

Analisis Profil Mahasiswa Politeknik Negeri Batam dengan Teknik Data Mining

Asosiasi dan Clustering

Sartikha¹, Maria², Festy Winda Sari³, Nurzahrati Jannah⁴

Informatics Engineering Study Program Batam State Polytechnics

Parkway Street, Batam Centre, Batam 29461, Indonesia

E-mail: sartikha@polibatam.ac.id¹, maria@polibatam.ac.id², festy@polibatam.ac.id³, nur.zahrati@polibatam.ac.id⁴

Abstrak

Saat ini, data mining menjadi topik penelitian yang banyak diminati. Berdekade setelah kita memasuki era komputasi, data menjadi tersedia luas dan tersimpan dalam jumlah yang sangat besar. Data mining dapat mengkonversi data mentah menjadi pengetahuan yang memiliki makna dengan memetakan data level rendah menjadi bentuk lain yang lebih ringkas, lebih abstrak, atau lebih bermakna. Penerapan data mining di dunia pendidikan juga semakin berkembang dan diminati. Educational Data Mining (EDM) merupakan bidang yang memanfaatkan algoritma statistika, machine-learning, dan data mining pada berbagai tipe data yang dalam bidang pendidikan. EDM berfokus dalam mengembangkan metode untuk mengeksplorasi tipe data pendidikan yang unik, untuk lebih memahami siswa dan pengaturan lingkungan di mana mereka belajar. Perguruan tinggi dapat juga memanfaatkan EDM ini untuk meningkatkan kualitas pendidikan dan luarannya. Ketersediaan data mahasiswa selama ini belum dapat dimanfaatkan secara optimal dalam pengambilan keputusan. Data mining dapat digunakan untuk mengelompokkan mahasiswa, mengidentifikasi pola tersembunyi dari gaya belajar mereka, menemukan perilaku mahasiswa yang tidak diinginkan, serta menganalisis profil mahasiswa. Dalam penelitian ini akan dilakukan analisis profil mahasiswa dengan menggunakan teknik data mining asosiasi dan clustering. Hasil dari penelitian ini dapat menjadi saran aksi strategis yang bisa diambil manajemen perguruan tinggi, terutama Politeknik Negeri Batam, untuk meningkatkan kualitas pendidikan dan luarannya. Penelitian ini dimulai dengan pengumpulan data dari mahasiswa Informatika angkatan 2010. Selanjutnya dilakukan preprocessing data yang kemudian didapat atribut yang akan digunakan untuk di analisis dengan menggunakan teknik data mining asosiasi dan clustering.

Kata kunci: *Data mining, educational data mining, asosiasi, clustering.*

Abstract

Currently, data mining became a topic of research is much in demand. Decades after we entered the era of computing, data becomes widely available, and stored in very large quantities. Data mining can convert raw data into knowledge that has meaning to a lower level map data into another form that is more compact, more abstract, or more meaningful. The application of data mining in the world of education is also growing and demand. Educational Data Mining (EDM) is a field that utilize statistical algorithms, machine-learning and data mining on a variety of data types in the field of education. EDM focuses on developing methods to explore the unique educational data type, to better understand students and setting up the environment in which they learn. The college can also take advantage of EDM is to improve the quality of education and luarannya. Student data availability has not been optimally utilized in decision making. Data mining can be used to classify students, identify hidden patterns of their learning styles, find undesirable student behavior, and analyzing student profiles. In this research will be the analysis of student profiles by using data mining techniques association and clustering. The results of this study may be to suggest strategic actions that can be taken college management, especially Batam Polytechnic, to improve the quality of education and superficial. The study began by collecting data from students of Informatics forces do preprocessing, 2010. Further data are then obtained attributes to be used for analysis using data mining techniques association and clustering.

Keywords: *Data mining, educational data mining, asosiasi, clustering.*

1 Pendahuluan

Dewasa ini, data mining menjadi topik penelitian yang banyak diminati. Berdekade setelah kita memasuki era komputasi, data menjadi tersedia luas dan tersimpan dalam

jumlah yang sangat besar. Data mining dapat mengkonversi data mentah menjadi pengetahuan yang memiliki makna dengan memetakan data level rendah menjadi bentuk lain yang lebih ringkas, lebih abstrak, atau lebih bermakna. Aplikasi data

mining telah sukses diimplementasikan di berbagai bidang operasi dan permasalahan dunia nyata berskala besar dalam bidang ilmu pengetahuan dan juga bisnis. Data mining dapat digunakan untuk meningkatkan keuntungan, meminimalisir biaya, menarik konsumen, mengatur ulang penawaran produk, dan lain-lain. Data mining juga dapat diterapkan di bidang astronomi, keuangan, manufaktur, pendidikan, telekomunikasi, kesehatan, dan masih banyak lagi (Witten & Frank, 2005).

Penerapan data mining di dunia pendidikan juga semakin berkembang dan diminati. Educational Data Mining (EDM) merupakan bidang yang memanfaatkan algoritma statistika, machine-learning, dan data mining pada berbagai tipe data yang dalam bidang pendidikan. EDM berfokus dalam mengembangkan metode untuk mengeksplorasi tipe data pendidikan yang unik, untuk lebih memahami siswa dan pengaturan lingkungan di mana mereka belajar (Romero, 2010).

Perguruan tinggi dapat juga memanfaatkan EDM ini untuk meningkatkan kualitas pendidikan dan luarannya. Ketersediaan data mahasiswa selama ini belum dapat dimanfaatkan secara optimal dalam pengambilan keputusan. Data mining dapat digunakan untuk mengelompokkan mahasiswa, mengidentifikasi pola tersembunyi dari gaya belajar mereka, menemukan perilaku mahasiswa yang tidak diinginkan, serta menganalisis profil mahasiswa (Parack, Zahid, & Merchant, 2012). Mengenali profil mahasiswa dengan lebih seksama dapat memberi petunjuk tentang potensi mahasiswa untuk berprestasi secara akademik. Dengan mengenali pola-pola profil mahasiswa, pihak perguruan tinggi, terutama dosen wali yang langsung berhubungan dalam pembimbingan mahasiswa, dapat memanfaatkan informasi tersebut untuk melakukan antisipasi dalam langkah pembimbingan dan pengarahan mahasiswa agar dapat mengoptimalkan prestasi dan meminimalisir potensi masalahnya.

2 Tinjauan Pustaka

(Romero, 2010) telah melakukan survey yang komprehensif mengenai penerapan data mining dalam bidang pendidikan (EDM) dan potensi-potensi pemanfaatannya. Pemanfaatan EDM dapat digolongkan ke dalam tiga bidang penerapan:

1. Pendidikan secara offline yang bertujuan menyalurkan pengetahuan melalui tatap muka serta studi psikologis mengenai bagaimana manusia belajar.
2. E-learning dan learning management system.
3. Intelligence tutoring system (ITS).

(Al-shargabi & Nusari, 2010) menerapkan data mining untuk

memprediksi prestasi akademis (keberhasilan dan kegagalan), drop out, dan perilaku pembayaran biaya kuliahnya. Metode yang digunakan adalah K-means clustering, Apriori, decision tree J48 dan Id3. Data yang digunakan adalah data akademik dari University of Science and Technology (UST) dari tahun 1993 sampai 2005. Mereka berhasil menemukan beberapa pola menarik mengenai prestasi akademik, drop out, dan perilaku pembayarannya serta melengkapinya dengan saran aksi strategis bagi pihak manajemen UST.

Analisis profil mahasiswa dengan data mining juga dilakukan oleh (Parack et al., 2012) dengan meneliti empat atribut utama, yakni nilai ujian, nilai mata kuliah, kehadiran, dan nilai praktek. Metode yang digunakan adalah Apriori dan K-means clustering.

Penerapan lain EDM dilakukan oleh (Kumar & Vijayalakshmi, 2012) untuk memprediksi prestasi akademik pada semester 5 berdasarkan pencapaian akademik dari 4 semester sebelumnya. Prediksi prestasi mahasiswa menggunakan data mining juga dilakukan oleh (Wook et al., 2009) dengan membandingkan penggunaan metode kombinasi clustering dan decision tree, dengan jaringan syaraf tiruan. Penelitian ini merupakan tahap awal untuk menentukan algoritma terbaik untuk memprediksi prestasi mahasiswa.

Clustering

Clustering adalah proses mengumpulkan objek yang mirip dari sumber data. Sumber data dapat berupa *file*, *paper*, OLTP, *Database System*, dan *Information Provider*. Teknik *clustering* dapat diklasifikasikan menjadi *hierarchical method* (*agglomerative* dan *disisive*), *partitioning method* (*probabilistic method*, *k-medoids method*, *k-means method*, *desity-based method*), *grid method*, dan teknik lainnya (P. Berkhin, 2002). *Clustering* yang baik akan menghasilkan kemiripan objek antar kelas sangat mirip dan mirip serta dapat menemukan pola tersembunyi dari data. Kemiripan objek ditentukan berdasarkan jarak, seperti Minkowski, Manhattan dan Euclidean. Adapun rumus untuk menghitung jarak objek, yaitu :

Jarak *Minkowski* dihitung menggunakan rumus 1. $d(i,j) = \sqrt[q]{(|x_{i1}-x_{j1}|^q + |x_{i2}-x_{j2}|^q + \dots + |x_{ip}-x_{jp}|^q)}$

Jarak *Manhattan* dihitung dengan rumus 2. jika $q=1$. $d(i,j) = (|x_{i1}-x_{j1}| + |x_{i2}-x_{j2}| + \dots + |x_{ip}-x_{jp}|)$

Jarak *Euclidean* dihitung dengan menggunakan rumus 3, jika $q=2$. $d(i,j) = (|x_{i1}-x_{j1}|^2 + |x_{i2}-x_{j2}|^2 + \dots + |x_{ip}-x_{jp}|^2)$

Data Asosiasi

Analisis asosiasi atau association rule mining adalah teknik data

mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (**frequent pattern mining**) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, **support** (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan **confidence** (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Algoritma Apriori

Ide utama pada algoritma apriori adalah dengan membaca database secara berulang. Langkah pada algoritma apriori adalah :

1. Mencari frequent itemset dari basis data transaksi
2. Menghilangkan itemset dengan frekuensi yang rendah berdasarkan level minimum *support* yang telah ditentukan sebelumnya
3. Membuat aturan asosiasi (*association rule*) dari itemset yang memenuhi ketentuan nilai minimum *confidence* dalam basis data.

Algoritma apriori merupakan suatu bentuk algoritma dalam data mining yang akan memberikan informasi tentang hubungan antar item dalam database yang dapat dimanfaatkan secara luas dalam proses bisnis khususnya dalam proses penjualan. *Knowledge* algoritma apriori terletak pada *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya (Erwin, 2009).

Terdapat dua proses utama pada algoritma apriori (Jiawei Han & Micheline Kamber, 2000), yaitu sebagai berikut :

1. *Join* (penggabungan). Dalam proses ini, setiap *item* dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi.
2. *Prune* (pemangkasan). Pada proses ini, hasil kombinasi item akan dipangkas dengan menggunakan minimum *support* yang telah ditentukan oleh pengguna.

Algoritma apriori bekerja dengan cara menghasilkan kandidat baru dari *k-itemset* pada frequent item set sebelumnya dan menghitung nilai *support-itemset* tersebut. *Item set* yang memiliki nilai *support* di bawah dariminsupakan dihapus. Langkah selanjutnya adalah menghitung minimal *confidence* mengikuti rumus sesuai yang telah ditentukan. *Support* tidak perlu dilihat lagi, karena generate *frequent itemset* didapatkan dari melihat minsup-nya. Bilanya yang didapatkan memenuhi batasan.

Metode Penelitian Pengumpulan Data

Metodologi penelitian dilakukan dengan mengumpulkan data terlebih dahulu. Proses pengumpulan data dilakukan dengan dua cara yaitu: studi kepustakaan dan studi lapangan. Studi kepustakaan dilakukan untuk memahami penerapan data mining seperti mengumpulkan teori-teori yang berhubungan dengan data mining dan studi lapangan dilakukan mendapatkan data penelitian. data yang diamati diharapkan adalah data profil dan data akademik mahasiswa Politeknik Negeri Batam angkatan 2010. Data penelitian seperti : Asal Sekolah, alamat, pekerjaan orangtua, IPK dan .

Pengolahan Data

Pada tahap pengumpulan data selanjutnya dilakukan tahap pengolahan data. Data profil yang telah di dapat dari Politeknik Negeri Batam angkatan 2010 di siapkan dan dipilih untuk dilakukan analisis dengan Asosiasi dan Clustering. Persiapan dan pemilihan data atau data *preprocessing*.

Data Preprocessing

Data preprocessing adalah melakukan pembersihan data dari data outlier atau data yang *error*, data yang hilang/kosong, kekurangan atribut yang sesuai dan data yang tidak konsisten atau berisi nilai yang berbeda. Data preprocessing ini sangat penting untuk menghasilkan kualitas mining yang baik. Ada beberapa teknik data preprocessing yang dilakukan untuk menghasilkan data tersebut, yaitu:

1. Data Cleaning

Proses data cleaning dilakukan berkaitan dengan mendeteksi, menghapus kesalahan dan data yang tidak konsisten untuk meningkatkan kualitas data.

Kesalahan data bisa berupa data tidak lengkap, kesalahan ejaan, data duplikat. Pada tahap ini dilakukan reduksi data tidak lengkap dan atribut yang tidak diperlukan, melengkapi data kosong dan memperbaiki data yang tidak konsisten. Berikut ini hal yang dilakukan untuk mengatasi masalah yang terdapat pada penelitian ini, yaitu:

a. Atribut yang tidak diperlukan

Beberapa atribut yang tidak digunakan dalam analisis asosiasi dan clustering, yaitu; Nomor, NIM, Nama, Tempat Lahir, Tanggal Lahir, Agama, Program Studi, Suku Bangsa, Golongan Darah, Nilai UN, Nilai Ijazah, Pekerjaan Ijazah, Pekerjaan Ibu, Hobi, Alamat Orang Tua.

b. Data tidak Konsisten

Data tidak konsisten terdapat pada atribut jenis kelamin dan asal sekolah. Pada atribut jenis kelamin ada data yang diberi spasi dan ada yang tidak, kemudian ada data yang menggunakan huruf kecil dan tidak. Sehingga data tersebut di sesuaikan agar hanya terdapat dua data saja. Laki-Laki dan Perempuan. Sedangkan untuk data asal sekolah, ada beberapa data yang sama namun berbeda penulisannya, misal SMKBBBS dan SMK Batam Business School. Kedua SMK tersebut adalah SMK yang sama namun beda penulisan, maka dilakukan perubahan data agar data dan penulisannya sama.

2. Diskretisasi

Membagi nilai data menjadi beberapa range data dengan menggunakan hirarki konsep yaitu mengelompokan nilai IPK berdasarkan range dan disesuaikan dengan predikatnya. Contoh range 3.5-4 mendapat predikat dengan pujian. Tabel contoh data

3. Data Transformasi

Tranformasi mengubah data yang kompleks dengan tidak menghilangkan isi, sehingga lebih mudah di olah.

a. Generalisasi

Generalisasi dilakukan pada atribut alamat mahasiswa. Alamat mahasiswa dibuat per grup berdasarkan kecamatannya contoh data bisa dilihat pada Tabel 4.1, asal sekolah mahasiswa dibuat berdasarkan kategori status negeri dan swasta, lalu di kategorikan lagi berdasarkan yang ada di Batam atau luar batam dan dikategorikan berdasarkan SMA atau SMK.

Hasil Yang Dicapai

Asosiasi

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan assosiatif antara suatu kombinasi item. Analisis asosisati digunakan untuk menemukan aturan asosiatif dari suatu kombinasi Item, menemukan rule anantara data demografis mahasiswa dengan data akademiknya. Kolaborasi items yang akan digunakan untuk data demografis akan menggunakan kategori sekolah, jenis kelamin, kategori asal sekolah dengan keterangan lulus dan predikat yang didapat setelah lulus. Analisis Asosiasi menggunakan menggunakan Weka 3.6 dengan aturan sebagai berikut:

- Jumlah data=93

- Minimum Support=30%
- Minimum Confidence=90%

Tabel 1. Kemunculan 1-Items

Nama Atribut	Item Atribut
Kategori SMA atau SMK	SMA
	SMK
Jenis Kelamin	Laki-Laki
	Perempuan
Kategori Status Sekolah Negeri/Swasta	Negeri
	Swasta
Keterangan Lulus	Lulus
	Berhenti
Predikat	Memuaskan
	Sangat Memuaskan
	Dengan Pujian
Kategori daerah sekolah dalam dan luar Batam	Batam
	Luar Batam

Tabel 2 Hasil Asosiasi F2

Kandidat aturan asosiasi	Confidence
(Predikat:Sangat Memuaskan)→Ket Lulus=Sudah Lulus	100%
(JK=Perempuan)→(Ket. Lulus=Sudah Lulus)	96%
(Kat.SMA/SMK=SMA) →(Kat.Status Sekolah= Negeri)	90%

Tabel 3 Hasil Asosiasi F3

Kandidat Aturan Asosiasi	Confidence
(JK=Perempuan), (Predikat=Sangat Memuaskan)→(Ket. Lulus=Sudah Lulus)	100%
(Kat.Daerah=Batam),(Predikat=Sangat Memuaskan)→(Ket.Lulus=Sudah Lulus)	100%
(JK=Perempuan), (Kat.Daerah=Batam) →(Ket.Lulus=Sudah Lulus)	97%
(JK=Perempuan), (Kat.SMA/SMK=SMA) →(Ket.Lulus=Sudah Lulus)	97%
(Kat.Daerah=Batam),(Kat.SMA/SMK=SMA)→(Kat.Status Sekolah= Negeri)	97%
(JK=Perempuan), (Kat.Status Sekolah= Negeri)→(Ket. Lulus=Sudah Lulus)	94%
(JK=Perempuan), (Kat.SMA/SMK=SMA) →(Kat.Status Sekolah= Negeri)	90%

Sehingga dari hasil analisa asosiasi apriori dengan menggunakan Weka 3.6, maka ditemukan 10 rule sebagai berikut.

1. Semua mahasiswa dengan predikat yang Sangat Memuaskan 100% sudah lulus dari Politeknik Negeri Batam
2. Mahasiswa Perempuan dan sudah lulus nilai confidencenya 96%
3. Kategori SMA dan Negeri nilai confidencenya 90%
4. Mahasiswa dengan jenis kelamin perempuan dengan predikat sangat memuaskan dan lulus nilai confidencenya 100%
5. Mahasiswa yang dari Batam dengan predikat sangat memuaskan dan Lulus nilai confidencenya 100%

6. Mahasiswa dengan jenis kelamin perempuan dari sekolah di Batam dan Lulus nilai confidencenya 97%
7. Mahasiswa dengan jenis kelamin perempuan dengan kategori SMA dan lulus nilai confidencenya 97%
8. Mahasiswa SMA di Batam dengan status negeri nilai confidencenya 97%
9. Mahasiswa dengan jenis kelamin perempuan yang bersekolah di sekolah negeri dan lulus nilai confidencenya 94%
10. Mahasiswa dengan jenis kelamin perempuan yang bersekolah di SMA negeri tingkat confidencenya 90%

Clustering

Clustering atau pengelompokan mahasiswa berdasarkan data demografis dan data akademiknya menggunakan K-Means. Beberapa tahapan yang harus dilakukan sebelum melakukan clustering yaitu melakukan diskretization dan transformasi ke numerik. Selanjutnya akan di clustering K-means dengan menggunakan Weka 3.6.

1. Data Diskretisasi

Dalam data diskretisasi dilakukan penggolongan atribut. Hal ini dilakukan untuk mempermudah proses data mining

Tabel 4 Diskretisasi Data

Nama Atribut	Data Asal	Data Setelah dilakukan pengelompokan Numerik
Kategori Sekolah	SMA	1
	SMK	0
Jenis Kelamin	Laki-Laki	1
	Perempuan	0
Kategori Asal Sekolah	Batam	1
	Luar Batam	0
Kategori status sekolah	Negeri	1
	Swasta	0
Predikat	Memuaskan	1
	Sangat Memuaskan	2
	Dengan Pujian	3
Keterangan Lulus	Berhenti/Masih Mahasiswa	0
	Lulus	2
	Masih mahasiswa	1
Gaji Orang Tua	>1000.000 dan <3000.000	1
	>=3000.000 dan <=4000.000	2
	>4000.000	3

2. Data Transformasi

Pada tahap ini dilakukan proses mengubah nilai data pada setiap atribut dengan rumus *MinMax Normalization* untuk menyamakan range atau jarak antara atribut. Berikut tabel perbedaan range antar atribut.

Tabel 5 Range Data Atribut

Nama Atribut	Min _n	Max _n	Range Data
Kategori Sekolah	0	1	1
Jenis Kelamin	0	1	1
Kategori Asal	0	1	1
Kategori status Sekolah	0	1	1
Predikat	0	3	3
Keterangan Lulus	0	2	2
IPK	0.3	3.92	3.62

Adapun atribut yang akan ditransformasi ke dalam rumus *min max normalization* adalah gaji ortu, kategori sekolah, jenis kelamin, kategori asal, kategori status, predikat dan keterangan lulus.

Adapun rumus *MinMax Normalization* adalah sebagai berikut:

$$v' = \frac{v - \min_n}{\max_n - \min_n} (\text{new_max}_n - \text{new_min}_n) + \text{new_min}_n$$

V merupakan nilai yang ingin dinormalisasi, v' adalah hasil data yang telah dinormalisasi, nilai new min adalah nilai minimum yang diinginkan yaitu 0. Sedangkan new max diperoleh dari pengurangan nilai maksimum dengan nilai minimum pada atribut IPK yaitu 3.92 dan 0.3.

Contoh:

Diketahui Gaji Ortu= 1, Kategori Sekolah=1, Jenis Kelamin=0, Kategori Asal=1, Kategori Status Sekolah=1, Predikat=3, Kategori Lulus=2, IPK=3.83. Pembahasan

Gaji ortu (G)	$v'G = (1-0)/(3-1) * ((3.62-0)+0) = 1.81$
Kategori Sekolah (KS)	$v'KS = (1-0)/(1-0) * ((3.62-0)+0) = 3.62$
Jenis Kelamin (JK)	$v'JK = (0-0)/(1-0) * ((3.62-0)+0) = 0.00$
Kategori Asal Sekolah (KA)	$v'KA = (1-0)/(1-0) * ((3.62-0)+0) = 3.62$
Kategori Status Sekolah (KSS)	$v'KSS = (1-0)/(1-0) * ((3.62-0)+0) = 3.62$
Predikat (P)	$v'P = (3-0)/(3-0) * ((3.62-0)+0) = 3.62$
Keterangan Lulus (KL)	$v'KL = (2-0)/(2-0) * ((3.62-0)+0) = 3.62$

Data yang telah diproses akan dilakukan *clustering* dengan menggunakan Weka 1.6 kemudian hasilnya akan dianalisis. Data akan dibagi menjadi 3 *cluster*, *cluster 0*, *cluster 1* dan *cluster2*. Berikut data yang sudah di cluster dengan menggunakan weka.

Berdasarkan data yang telah di-*clustering* maka kelompok yang muncul pada *cluster* mempunyai tipenya masing-masing: *Cluster 0* Gaji orang tua 2-5 juta, Cenderung Perempuan yang bersekolah di SMK kebanyakan swasta di Batam IPK rata-rata tinggi dengan predikat memuaskan sampai sangat memuaskan

dan kebanyakan mereka lulus tepat waktu.

Cluster 1 Gaji orang tua 2-5 juta, Kebanyakan Perempuan yang bersekolah di SMA Negeri cenderung di Batam, IPK tinggi dengan predikat sangat memuaskan dan kebanyakan mereka lulus tepat waktu.

Cluster 2 Gaji orang tua 2-3 juta, Kebanyakan laki-laki cenderung SMA Negeri di luar Batam dengan IPK rendah predikat memuaskan dan cenderung belum lulus atau tidak lulus tepat waktu.

Kesimpulan

Dari hasil analisis data mahasiswa Politeknik Negeri Batam jurusan Informatika tahun 2010 dengan metode asosiasi dan *clustering* dapat disimpulkan sebagai berikut:

1. Dengan Metode Asosiasi dengan tingkat *confidence* $\geq 90\%$ ditemukan 10 rule penting yang tersembunyi dalam data mahasiswa Politeknik Negeri Batam jurusan Informatika tahun 2010. Pola data yang ditemukan bahwa mahasiswa yang lulus tepat waktu dengan predikat sangat memuaskan kebanyakan berjenis kelamin perempuan dari SMA Negeri di Batam.
2. Dengan Metode *Clustering* terdapat 3 *cluster* yang terbentuk yaitu:
 - a. Cluster 0 Gaji orang tua 2-5 juta, Cenderung Perempuan yang bersekolah di SMK kebanyakan swasta di Batam IPK rata-rata tinggi dengan predikat memuaskan sampai sangat memuaskan dan kebanyakan mereka lulus tepat waktu.
 - b. Cluster 1 Gaji orang tua 2-5 juta, Kebanyakan Perempuan yang bersekolah di SMA Negeri Cenderung di Batam, IPK tinggi dengan predikat sangat memuaskan dan kebanyakan mereka lulus tepat waktu.
 - c. Cluster 2 Gaji orang tua 2-3 juta, Kebanyakan laki-laki Cenderung SMA Negeri di luar Batam dengan IPK rendah predikat memuaskan dan cenderung belum lulus atau tidak lulus tepat waktu.

Daftar Rujukan

Al-shargabi, A. a, & Nusari, A. N. (2010). Discovering vital patterns from UST students data by applying data mining techniques. In *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)* (pp. 547–551). Ieee. doi:10.1109/ICCAE.2010.5451653

Kumar, S. A., & Vijayalakshmi, M. N. (2012). Mining of student academic evaluation records in higher education. In *2012 International Conference on Recent Advances in Computing and Software Systems* (pp. 67–70). Ieee. doi:10.1109/RACSS.2012.6212699

Parack, S., Zahid, Z., & Merchant, F. (2012). Application of data mining in educational databases for predicting academic trends and patterns. In *2012 IEEE International Conference on Technology Enhanced Education (ICTEE)* (pp. 1–4). Ieee. doi:10.1109/ICTEE.2012.6208617

Romero, C. (2010). Educational Data Mining : A Review of the State of the Art. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 40(6), 601–618.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). San Francisco: Morgan Kaufman.

Wook, M., Yahaya, Y. H., Wahab, N., Isa, M. R. M., Awang, N. F., & Seong, H. Y. (2009). Predicting NDUM Student's Academic Performance Using Data Mining Techniques. In *2009 Second International Conference on Computer and Electrical Engineering* (pp. 357–361). Ieee. doi:10.1109/ICCEE.2009.168

Kantardzic, Mehmed. 2003. *Data Mining Concepts Models, Methods, and Algorithms*. New Jersey: IEEE.

Berkhin, P. 2002. *Survey of Clustering Data Mining Techniques*, Accrue Software, 1045 Forest Knoll Dr. San Jose.

MacQueen, J. B. (1967). *Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, University of California Press, 1: 281-297.

Han Jiawei, and M. Kamber. 2006. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, USA.