



ANALISIS DATA MAHASISWA DENGAN ALGORITMA K-MEAN UNTUK MENDUKUNG STRATEGI PROMOSI STMIK HANG TUAH PEKANBARU

Eka Sabna¹⁾, Afifa²⁾, Ria Mutiara Sari³⁾
Teknik Informatika, STMIK Hang Tuah Pekanbaru^{1,2,3}

Email:

es3jelita@yahoo.com¹⁾, afifa@alumni.pcr.ac.id²⁾, riamutiara6@gmail.com³⁾

Abstract

Data of new students owned by STMIK Hang Tuah Pekanbaru are already in large numbers with many variations of attributes stored in the database. However, the data data has not been utilized optimally, therefore a method is needed that can explore and extract the new student data into information that is of strategic value. This study aims to group data on New Student STMIK Hang Tuah Pekanbaru by using Data Mining Engineering. The Data Mining technique used is the Clustering Technique with the K-Means algorithm and for the process used the CRISP-DM method. This research is used to determine the right Promotion Strategy. Determination of the right promotion strategy will be able to reduce the cost of promotion and achieve the right promotional goals.

Keywords: *New Student Data, Promotion, Data Mining, K-Means, CRISP-DM.*

Abstrak

Data mahasiswa baru yang dimiliki oleh STMIK Hang Tuah Pekanbaru sudah dalam jumlah besar dengan banyak variasi atribut yang tersimpan dalam basis data (database). Namun data data tersebut belum di manfaatkan secara optimal, oleh karena itu di perlukan suatu metode yang dapat menggali dan mengekstrak data mahasiswa baru tersebut menjadi suatu informasi yang bernilai strategis. Penelitian ini bertujuan untuk melakukan pengelompokkan terhadap data Mahasiswa Baru STMIK Hang Tuah Pekanbaru dengan menggunakan Teknik Data Mining. Teknik Data Mining yang digunakan adalah Teknik Clustering dengan algoritma K-Means dan untuk proses digunakan metode CRISP-DM. Penelitian ini dimanfaatkan untuk menentukan Strategi Promosi yang tepat. Penentuan strategi promosi yang tepat akan dapat mengurangi biaya promosi dan mencapai sasaran promosi yang tepat.

Kata Kunci : Data Mahasiswa Baru, Promosi, Data Mining, K-Means, CRISP-DM..

PENDAHULUAN

STMIK Hang Tuah Pekanbaru merupakan salah satu Sekolah Tinggi Kesehatan yang terkemuka di Pekanbaru. Mahasiswa STMIK Hang Tuah Pekanbaru berasal dari berbagai daerah di Riau seperti Pekanbaru, Duri, Bangkinang, Siak dan juga di luar daerah seperti Sumatra Barat. Untuk Provinsi Riau sendiri Sekolah Tinggi yang sejenis banyak bermunculan, hal ini akan menimbulkan persaingan antar sekolah tinggi akan semakin tinggi. Oleh karena itu untuk menghadapi persaingan ini diperlukan strategi promosi yang tepat. Proses penerimaan mahasiswa baru dalam suatu Perguruan Tinggi menghasilkan data yang berlimpah berupa profil dari mahasiswa baru tersebut. Data data tersebut kemudian di olah untuk mengetahui pola dengan menggunakan teknik Data Mining Klustering. Algoritma yang digunakan adalah *k-means* dan proses data mining nya menggunakan CRISP-DM. Dengan adanya pengelompokan-pengelompokan data seperti ini, diharapkan bagian marketing dapat melakukan pemasaran dengan strategi yang tepat untuk mendapatkan calon mahasiswa baru.

Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data mahasiswa STMIK Hang Tuah Pekanbaru dengan memanfaatkan proses data mining dengan menggunakan teknik Clustering. Berdasarkan penjelasan yang telah diuraikan diatas maka pokok permasalahan yang akan dirumuskan dalam penelitian ini adalah Bagaimana menerapkan metode *data mining Clustering* dengan Algoritma *K-means* untuk menentukan sebaran mahasiswa sebagai informasi penunjang strategi promosi calon mahasiswa baru yang efektif dan efisien ?

METODE PENELITIAN

Pada penelitian ini menggunakan 2 metode/pendekatan yaitu (1) Metode Pengumpulan Data dan (2) Metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) untuk proses dalam Data Mining nya.

Metode Pengumpulan Data

Untuk mendapatkan data, seorang peneliti dapat melakukan berbagai metode. Data dapat diperoleh melalui wawancara dan dokumentasi.

- Wawancara dilakukan dengan cara melakukan percakapan langsung dengan responden dan narasumber.
- Dokumentasi dapat ditempuh dengan cara meminta data pada lembaga atau instansi yang terkait, dalam hal ini adalah data mahasiswa STKes Hang Tuah Pekanbaru.

Selain dengan cara diatas juga digunakan metode lain untuk mendukung penelitian yang dilakukan yaitu dengan menggunakan data sekunder. Data Sekunder adalah data yang diperoleh secara tidak langsung bersumber dari dokumentasi, literature, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang di teliti. Data Sekunder pada penelitian ini adalah Buku-buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti. Data sekunder pada penelitian ini adalah buku-buku, jurnal tentang Algoritma *K-Means* dan data mining serta data mahasiswa STMIK Hang Tuah Pekanbaru .

Metode CRISP-DM (*Cross Industry Standard Process for Data Mining*)

Penelitian ini menggunakan metode CRISP-DM dalam proses Data Mining. Metode ini terdiri dari 6 tahapan :

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Pemahaman terhadap bisnis (*Business/Research Understanding Phase*)

Data yang diperoleh dari database Mahasiswa STMIK Hang Tuah Pekanbaru yaitu berupa data mahasiswa baru Tahun 2018/2019. Data-data tersebut berisi data diri mahasiswa yang terdiri dari atribut Nama Mahasiswa, Asal Sekolah dan Program Studi. Berikut ini adalah data mahasiswa baru pada tahun 2018/2019.

Tabel-1 Contoh Data Mahasiswa baru Tahun 2018/2019

NAMA	ASAL SEKOLAH
Gunawan Syahputra	SMKN5 PKU
Fakhriyah	MA HASANAH
Nizanil Fadli	SMK PPYHM PKL.KERINCI
Aisah Lestari	SMA N 1 PASIR PENYU
Febby Zurian	SMA N 1 TELUK KUANTAN
Sari Bunga	SMK N PERTANIAN TERPADU PROV. RIAU
Arnanda Afrinaldi	SMA N 1 KATEMAN
Nikita Salwa	SMAN1 SUNGAI LALA
Malik Fajar	SMAN 1 KUANTAN HILIR SEBERANG
Yurike Safitri	SMK NEGERI 3
Udi Yono	SMAN 1 MANDAH
Bayu Fardilla	SMK N 1 TEMBILAHAN KOTA

Rindianisesa Perdanosri	SMKN 2 PEKANBARU
Yoki Rahmad Afdal	SMK N 6 PEKANBARU
Nadia Putri Amalyah	SMA N 2 SIAK HULU
Syah Wira Wadi	SMA N 2 BANGKO
Liana Putri Siregar	SMA NEGERI 1 MINAS
Pebriya Kalsum	SMAN 2 TEMBILAHAN
Muhd Ali Kajufri	SMK MUHAMMADIYAH PEKANBARU
Teguh Tri Laksono	SMK NEGERI 7 PEKANBARU
Refni Ayu Hesti	SMK ISLAM INAYAH UJUNG BATU
Marsaulina Magdalena Siahhan	SMA NEGRI 3 TAMBUSAI
Bayu Putra Haryanto	SMK NEGERI 2 PEKANBARU
Alfitra Pratama	SMK HASANAH
Ranto Hasurungun	SMA YLPI
Mohammad Randy Dwinugraha	SMA WIDYA GRAHA PEKANBARU

Fase Pemahaman Data(Data Understanding Phase)

Data di peroleh dari database mahasiswa STMIK Hang Tuah Pekanbaru sebanyak 136 mahasiswa. Atribut yang digunakan yaitu NAMA, Asal Sekolah dan Program Studi. Data data tersebut nanti nya akan diproses untuk di kelompokkan.

Persiapan data (Data Preparation)

Dari 136 data mahasiswa maka akan dilakukan teknik preprocessing agar kualitas data diperoleh lebih baik dengan cara :

1. *Data Validation*, untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsisten dan data yang tidak lengkap (missing value).
2. *Data Integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. data ditransformasi ke dalam angka.
3. *Data Size Reduction and Discretization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informative. Di dalam *data training* yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi.

Fase Pemodelan (Modeling Phase)

Pada tahapan ini merupakan tahapan pemrosesan data, Pada penelitian ini menggunakan *algoritma K-Means*.

Fase Evaluasi (Evaluation Phase)

Pada fase ini dilakukan pengujian terhadap model kluster dengan algoritma *k-means*.

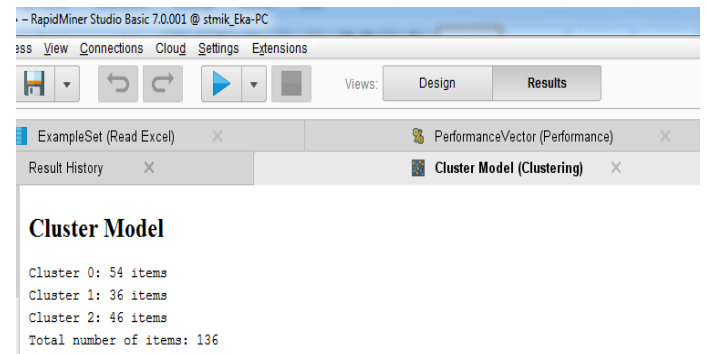
Fase Penyebaran (Deployment Phase)

Setelah pembentukan model selanjutnya dapat digunakan untuk me-kluster data mahasiswa baru di lingkungan STMIK Hang Tuah Pekanbaru.

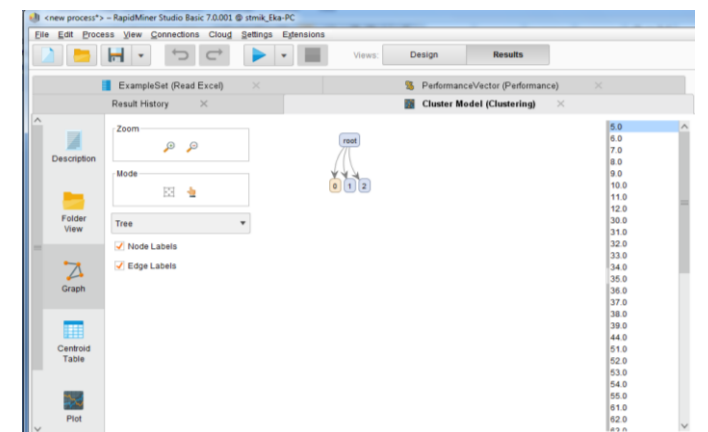
HASIL DAN PEMBAHASAN

Hasil Clustering

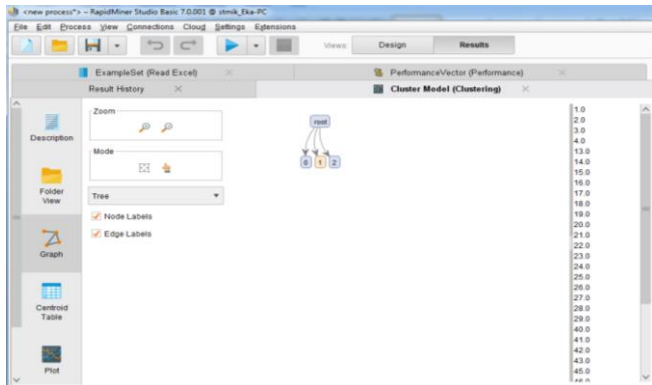
Setelah semua data mahasiswa di transformasikan kedalam bentuk angka maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma K-Means Clustering. Untuk mengolah data tersebut digunakan Software Rapidminer dengan hasilnya sebagai berikut :



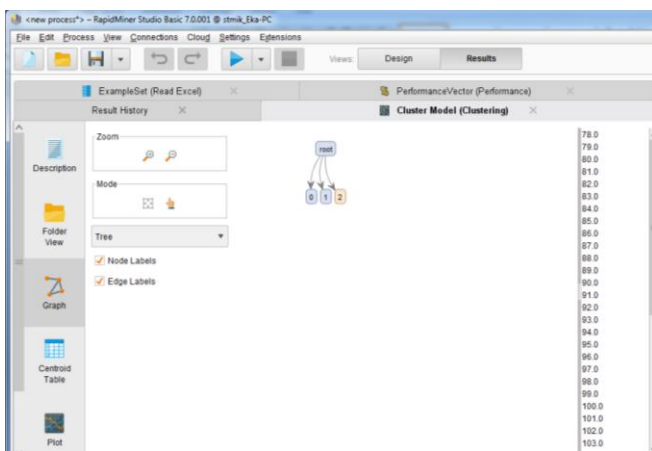
Gambar 4.1 Hasil Cluster 0,1 dan 2



Gambar 4.2 Rincian Cluster 0



Gambar 4.3 Rincian Cluster 1



Gambar 4.3 Rincian Cluster 2

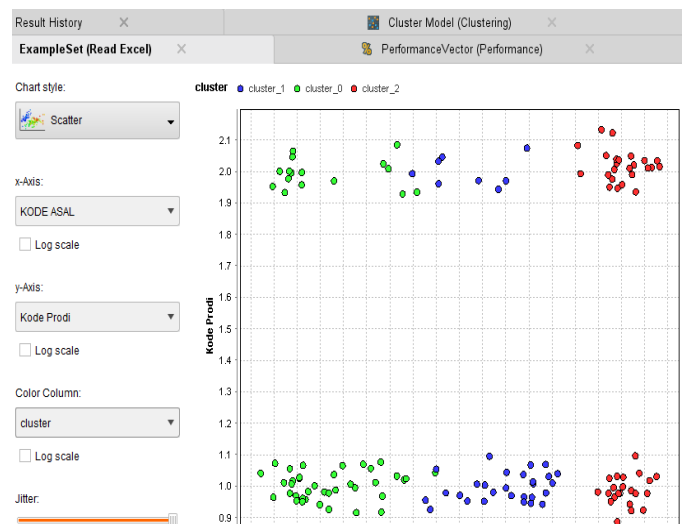
Tabel-2 Tabel Cluster

<p>Cluster 0 terdiri dari 54 orang yang berasal dari kota :</p> <ul style="list-style-type: none"> - Bangko - Duri - Indragiri Hulu - Jambi - Kampar Hilir - Kec Batang Kapas - Kec Bathi Solapan - Kec Kateman - Kec Kota Ilir - Kec Kuantan Tengah - Kec Nanggalo - Kec Pelangiran - Kec Reteh - Kec Rokan - Kec Rumbai - Kec Sail - Kec Tapung Hulu - Kec Tualang - Lipat Kain - Minas - Muaro Bungo - Payakumbuh - Pangkalan Kerinci - Rengat - Rimbo Bujang - Teluk Kuantan - Teluk Pinang - Kec Kapur - Kec Kelayang - Kec Tambang - Siak - Tembilahan 	<p>Cluster 1 terdiri dari 36 orang yang berasal dari kota :</p> <ul style="list-style-type: none"> - Bangkinang - Bengkalis - Kec Singingi - Kec Pangkalan Kerinci 	<p>Cluster 2 terdiri dari 46 orang yang berasal dari kota :</p> <ul style="list-style-type: none"> - Pekanbaru
--	--	---

Pembahasan

Dari hasil pengelompokan berdasarkan daerah asal diperoleh hasil bahwa Cluster 2 adalah mahasiswa STMIK Hang Tuah didominasi (Jumlah Mahasiswa dikategori Tinggi)) oleh mahasiswa yang berasal dari Kota Pekanbaru. Cluster 1 adalah mahasiswa yang berasal dari Bangkinang, Kec Bengkalis, Kec Kuantan Singingi dan Kec Pangkalan Kerinci (Jumlah mahasiswa dikategori sedang). Dan Cluster 0 adalah mahasiswa yang berasal dari Bangko, Duri, Indragiri Hulu, Jambi, Kampar Hilir, Kec Batang Kapas, Kec Kateman sampai dengan Tembilahan.

Berdasarkan Program Studi (sumbu Y) dan Kota Asal (sumbu X) hasilnya sebagai berikut :



Gambar 4.4 Rincian Cluster berdasarkan Program Studi dan Kota Asal

Hasilnya adalah jumlah Mahasiswa Sistem Informasi yang terbanyak berasal dari Kota asal Cluster 1 sedangkan Mahasiswa Teknik Informatika lebih banyak berasal dari Kota asal Cluster 2.

KESIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa ada 2 strategi promosi yang dapat dilakukan yaitu :

1. Melakukan promosi yang lebih banyak dan gencar terhadap kota-kota yang masih sedikit jumlah mahasiswanya yaitu kota-kota pada kluster 0.
2. Untuk kota-kota pada kluster 1 dan 2 tetap melakukan promosi meskipun jumlah mahasiswa banyak berasal dari kota-kota pada kluster 1 dan 2.

DAFTAR PUSTAKA

- Daniel T. Larose, 2005, *Discovering Knowledge in Data: an Introduction to Data Mining*, John Wiley & Sons
- Erwin, 2009, Analisis Market Basket dengan Algoritma Apriori dan FP-Growth, *Jurnal Vol 4*
- Ian H. Witten, Frank Eibe, Mark A. Hall, 2011, *Data mining: Practical Machine Learning Tools and Techniques 3rd Edition*, Elsevier
- Jiawei Han and Micheline Kamber, 2006, *Data Mining: Concepts and Techniques Second Edition*
- Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques Third Edition*, Elsevier, 2012
- Kurniawan, Harris, 2014, Analisa Pola Data Penyakit Rumah Sakit dengan Menerapkan Metode Association Rule Menggunakan Algoritma Apriori, *Seminar Nasional Informatika*
- Ririanti, 2014, Implementasi Algoritma FP-Growth pada Aplikasi Prediksi Persediaan Sepeda Motor, *Pelita Informatika Budi Darma*, Medan
- Ramadhani, Nilam, Badar, 2014, Analisis Pola Asosiasi dan Sekuensial Data Rekam Medis RSUD DR.H.Slamet Martodirdjo Pamekasan dengan Teknik Data Mining Menggunakan Algoritma Apriori, *Seminar Nasional Sistem Informasi*
- Markus Hofmann and Ralf Klinkenberg, 2014, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, CRC Press Taylor & Francis Group