



# ANALISIS DATA MAHASISWA DENGAN ALGORITMA K-MEAN UNTUK Mendukung Strategi Promosi STIKes HANG TUAH PEKANBARU

Eka Sabna

Teknik Informatika, STMIK Hang Tuah Pekanbaru Hang Tuah Pekanbaru

Email :

[es3jelita@yahoo.com](mailto:es3jelita@yahoo.com)

## Abstract

New student data owned by STIKes Hang Tuah Pekanbaru has many numbers with many variations of attributes stored in the database (database). But data data has not been utilized optimally, therefore required a method that can dig and extract new student data into a valuable information strategy. This study aims to classify the data of New Student STIKes Hang Tuah Pekanbaru by using Data Mining Technique. Mining Engineering The data used are Clustering technique with K-Means algorithm and for process using CRISP-DM method. This research is used to know the proper Promotion Strategy. Determining the right promotional strategy will be able to reduce promotional costs and achieve the right promotional goals. The results of the grouping of student data through the distribution of the region based on the Study Program and the City of Origin using K-Means clustering formed three clusters namely Cluster 2 with the number of 367 students coming from the city with code 1-10, Cluster 1 with the number of students 221 coming from the city with code 4-18 and cluster 0 with 63 students coming from the city code 14-37.

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

## Abstrak

Data mahasiswa baru yang dimiliki oleh STIKes Hang Tuah Pekanbaru telah banyak jumlahnya dengan banyak variasi atribut yang tersimpan dalam database (database). Namun data datanya belum dimanfaatkan secara optimal, oleh karena itu diperlukan suatu metode yang dapat menggali dan mengekstrak data siswa baru menjadi strategi informasi yang berharga. Penelitian ini bertujuan untuk mengklasifikasikan data Mahasiswa Baru STIKes Hang Tuah Pekanbaru dengan menggunakan Teknik Data Mining. Teknik Penambangan Data yang digunakan adalah teknik Clustering dengan algoritma K-Means dan untuk proses menggunakan metode CRISP-DM. Penelitian ini digunakan untuk mengetahui Strategi Promosi yang tepat. Menentukan strategi promosi yang tepat akan dapat mengurangi biaya promosi dan mencapai tujuan promosi yang tepat. Hasil pengelompokan data mahasiswa melalui persebaran wilayah berdasarkan Program Studi dan Kota Asal menggunakan K-Means clustering terbentuk tiga Cluster yaitu Cluster 2 dengan jumlah 367 mahasiswa yang berasal dari kota dengan kode 1-10, Cluster 1 dengan jumlah mahasiswa 221 yang berasal dari kota dengan kode 4-18 dan cluster 0 dengan jumlah mahasiswa 63 orang yang berasal dari kota dengan kode 14-37.

**Keywords:** Data Siswa Baru, Promosi, Data Mining, K-Means, CRISP-DM

## PENDAHULUAN

STIKes Hang Tuah Pekanbaru merupakan salah satu Sekolah Tinggi Kesehatan yang terkemuka di Pekanbaru. Mahasiswa STIKes 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 Perguruan 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 Clustering. 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 dengan Data mining untuk mendukung strategi promosi Perguruan Tinggi sebelumnya sudah pernah dilakukan, tetapi dengan tempat dan algoritma yang digunakan berbedabeda, salah satunya Penelitian yang menggunakan studi kasus President University dalam penerapan algoritma K-Means untuk menentukan strategi promosi ke daerah di seluruh Indonesia agar lebih efektif dan efisien.

## METODE PENELITIAN

### 1. Clustering

Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised) maksudnya metode ini diterapkan tanpa adanya latihan (training) dan tanpa ada guru (teacher) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu Hierarchical Clustering dan Non Hierarchical Clustering (Santosa, 2007). [3]

### 2. K-Means

K-Means clustering merupakan salah satu metode data clustering non hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007). [1]

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut :

- Pilih jumlah cluster k.
- Inisialisasi k
- Alokasi semua data/objek ke cluster terdekat. Kedekatan dua objek di tentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu di hitung jarak tiap data ke tiap pusat cluster. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut :

$$D(i, j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2}$$

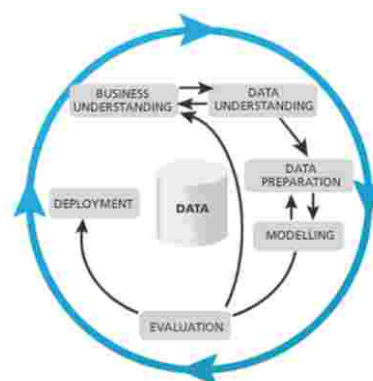
Dimana :

- D(i,j) = Jarak data ke i ke pusat cluster j  
 Xki = Data ke i pada atribut data ke k  
 Xkj = Titik pusat ke j pada atribut ke k

- Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata rata dari semua data/objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu satunya ukuran yang bisa dipakai.
- Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi. [3]

### 3 CRISP-DM

CRISP-DM merupakan singkatan dari Cross Industry Standard Process for Data Mining. CRISP-DM merupakan standarisasi data mining yang disusn oleh tiga penggagas data mining market yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR (Larose,2005). Pada metodologi ini dilakukan pembagian siklus proses data mining menjadi 6 tahap, dimana ketergantungan antara setiap tahap digambarkan dengan panah. Berikut gambaran dari metode CRISP-DM. [2]



Gambar 1 Tahapan CRISP-DM

## HASIL & PEMBAHASAN

### Tahapan CRISP-DM

#### 1. Pemahaman Bisnis (Business Understanding)

Tahap Pertama dari proses CRISP-DM adalah Pemahaman Bisnis (penelitian) dalam hal ini yang dilakukan adalah menentukan tujuan. Ada pun tujuan dari penelitian ini adalah untuk menganalisis data mahasiswa dengan algoritma k-means untuk mendukung strategi promosi STIKes Hang Buah Pekanbaru.

#### 2. Pemahaman Data (Data Understanding)

Data di peroleh dari database mahasiswa baru 2017 STIKes Hang Buah Pekanbaru sebanyak 651 mahasiswa. Atribut yang digunakan adalah Nama, Program Studi dan Tempat Tinggal.

#### 3. Persiapan Data (Data Preparation)

Dari 651 data mahasiswa maka akan dilakukan Teknik Data Preparation agar kualitas data diperoleh lebih baik dengan cara :

- 1.Data Validation, mengidentifikasi dan menghapus data yang tidak konsisten dan data yang tidak lengkap.
- 2.Data Itegration and Transformation, meningkatkan akurasi dan efisiensi algoritma. Pada Kota Asal (Tempat Tinggal) dilakukan peng-kodean .

Tabel 1. Inisial Kode Kota

ASAL	KODE
Pekanbaru	1
Indragiri Hilir	2
Kepulauan Meranti	3
Kampar	4
Lingga	5
Mandau	6
Bengkalis	7
Siak	8
Kepri	9
Dumai	10
Rahul	11
Kuantan Singingi	12
Pelalawan	13
Indragiri Hulu	14
Bintan	15
Tanjung Pinang	16
Batam	17
Rohil	18
50 Kota	19
Pangkalan Kerinci	20

ÇÄÅll 5 ÄÄn	وو
{j LÄ+	وو
Y šnll	یو
Y Änll	یو
[ ÄllÄ	یو
! Pš ll	یو
b ÄÇ Ä	یو
! ÄllÄ	یو
! ÄllÄ	لاو
h YÜ	هی
. ÄÄ+ / sÄff	وی
Yj Ä+	وی
{ ÄÄ+ ff	یو
ÇsÄllÄ	یو
{ ÄÄ	یو
. ÄÄ+	یو
! Pš ll { sÄÄ	یو

Pada Program Studi dilakukan peng-kodean seperti tabel berikut ini :

Tabel-2 Inisial Kode Program Studi

t wh Dw! a STUDI	Yh 5 9
a Y9 {	و
IYa .	و
DIDL	ی
IYa	ی
t IY9 {	ی
t {IY	ی
. I5 ! b	ی

#### 4.Fase Pemodelan (Modelling Phase)

Pada tahapan ini merupakan tahapan pemrosesan data training yang diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini menggunakan algoritma k-means.

#### 5. Fase Evaluasi (Evaluation Phase)

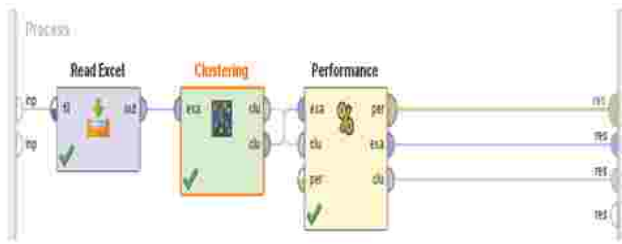
Pada fase ini dilakukan pengujian terhadap model klustering dengan algoritma k-means. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai dalam tahap business understanding.

#### 6.Fase Penyebaran

Berisi tentang hasil pengelompokkan berdasarkan kedekatan jarak antar titik pusat dengan data mahasiswa pada setiap atribut.

Implementasi dengan Rapidminer

Berikut adalah pengolahan data dengan menggunakan k-means pada RapidMiner:



Gambar 2 Pemodelan K-Means pada Rapidminer

Dengan menggunakan pemodelan k-means clustering seperti Gambar 3 diatas, dengan inialisasi jumlah cluster sebanyak 3, maka didapat hasil sbb :

**Cluster Model**

Cluster 0: 63 items  
 Cluster 1: 221 items  
 Cluster 2: 367 items  
 Total number of items: 651

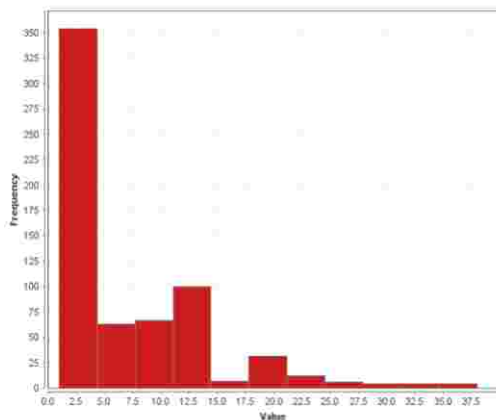
Gambar 3 Hasil cluster K-Means dalam implementasi Rapidminer

Cluster 0 adalah jumlah kota yang paling sedikit mahasiswa masuk, Cluster 1 adalah jumlah kota ke 2 mahasiswa masuk

Attribute	cluster_0	cluster_1	cluster_2
KodeKota	22.921	10.842	1.992
KodeProdi	4.730	4.837	4.627

Gambar 4 Hasil perhitungan jarak cluster dengan centroid dalam implementasi Rapidminer

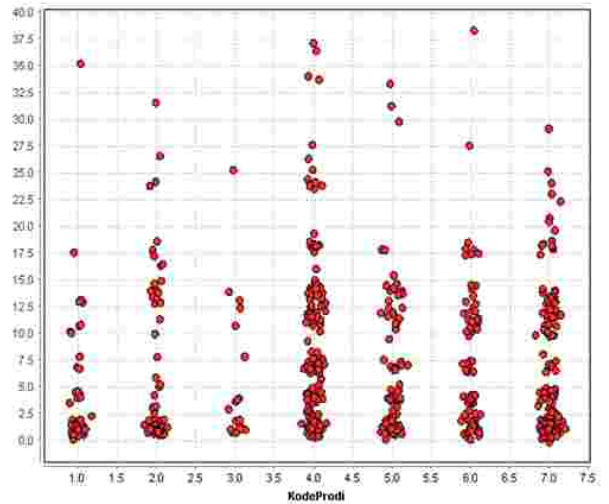
Analisis cluster dengan menggunakan chart adalah sbb :



Gambar 5 Cluster dalam implementasi RapidMiner berdasarkan Kota

Berdasarkan Chart diatas dapat dilihat 4 kota asal mahasiswa yang paling tinggi (terurut) yaitu :

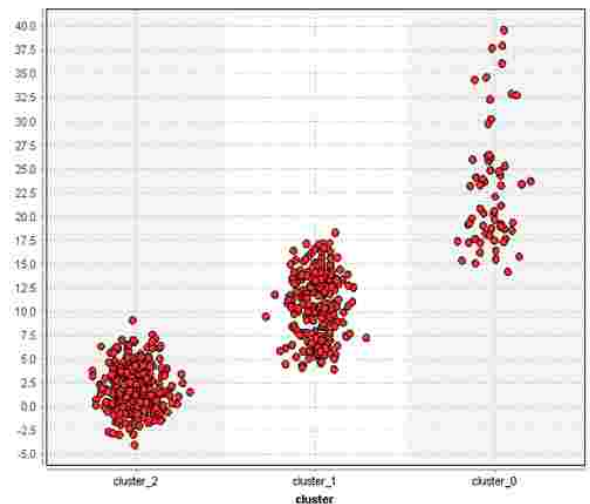
- 1.Asal kota mahasiswa dengan kode kota 1-4 yaitu Pekanbaru, INHIL, Kepulauan Meranti dan Kampar.
- 2.Asal kota mahasiswa dengan kode kota 11-14 yaitu Rohul, Kuantan Singingi, Pelalawan dan INHU.
- 3.Asal kota mahasiswa dengan kode kota 7-10 yaitu Bengkalis, Siak, Kepri dan Dumai
- 4.Asal kota mahasiswa dengan kode kota 5-6 yaitu Lingga dan Mandau.



Gambar 6 Cluster berdasarkan Kode Prodi

Berdasarkan Chart diatas dapat diperoleh informasi berdasarkan Program Studi pilihan mahasiswa (3 urutan tertinggi) adalah :

- a.Kode prodi 4 yaitu IKM.
- b.Kode Prodi 7 yaitu BIDAN
- c.Kode Prodi 5 yaitu PIKES



Gambar 7 Cluster berdasarkan Kode Kota

Berdasarkan Chart diatas dapat diperoleh informasi Cluster dari pengolahan data yaitu :

- a.Cluster 0 dengan jumlah mahasiswa yang paling sedikit yaitu 63 yang berasal dari kota dengan kode 14-37.
- b.Cluster 1 dengan jumlah 221 yang berasal dari kota dengan kode 4-18.
- c.Cluster 2 dengan jumlah 367 yang berasal dari kota dengan kode 1-10.

## SIMPULAN DAN SARAN

Setelah dilakukan pengelompokan data mahasiswa melalui persebaran Kota Asal menggunakan K-Means clustering maka dapat di simpulkan sebagai berikut :

- a. Setelah dilakukan pengelompokan data mahasiswa melalui persebaran wilayah berdasarkan Program Studi dan Kota Asal menggunakan K-Means clustering terbentuk tiga Cluster yaitu Cluster 2 dengan jumlah 367 mahasiswa , Cluster 1 dengan jumlah mahasiswa 221 dan cluster 0 dengan jumlah mahasiswa 63 orang.
- b. Dari Pengelompokan data maka dapat ditentukan Strategi Promosi untuk calon mahasiswa baru yang tepat sasaran yaitu dengan mengirimkan tim promosi yang sesuai dengan asal kota mahasiswa dan berdasarkan Program Studi yang di minati.

## DAFTAR PUSTAKA

- Agusta, Y. 2007. K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika Vol 3* (Februari 2007): 47-60.
- Arlinda, Zefrr and Arie, Vatesia and Boko, susilo. Implementasi Algoritma K-Means untuk mengetahui Sebaran Mahasiswa Sebagai Informasi Penunjang Promosi Universitas (Studi Kasus : Mahasiswa Fakultas Keguruan dan Ilmu Pendidikan Universitas Bengkulu).
- Daniel T. Larose, 2005, *Discovering Knowledge in Data: an Introduction to Data Mining*, John Wiley & Sons
- Han,J . and Kamber, M. *Data Mining Concept and Technique*, 2nd.
- Hidayat, T., dan Istiadah, N.2011. *Panduan Lengkap Menguasai SPSS 19 untuk mengolah Data Statistik Penelitian*. Jakarta : Media Kita
- Johan Oscar Ong. Implementasi Algoritma K-Means untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknikn Industri*.2013; vol 12 (no 1): 10-20.
- Larose, Daniel d. 2006. *Data Mining Methods and Models*. Hoboken New Jersey: Jhon Wiley & Sons, Inc.

Kusrini dan Emha Taufiq Luthfi.2009. *Algoritma Data Mining*. Jakarta : Graha Andi.

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

Rima, Ramadhani Dias. 2014. *Data Mining Menggunakan Algorithma K-Means Clustering untuk menentukan strategi Universitas Dian Nuswantoro*. Skripsi. Fakultas Ilmu Komputer

Turban,E.dkk. 2005. *Decision Support System and Intelligent System: andi Offset*.



- itSMF. (2007). An Introductory Overview of ITIL v3, Version 1.0. London : The UK Charter of the itSMF
- Mallete, Debra. (2012). COBIT 5 Process Assessment Methode (PAM). Paper presented at 2012 SF ISACA Fall Conference, San Fransisco, CA
- Lulu, Y. D. (2013). Analisa Teori IT Governance menggunakan COBIT 5. Jurnal Teknik Elektro dan Komputer Instiut Teknologi Sepuluh November
- Saputra, I. N. (2013). Pengukuran Tingkat Kapabilitas dan Perbaikan Tata Kelola Teknologi Informasi Berdasarkan Kerangka Kerja COBIT 5 dan ITIL V3 2011 : Studi Kasus PT. XYZ. Karya Akhir MTI Universitas Indonesia