160010042	I GEDE ARYA PREDHI KARYADNYA	A	AB	A	A	AB	AB	
160010059	A A SATRIA WIRAWAN	A	A	C	A	A	BC	
160010144	I WAYAN TEGUH PRATAMA	A	C	C	A	C	BC	
160010179	RENDI KHOIRULHUDA	A	AB	A	A	AB	A	
160010231	ANGGIE RAPHA SUCIPTO	A	AB	AB	A	AB	A	
160010038	I MADE ADI DARMA YASA	A	B	C	A	B	B	
160010087	I WAYAN RAHMAWAN	A	BC	C	A	BC	C	
160010263	FADHIL AFIF	A	B	B	A	B	B	
160010009	I MADE SOMANDIKA YANA	A	AB	A	A	AB	A	
160010041	I MADE BAYU SUDARSANA WIGUNA	A	A	A	A	A	AB	
160010090	A A GDE AGUNG ARYSIWA ISWARA	A	B	BC	A	B	AB	
160010249	NI LUH GEDE WIWIN HERLINDA	A	C	C	A	C	B	
160010266	GREYDIOS NANDA VANDITA	A	AB	B	A	AB	AB	
160010346	BRYAN GERRY RUPULU	A	B	C	A	B	C	
160010040	A A BAGUS KUSUMA DANA SWARA	A	BC	C	A	BC	C	
160010177	I KADEK DONNY SURYANATA	A	C	A	A	C	A	
160010265	DEDE ARI SAPUTRA	A	B	B	A	B	BC	
160010345	I WAYAN HAGY PRIADI	A	B	B	A	B	B	
160010011	KETUT ARI JAYA KUSUMA	A	A	BC	A	A	B	
160010027	I KOMANG WENTEN AGUS JUNIANTARA	A	B	AB	A	B	A	
160010043	PUTU REZA SETYAWAN	A	AB	A	A	AB	A	
160010127	I KADEK SURYA ADI PUTRA	A	B	A	A	B	A	

Gambar 4. Data Set Nilai Mahasiswa

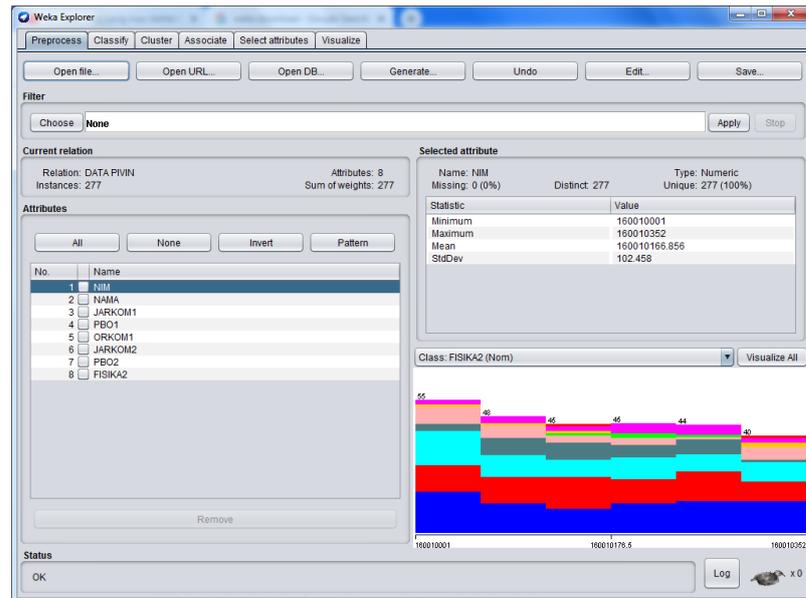
Penelitian ini membahas mengenai pengelompokan nilai dari matakuliah konsentrasi mahasiswa, hal ini dilakukan untuk mengetahui seberapa kelompok yang terbentuk dari data yang ada. Proses clustering yang digunakan menggunakan data set nilai dari matakuliah prasyarat konsentrasi. Dataset yang digunakan pada penelitian ini adalah dataset dari data nilai mahasiswa STIKOM Bali program studi Sistem Komputer Angkatan 2016. Data nilai matakuliah yang digunakan adalah Nilai Matakuliah Jaringan Komputer, Pemrograman Berorientasi Obyek, Fisika, dan Organisasi komputer. Contoh format dataset nilai mahasiswa yang digunakan ditunjukkan pada gambar 4 diatas.

Berdasarkan hasil analisis kebutuhan sistem, maka tahap selanjutnya adalah mendesain sistem yang dilakukan dengan perancangan alur metode *feature selection* untuk diterapkan pada permasalahan analisis nilai dari tiap tiap prasyarat matakuliah yang berpengaruh terhadap konsentrasi mahasiswa. Setelah dirancang, langkah selanjutnya adalah melakukan implementasi dan uji coba dengan memanfaatkan fungsi "Select Attribut". Evaluasi dilakukan dengan menggunakan algoritma K-means Clustering untuk melakukan klasifikasi berdasarkan atribut yang terpilih pada tahap seleksi fitur. Berikut alur penerapan feature selection .



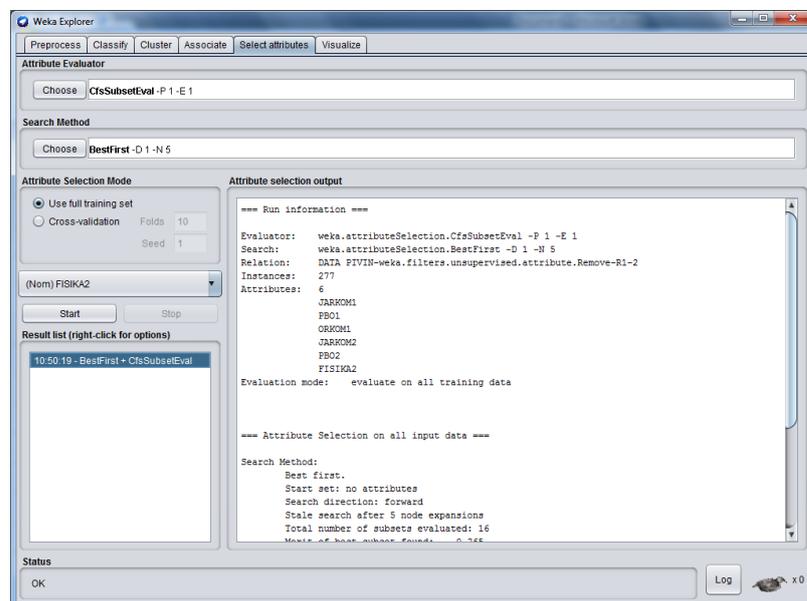
Gambar 5. Alur penerapan feature selection

Tahapan feature selection atau seleksi fitur merupakan tahapan yang umum digunakan pada tahapan pre-processing. Teknik ini digunakan untuk mengurangi kompleksitas atribut yang akan dikelola pada processing dan analisis. Dengan cara ini akan membantu mempercepat proses clustering. Untuk tahapan preprocess kita memilih dataset dengan format .csv, kemudian disini kita dapat memilih atribut mana saja yang akan digunakan.



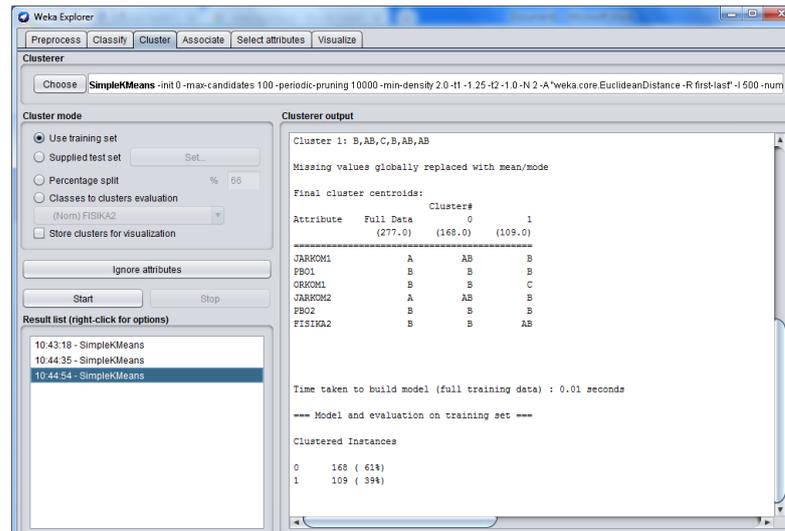
Gambar 6. Proses Preprocess untuk data konsentrasi

Tahapan select atribut, pada tahapan ini dilakukan seleksi terhadap fitur yang digunakan dalam dataset. Seleksi fitur yang dipilih adalah teknik CfSubsetEval dengan search Method Best First. Teknik ini menghitung korelasi antara masing-masing atribut dan variabel output, lalu memilih atribut yang mempunyai nilai korelasi dan membuang atribut yang memiliki nilai korelasi rendah). CFS menggunakan kinerja prediktif dan inter-korelasi fitur untuk mencari sekumpulan fitur yang bagus. Atribut yang dipilih adalah Jarkom1, PBO1, Orkom, Jarkom2, PBO2, Fisika.



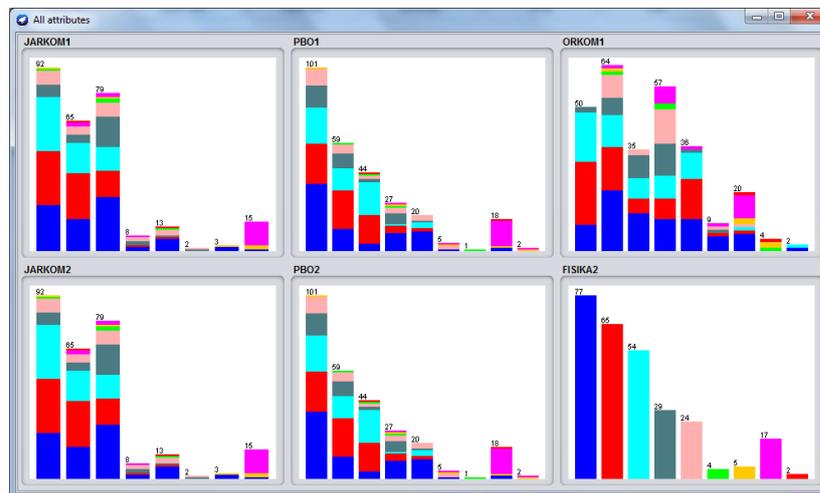
Gambar 7. Proses Select Atribut

Setanjutnya tahapan clustering, disini terlebih dahulu di pilih metode yang digunakan yaitu Simple K-means Clustering. Dengan memanfaatkan fungsi "Select Atribut" yang sebelumnya. Evaluasi dilakukan dengan menggunakan algoritma K-Means Clustering untuk melakukan pengelompokan berdasarkan atribut yang terpilih pada tahap seleksi fitur. Nilai yang diukur berdasarkan output dari aplikasi WEKA.



Gambar 8. Proses Clustering K-Means dengan WEKA

Berikut diagram batang yang menggambarkan prosentase bobot nilai dari masing-masing matakuliah yang dijadikan atribut dalam menentukan clustering.



Gambar 9. Diagram clustering dari masing-masing atribut

Berdasarkan hasil clustering menggunakan K-Means pada dapat diketahui bahwa penerapan feature selection secara umum meningkatkan akurasi dari algoritma Clustering. Dari hasil clustering dengan metode K-Means dengan WEKA maka dihasilkan 2 cluster yang terbentuk yaitu cluster 0 dan cluster 1, dimana cluster 0 ini memiliki nilai 61 % dengan data 168 dan cluster1 dengan nilai 39 % dengan data 109.

4. Simpulan

Kesimpulan yang didapat dari penelitian ini adalah, Penerapan feature selection secara umum dapat meningkatkan akurasi dari algoritma K-Means Clustering. Dari hasil pengolahan data maka dihasilkan 2 cluster yang terbentuk yaitu cluster 0 dan cluster 1, dimana cluster 0 ini memiliki nilai 61 % dengan data 168 dan cluster1 dengan nilai 39 % dengan data 109, dengan Seleksi fitur yang dipilih adalah teknik CfSubsetEval dengan search Method Best First. Teknik ini menghitung korelasi antara masing-masing atribut dan variabel output

Daftar Rujukan

- Annas, S., kanai, T. & S, K., 2007. Principal Component Analysis (PCA) and Self Organizing Map (SOM) for Visualizing and Classifying Fire Riks in Forest Region. *Agricultural Information Research*, 16(2), pp. 44-51.
- Andri.,Marlindawati.,2015. Pengelompokan Minat Belajar Mahasiswa Menggunakan Teknik Data Mining Dengan Metode Clustering Jurnal Ilmiah Matrik Vol.17 No.1, April 2015: 67 – 76
- Asroni, Ronald, A., 2015. Penerapan Metode K-Means untuk Clustering Mahasiswa berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang. *Jurnal Ilmiah Semesta Teknika*.
- Han, J., Kamber, M. & Pei, J., 2012. *Data Mining. Concepts and Techniques, Third Edition*. United States of America: Morgan Kaufmann Publishers.
- Fina, N., Surya, D., 2015. Penerapan Metode K-Means Clustering pada Data Penerimaan Mahasiswa Baru (Universitas Potensi Utama). Makalah diseminarkan pada Seminar Nasional Teknologi Informasi dan Multimedia, 2015.
- Larose, T.Daniel., 2005. *Discovering Knowledge in Data : an Introduction to Data Mining*. Jhon Wiley & Sons Inc
- Setiawan, Dian. 2015. Perancangan Aplikasi K-Means Sebagai Penentu Konsentrasi Bagi Mahasiswa Informatika UMS. (Online), tersedia: eprints.ums.ac.id.
- Sreedevi, K., dan Chandra, S.P., 2014. Analyzing the Student's Academic Performance by Using Clustering Methods in Data Mining. *International Journal of Scientific & Engineering Research*
- Turkhamun Adi K.,Bosar Panjaitan., Rama Guntara.,2018. Implementasi Data Mining Clustering Data Mahasiswa Teknik Informatika Menggunakan Algoritma K-Means. Makalah diseminarkan pada Seminar Nasional Cendekiawan ke 4 Tahun 2018.