

# Analisa Potensi Mahasiswa Di Daerah Bali Menggunakan Pendekatan K-Mean Clustering

**Anggun Nugroho**

Program Studi Sistem Komputer STMIK STIKOM Bali  
Jl. Raya Puputan No. 86 Renon, Denpasar, telp/fax 0361 24445/0361 264773  
e-mail: [anggun.bali@gmail.com](mailto:anggun.bali@gmail.com)

## *Abstrak*

*Dalam era globalisasi yang sarat akan tantangan, persaingan dan perkembangan ilmu pengetahuan serta teknologi mengakibatkan kebutuhan akan teknologi semakin besar. Sebagai salah satu kampus yang terbesar di pulau Bali, tentunya juga harus mengetahui potensi mahasiswa disetiap daerah di pulau Bali ini untuk keperluan analisis, baik analisis untuk marketing ataupun pengembangan lainnya. Sehubungan dengan hal tersebut di atas, maka peneliti akan melakukan sebuah perhitungan potensi di setiap daerah di Bali untuk mahasiswa STIKOM Bali selama 3 tahun terakhir. Untuk membuat hasil analisa menjadi akurat diperlukan sebuah metode untuk melakukan analisa ini, yaitu dengan menggunakan metode K-Mean Clustering. Dimana hasil dari penelitian ini adalah sebuah hasil analisa sistem pengelolaan potensi mahasiswa STIKOM Bali di setiap daerah di pulau Bali.*

**Kata kunci:** *Analisis, K-Mean Clustering, Mahasiswa, Pulau Bali*

## *Abstract*

*In this era of globalization which will be full of challenges, competition and the development of science and technology resulting in greater demand for the technology. , As one of the largest campuses in the island of Bali, of course, also need to know the potential of students in each region on the island of Bali for the purposes of analysis, both the analysis for marketing or other development. In connection with the foregoing, the researchers will perform a calculation of the potential in each area in Bali for students STIKOM Bali over the last 3 years. To make an accurate analysis results become required a method to perform this analysis, ie using K-Mean Clustering. Where the results of this study are a potential student management system design STIKOM Bali in every region on the island of Bali.*

**Keywords:** *Analysis, K-Mean Clustering, Student, Bali Island*

## **1. Pendahuluan**

Dalam era globalisasi yang sarat akan tantangan, persaingan dan perkembangan ilmu pengetahuan serta teknologi mengakibatkan kebutuhan akan teknologi semakin besar. Salah satunya adalah kebutuhan akan komputer oleh berbagai kalangan maupun organisasi. Disamping itu komputer juga digunakan sebagai sarana untuk pengolahan data serta pendataan pada sebuah permasalahan, hingga pembuatan laporan yang nantinya akan dijadikan sebagai tolak ukur perkembangan dari permasalahan tersebut.

STIKOM Bali sebagai Sekolah Tinggi Komputer pertama di Bali, didirikan untuk mengantisipasi era atau abad informasi dan tentu saja membutuhkan SDM (Sumber Daya Manusia) yang memahami dunia teknologi informasi dan komputer. Sebagai salah satu kampus yang terbesar di pulau Bali, tentunya juga harus mengetahui potensi mahasiswa disetiap daerah di pulau Bali ini untuk keperluan analisis, baik analisis untuk marketing ataupun pengembangan lainnya.

Sehubungan dengan hal tersebut di atas, maka peneliti akan melakukan sebuah perhitungan potensi di setiap daerah di Bali untuk mahasiswa STIKOM Bali selama 3 tahun terakhir. Sehingga dalam penyajian laporan dan informasi dapat diperoleh data distribusi potensi mahasiswa di masing-masing daerah dengan maksud untuk kegiatan marketing di tahun-tahun berikutnya secara cepat, tepat dan

lengkap tanpa harus melalui proses pencatatan yang berulang-ulang. Untuk membuat hasil analisa menjadi akurat diperlukan sebuah metode untuk melakukan analisa ini, yaitu dengan menggunakan metode K-Mean Clustering.

**2. Metode Penelitian**

Metode Penelitian memberikan penjelasan tentang langkah-langkah, data, lokasi penelitian, metode evaluasi yang digunakan serta penjelasan terstruktur tentang algoritma atau metode dari penelitian yang dibahas.

**2.1. Metode Pengumpulan Data**

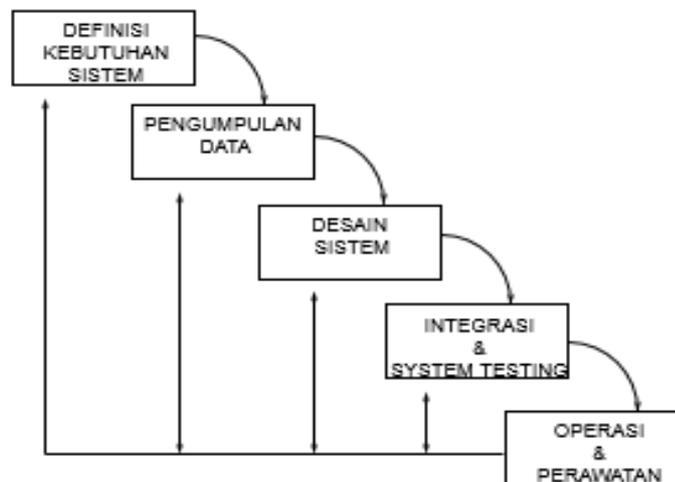
Metode yang digunakan dalam rangka pengumpulan data-data yang diperlukan dalam penelitian ini adalah Studi literatur dari sumber-sumber kepustakaan sebagai landasan dalam menganalisis permasalahan yang disusun dalam penelitian ini. Dalam penyusunan penelitian ini, metode yang dipakai oleh penulis adalah *Literatur Review*. Metode pengumpulan data dan informasi dengan cara menggali pengetahuan atau ilmu dari sumber-sumber seperti buku, karya tulis, jurnal ilmiah, makalah, dan sumber lain yang berhubungan dengan objek penelitian khususnya terkait dengan permasalahan

**2.2. Data**

Penyusunan penelitian ini menggunakan data-data yang mendukung pelaksanaan dari proses penelitian yang dilakukan. Adapun hal-hal yang menyangkut data-data tersebut adalah Sumber data yang digunakan dalam penelitian ini diperoleh dari hasil pengumpulan data dan simulasi, dan pada penelitian ini digunakan jenis data sekunder yaitu data yang didapatkan dari studi kepustakaan.

**2.3 Alur Analisis**

Berdasarkan data yang telah diperoleh, maka pembahasan penelitian ini dilakukan dengan langkah-langkah sebagai berikut:



Gambar 1. Langkah-langkah Penelitian

Proses pelaksanaan penelitian yang akan dilaksanakan dalam waktu 5 (lima bulan). Sedangkan untuk tempat penelitian ini dilaksanakan di STIKOM Bali, Jl.Raya Puputan No.86 Renon Denpasar.

Penelitian ini memerlukan data-data penunjang yang digunakan adalah jumlah data mahasiswa per kabupaten selama 5 tahun terakhir. Adapun jenis data yang dipakai dalam

- a. Data primer  
Data primer dalam penelitian ini adalah jumlah data mahasiswa yang masuk ke STIKOM Bali selama 5 tahun terakhir
- b. Data sekunder  
Data sekunder meliputi gambaran umum K-Mean Clustering, dan hal lain yang ada kaitannya.

Metode pengumpulan data merupakan bagian paling penting dalam sebuah penelitian. Ketersediaan data akan sangat menentukan dalam proses pengolahan dan analisa selanjutnya. Karenanya, dalam pengumpulan data harus dilakukan teknik yang menjamin bahwa data diperoleh itu benar, akurat dan bisa dipertanggungjawabkan sehingga hasil pengolahan dan analisa data tidak bias.

Data yang dikumpulkan dalam penelitian ini merupakan data primer dan sekunder yang diperoleh dari berbagai sumber. Teknik pengumpulannya dilakukan melalui beberapa langkah yakni:

- a. Data primer diperoleh melalui :
  - Observasi, yaitu dilakukan dengan melakukan pengamatan secara langsung terhadap obyek penelitian, selama periode waktu tertentu.
- b. Data sekunder meliputi gambaran umum sistem informasi. Data sekunder diperoleh melalui:
  1. Studi dokumentasi  
Studi dokumentasi digunakan untuk mencari data-data sekunder yang dibutuhkan dalam pembuatan sistem informasi.
  2. Akses internet  
Akses internet digunakan untuk mencari data-data pendukung dari berbagai buku, ebook, maupun jurnal-jurnal yang disediakan di internet.
  3. Studi yang relevan  
Studi yang relevan ini digunakan sebagai acuan dalam melakukan penelitian.

### 3. Hasil dan Pembahasan

#### 3.1 Algoritma K-Mean Clustering

K-means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam cluster yang lain [1].

Clustering adalah suatu alat untuk analisa data, yang memecahkan permasalahan penggolongan. Obyek nya adalah untuk kasus sebaran mahasiswa STIKOM berdasarkan geografis di pulau Bali ke dalam kelompok, sedemikian sehingga derajat tingkat keterhubungan antar anggota cluster yang sama adalah kuat dan lemah antar anggota dari cluster yang berbeda [2]. Hasil dari analisis cluster definisi dari penyebaran data mahasiswa baru selama 3 tahun terakhir di STIKOM Bali. Hasilnya adalah akan menjadi beberapa kluster berdasarkan kabupaten/kotamadya di Bali.

Objective function yang berusaha diminimalkan oleh k-means adalah:

$$J(U, V) = \sum_{k=1}^N \sum_{i=1}^c (a_{ik} * (x_k - v_i)^2)$$

dimana:

U : Matriks keanggotaan data ke masing-masing cluster yang berisikan nilai 0 dan 1

V : Matriks centroid/rata-rata masing-masing cluster

N : Jumlah data

c : Jumlah cluster

$a_{ik}$  : Keanggotaan data ke-k ke cluster ke-i

$x_k$  : data ke-k

$v_i$  : Nilai centroid cluster ke-i

Prosedur yang digunakan dalam melakukan optimasi menggunakan k-means adalah sebagai berikut:

Step 1. Tentukan jumlah cluster

Step 2. Alokasikan data ke dalam cluster secara random

Step 3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster.

Step 4. Alokasikan masing-masing data ke centroid/rata-rata terdekat

Step 5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan, di atas nilai threshold yang ditentukan

Centroid/rata-rata dari data yang ada di masing-masing cluster yang dihitung pada Step 3. didapatkan menggunakan rumus sebagai berikut:

$$v_{ij} = \sum_{k=0}^{N_i} (x_{kj}) / N_i$$

dimana:

i,k : indeks dari cluster

j : indeks dari variabel

$v_{ij}$  : centroid/rata-rata cluster ke-i untuk variabel ke-j

$x_{kj}$  : nilai data ke-k yang ada di dalam cluster tersebut untuk variabel ke-j  
 $N_i$  : Jumlah data yang menjadi anggota cluster ke-i

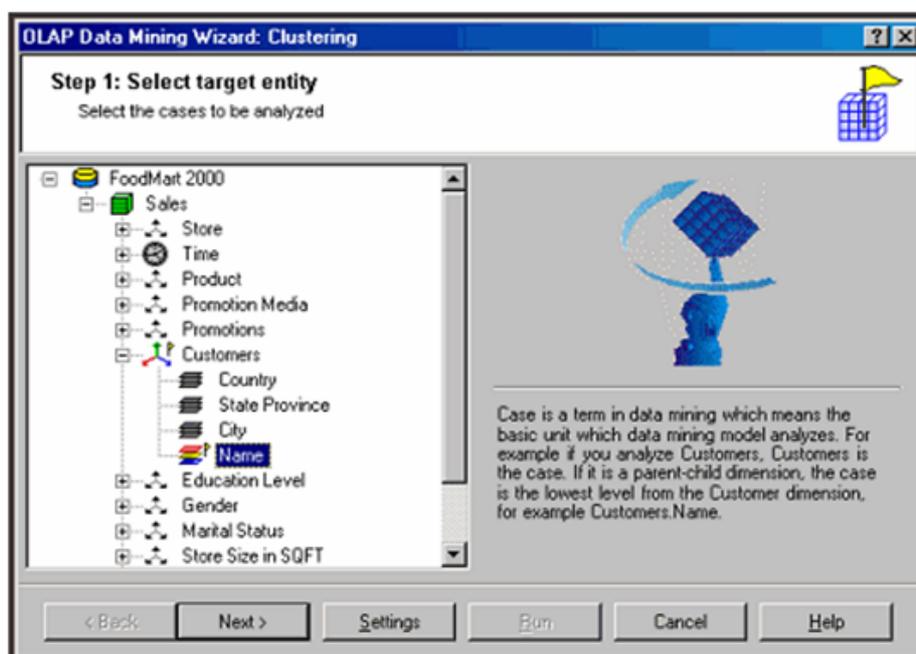
Sedangkan pengalokasian data ke masing-masing cluster yang dilakukan pada Step 4. dilakukan secara penuh, dimana nilai yang memungkinkan untuk  $a_{ik}$  adalah 0 atau 1. Nilai 1 untuk data yang dialokasikan ke cluster dan nilai 0 untuk data yang dialokasikan ke cluster yang lain. Dalam menentukan apakah suatu data teralokasikan ke suatu cluster atau tidak, dapat dilakukan dengan menghitung jarak data tersebut ke masing-masing centroid/rata-rata masing-masing cluster. Dalam hal ini,  $a_{ik}$  akan bernilai 1 untuk cluster yang centroidnya terdekat dengan data tersebut, dan bernilai 0 untuk yang lainnya[3].

Terdapat Database Mahasiswa dengan range 3 tahun terakhir. Kita akan mencari dan menemukan 3 segmen kabupaten dari data mahasiswa yang diambil sampling secara random. Tujuannya adalah untuk meningkatkan pemasaran atau promosi ke kabupaten tersebut agar serapan nya bisa bertahan atau terus meningkat. Kita akan menggunakan algoritma clustering di database dimana segmen pelanggan ada di OLAP dalam tiga kategori yang berdasarkan pada informasi berikut: Jenis kelamin, Desa, Kecamatan, Banjar, Kabupaten, atau Kotamadya.

### 3.2 Simulasi clustering

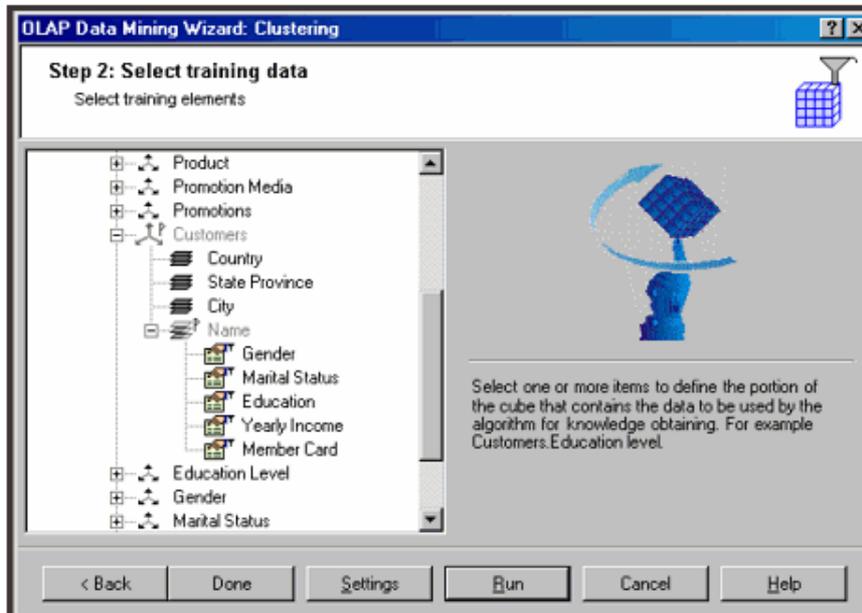
Untuk mendapatkan cluster dari data training menggunakan OLAP Data Mining Wizard yang sudah ada di dalam database maka langkah-langkah berikut yang harus dilakukan :

1. Ketika akan menggolongkan mahasiswa kita pilih customer (sebagai simulasi) pada OLAP Data.



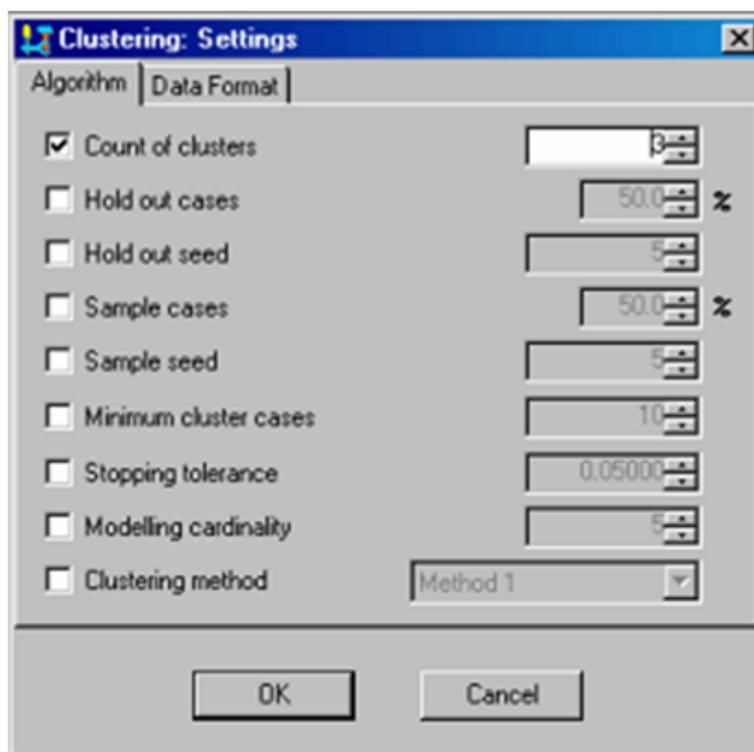
Gambar 2. Langkah 1 Simulasi menggunakan OLAP Data Mining Wizard

2. Pilih karakter (dalam hal ini kita asumsikan sebagai kategori) yang akan diproses oleh algoritma tersebut.



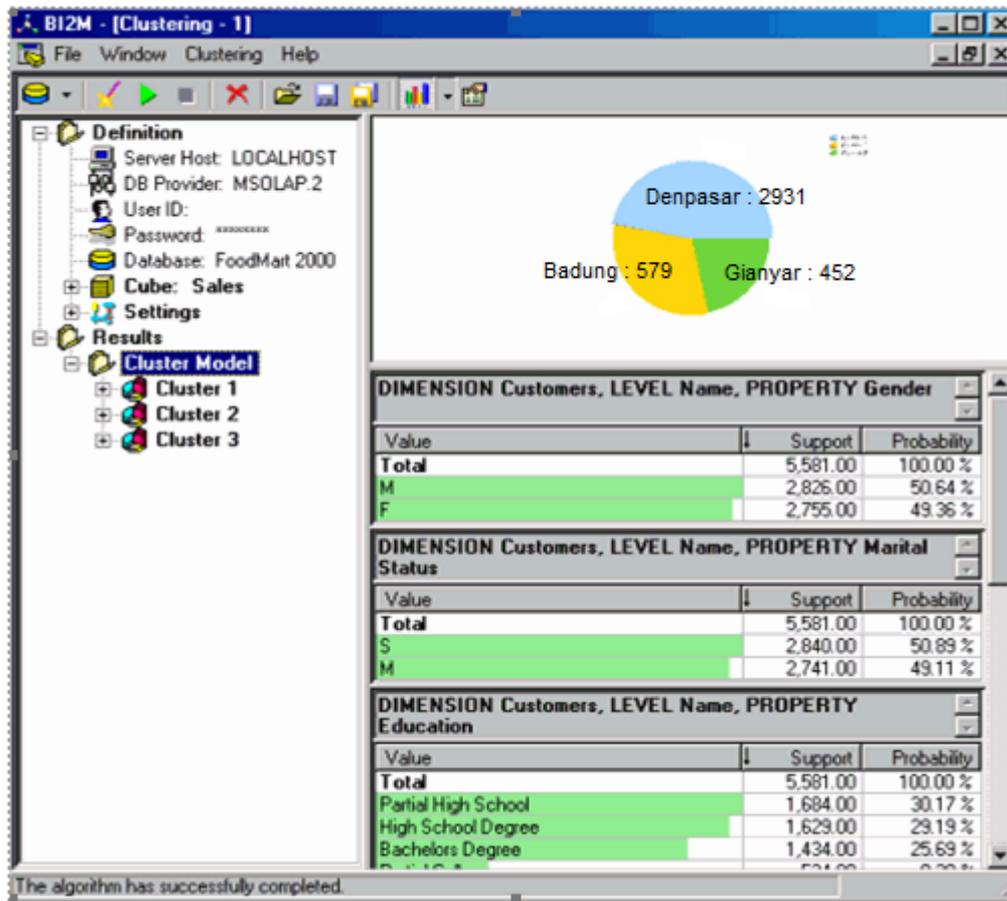
Gambar 3. Langkah 2 Simulasi menggunakan OLAP Data Mining Wizard

3. Tentukan banyaknya cluster yang akan terbentuk berdasarkan kabupaten/Kotamadya.



Gambar 4. Langkah 3 Simulasi menggunakan OLAP Data Mining Wizard

4. Keluar hasil dari simulasi clustering dari data sampling yang diambil secara acak.



Gambar 5. Langkah 4 Simulasi menggunakan OLAP Data Mining Wizard

Dengan menggunakan OLAP Datamining Wizard dihasilkan data bahwa Kabupaten/Kotamadya terbesar dalam penyerapan mahasiswa baru ke STIKOM pada rentan waktu 3 tahun terakhir adalah seperti yang tersaji pada gambar 4. Simulasi ini dilakukan dengan cara melakukan import data sampling yang mahasiswa STIKOM Bali selama 5 Tahun terakhir, dan aplikasi melakukan proses clustering sehingga didapatkan informasi seperti ini.

#### 4. Simpulan

Dari penelitian yang dilakukan maka didapatkan beberapa kesimpulan diantaranya adalah :

1. Simulasi data penyebaran mahasiswa dengan menggunakan data 5 tahun terakhir.
2. Data yang dicluster hanya data 3 tahun terakhir dari data yang diimport yaitu 5 tahun terakhir.
3. Simulasi menggunakan OLAP Data Mining Wizard dan didapatkan hasil Kotamadya Denpasar menempati posisi pertama, dan setelah itu disusul kabupaten Badung dan Kabupaten Gianyar.

#### Daftar Pustaka

- [1] Beddo, Vanessa, Application of Parallel Programming in Statistics, California: University of Los Angeles, 2002.
- [2] Liao, Wei-Keng, Parale; K-Means Data Clustering, [Online]. Available : <http://users.eecs.northwestern.edu/~wkliao/Kmeans/>). Diakses pada tanggal 22 Nopember 2015.
- [3] Teknomo, Kardi, K-Means Clustering Tutorial, [Online]. Available : (<http://people.revoledu.com/kardi/tutorial/kMean/>). Diakses pada tanggal 21 Mei 2013.
- [5] Yuan, M., & Vogel, D. (2006). Cultural Impact on Intergroup Coordination in Software Development in China: A Qualitative Analysis. Proceedings of the 39th Hawaii International Conference on System Sciences .