

BAB II

LANDASAN TEORI

2.1. Sistem Informasi

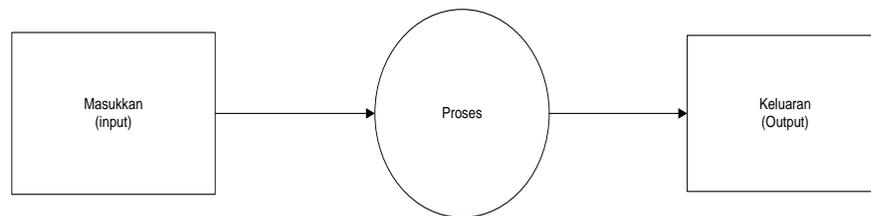
Pengertian Sistem Informasi pada dasarnya merupakan hasil dari dua arti, yakni sistem dan informasi yang digabungkan. Berikut definisi sistem menurut para ahli ini, Menurut (Gordon B. Davis, 2002) "Secara sederhana sistem dapat diartikan sebagai suatu kumpulan atau himpunan dari unsur, komponen, atau variabel-variabel yang terorganisasi, saling berinteraksi, saling tergantung satu sama lain dan terpadu. Sistem bisa berupa abstraksi atau fisis." Sedangkan definisi sistem informasi menurut (Tata Sutabri, 2004) mengemukakan, "Sistem yang abstrak adalah susunan yang teratur dari gagasan-gagasan atau konsepsi yang saling tergantung. Sedangkan sistem yang bersifat fisis adalah serangkaian unsur yang bekerjasama untuk mencapai suatu tujuan."

Definisi Informasi disini merujuk pada data-data yang telah dibuat sedemikian rupa sehingga menjadi sebuah informasi, seperti pernyataan ahli berikut, "Data adalah kumpulan fakta yang tidak terorganisir. Pengolahan data akan mengubah data mentah menjadi informasi" menurut (Nurwono, 1994). Sedangkan pengertian informasi sendiri, menurut (Tata Sutabri, 2004) "Informasi adalah hasil pengolahan data yang diperoleh dari setiap elemen sistem menjadi bentuk yang mudah dipahami oleh penerimanya dan informasi ini menggambarkan kejadian-kejadian nyata untuk menambah pemahamannya terhadap fakta-fakta yang ada, sehingga dapat digunakan untuk pengambilan suatu keputusan."

Pada pendapat yang telah dikemukakan mengenai sistem dan informasi di atas, telah dirumuskan mengenai pengertian sistem informasi seperti berikut, Menurut John F. Nash "Sistem Informasi adalah kombinasi dari manusia, fasilitas atau alat teknologi, media, prosedur dan pengendalian yang bermaksud menata jaringan komunikasi yang penting, proses atas transaksi-transaksi tertentu dan rutin,

membantu manajemen dan pemakai intern dan ekstern dan menyediakan dasar pengambilan keputusan yang tepat.”

Sistem merupakan kumpulan elemen yang saling terkait dan bekerja sama, saling berhubungan dan berinteraksi untuk mengolah data masukan (input) untuk mencapai tujuan tertentu sampai menghasilkan keluaran (output) yang diinginkan (Kristanto, 2003).



Gambar 2.1. Model Dasar Sistem

(Sumber : Kristanto, 2003)

Definisi-definisi para ahli di atas, kemudian dapat kita rangkum tentang definisi sistem informasi, merupakan kombinasi dari teknologi informasi dan aktivitas orang yang menggunakan teknologi itu untuk mendukung operasi dan manajemen. Penggunaan teknologi disini merujuk pada istilah yang digunakan Teknik Informasi dan Komunikasi (TIK) pada penggunaan database sebagai basis data. Kemudian jika kita melihat pengertian sistem informasi yang berinteraksi dengan proses bisnis dan melihat pada pengertian organisasi sendiri, sistem informasi merupakan sekumpulan informasi pada sebuah basis data, yang menggunakan model dan media teknologi informasi yang berguna dalam pengambilan keputusan bisnis, pada sebuah organisasi.

2.2. Pengelompokan Data (*Clustering*)

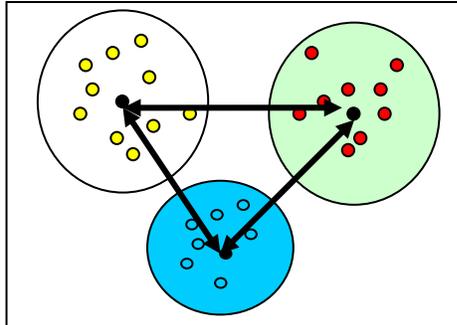
Clustering (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. *Clustering* membagi kumpulan data ke dalam beberapa

kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar daripada diantara kelompok-kelompok (Rui Xu dan Donald. 2009). Gagasan mengenai pengelompokan data atau *clustering*, memiliki sifat yang sederhana dan dekat dengan cara berpikir manusia, kapanpun kepada kita dipresentasikan jumlah data besar ini ke dalam sejumlah kecil kelompok-kelompok atau kategori-kategori untuk memfasilitasi analisisnya lebih lanjut. Selain dari itu sebagian besar data yang dikumpulkan dalam banyak masalah terlihat memiliki beberapa sifat yang melekat yang mengalami pengelompokan-pengelompokan natural (Hammuda dan Karay, 2003).

Algoritma-algoritma *clustering* digunakan secara ekstensif tidak hanya untuk mengorganisasikan dan mengkategorikan data, akan tetapi juga sangat bermanfaat untuk kompresi data dan konstruksi model. Melalui pencarian kesamaan dalam data, seseorang dapat mempresentasikan data yang sama dengan lebih sedikit simbol misalnya. Juga, jika kita dapat menemukan kelompok-kelompok data, kita dapat membangun sebuah model masalah berdasarkan pengelompokan-pengelompokan ini (Dubes dan Jain, 1988). *Clustering* sering dilaksanakan sebagai langkah pendahuluan dalam proses pengumpulan data. Dengan *cluster-cluster* yang dihasilkan digunakan sebagai input lebih lanjut ke dalam sebuah teknik yang berbeda, seperti natural diatas dapat diperoleh sebagai jarak dari pembaharuan formula Lance-Williams (Lance & Williams, 1967).

Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. Solusi analisis *cluster* bersifat tidak unik, anggota *cluster* untuk tiap penyelesaian/solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih. Solusi *cluster* secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan variabel-variabel yang relevan dapat mempengaruhi substansi hasil analisis *cluster*.

2.3. *K-Means Clustering*



Gambar 2.3. *Clustering*

(Sumber : Yudi Agusta, 2007)

Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. *K-Means* merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster/* kelompok.

Metode ini mempartisi data ke dalam *cluster/* kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*. Manfaat *Clustering* adalah sebagai Identifikasi *Object (Recognition)* misalnya dalam bidang *Image Processing*, *Computer Vision* atau *robot vision*. Selain itu adalah sebagai Sistem Pendukung Keputusan dan *Data Mining* seperti Segmentasi pasar, pemetaan wilayah, Manajemen marketing dll.

Data *clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut (Yudi Agusta, 2007) :

1. Tentukan jumlah *cluster*
2. Alokasikan data ke dalam *cluster* secara *random*
3. Hitung *centroid*/ rata-rata dari data yang ada di masing-masing *cluster*
4. Alokasikan masing-masing data ke *centroid*/ rata-rata terdekat
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.

Karakteristik *K-means* :

1. *K-means* sangat cepat dalam proses *clustering*.
2. *K-means* sangat sensitive pada pembangkitan *centroid* awal secara random.
3. Memungkinkan suatu *cluster* tidak mempunyai anggota.
4. Hasil *clustering* dengan *K-means* bersifat unik (selalu berubah-ubah, terkadang baik, terkadang jelek).

Distance Space Untuk Menghitung Jarak Antara Data dan Centroid

Beberapa *distance space* telah diimplementasikan dalam menghitung jarak (*distance* antara data dan *centroid*) termasuk di antaranya L1 (*Manhattan/ City Block distance space*), L2 (*Euclidean distance space*), dan Lp (*Minkowski distance space*). Jarak antara dua titik x_1 dan x_2 pada *Manhattan/City Block distance space* dihitung dengan menggunakan rumus sebagai berikut (Yudi Agusta, 2007)

$$D_{L_1}(x_2, x_1) = \|x_2 - x_1\|_1 = \sum_{j=1}^p |x_{2j} - x_{1j}| \dots\dots\dots (2.1)$$

Sedangkan untuk L2 (*Euclidean distance space*), jarak antara dua titik dihitung menggunakan rumus sebagai berikut:

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \dots\dots\dots (2.2)$$

Dimana :

D_{L2} = jarak kuadrat *Eucliden* antar objek ke x2 dengan x1.

P = jumlah variabel *cluster*.

x_{2j} = nilai atau data dari objek ke-2 pada variabel ke- j .

x_{1j} = nilai atau data dari objek ke-1 pada variabel ke- j (Everitt, 1993).

2.4. Definisi Beasiswa

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan, mahasiswa atau pelajar yang digunakan demi keberlangsungan pendidikan yang ditempuh. (Menurut Murniasih (2009) beasiswa diartikan sebagai bentuk penghargaan yang diberikan kepada individu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi. Penghargaan itu dapat berupa akses tertentu pada suatu institusi atau penghargaan berupa bantuan keuangan. Pada dasarnya, beasiswa adalah penghasilan bagi yang menerimanya. Hal ini sesuai dengan ketentuan pasal 4 ayat (1) Undang-undang PPh/2000. Disebutkan pengertian penghasilan adalah tambahan kemampuan ekonomis dengan nama dan dalam bentuk apa pun yang diterima atau diperoleh dari sumber Indonesia atau luar Indonesia yang dapat digunakan untuk konsumsi atau menambah kekayaan Wajib Pajak. Karena beasiswa bisa diartikan menambah kemampuan ekonomis bagi penerimanya, berarti beasiswa merupakan penghasilan.

Beasiswa dapat diberikan oleh lembaga pemerintah, perusahaan ataupun yayasan. Pemberian beasiswa dapat dikategorikan pada pemberian cuma-cuma ataupun pemberian dengan ikatan kerja (biasa disebut ikatan dinas) setelah selesainya pendidikan. Lama ikatan dinas ini berbeda-beda, tergantung pada lembaga yang memberikan beasiswa tersebut. beasiswa juga banyak diberikan kepada perkelompok (group) misalnya ketika ada event perlombaan yang diadakan oleh lembaga pendidikan, dan salah satu hadiahnya adalah beasiswa.

2.5. Studi Literatur

Penggunaan metode *clustering k-means* ada di beberapa sumber jurnal pendukung, seperti :

1. Dalam jurnal SNTIKI, vol (5), Hal 395-398, oleh Nengsih W, Febiyanto pada tahun 2012 dengan judul **“Data Mining Analysis Pengelompokan Penerima Beasiswa Menggunakan Teknik *Clustering K-Means* (Studi Kasus Universitas X)”** seperti abstrak dibawah ini.

Kehadiran beasiswa pada perguruan tinggi memberikan solusi untuk membantu biaya pendidikan mahasiswa. Jenis beasiswa yang ditawarkan beragam memiliki kriteria dan persyaratan masing-masing. Antusias mahasiswa untuk mendaftar jalur beasiswa sangat tinggi, hal itu terlihat dari jumlah calon pendaftar beasiswa. Sehingga untuk memudahkan dalam proses pengelompokan penerima beasiswa digunakanlah teknik *data mining clustering k-means analysis* berdasarkan beberapa atribut yang memiliki pola kemiripan yang sama. Sementara untuk variabel kemiripan pola yang digunakan untuk proses *clustering* penerima beasiswa adalah IPK, semester, kegiatan ekstrakurikuler dan tingkatan ekonomi. Aplikasi ini dapat mempermudah pihak yang berwenang dalam pengelolaan beasiswa pada universitas serta meminimalisir kesalahan dalam hal penempatan beasiswa. Dari implementasi metode *k-means* yang sudah dilakukan dapat diambil kesimpulan bahwa 4 mahasiswa berada pada *cluster* 1 kategori beasiswa KBM, 16 mahasiswa berada pada *cluster* beasiswa PPA. Sementara tidak ada mahasiswa dengan karakteristik yang di isyaratkan untuk mendapatkan jenis beasiswa C2, kategori beasiswa bantuan TA. (*Studi Kasus Universitas X*).

Kata Kunci : *Clustering, K-Means, Beasiswa, Data Mining*

2. Dalam jurnal SNTIKI, vol (4), hal 2085-9902, oleh Mustakim, pada tahun 2012, dengan judul **“Pemetaan Digital dan Pengelompokan Lahan Hijau di Wilayah Provinsi Riau Berdasarkan *Knowledge Discovery in***

Databases (KDD) Dengan Teknik K-Means Mining” seperti abstrak dibawah ini.

Perkembangan teknologi semakin meningkat disemua kalangan, baik di lingkungan pendidikan, instansi maupun perusahaan, teknologi tersebut salah satunya adalah internet dan pemetaan. Peranan penting peta adalah memberikan informasi kepada pembaca atau akses digital, pada kasus ini dilakukannya pemetaan digital untuk pengelompokan lahan hijau di Riau dengan menggunakan KDD teknik *data mining*. Lahan hijau yang dimaksudkan pada topik yang dibahas adalah Hutan. Sistem ini memberikan informasi berupa pengelompokan pada daerah kecamatan di Riau, dengan hasil akhir sebuah peta digital lahan hijau, selain itu detail informasi mengenai atribut daerah akan tampil sesuai dengan informasi masing-masing daerah kecamatan. Sistem dikembangkan dengan bahasa pemrograman PHP dengan bantuan mapping pada google maps. Pengguna dari sistem tersebut terdiri atas admin, pimpinan, dan pengguna biasa berdasarkan hak aksesnya masing-masing.

Kata Kunci : *Data Mining, K-Means, Knowledge Discovery in Databases (KDD), Pemetaan digital*

3. Dalam jurnal SNTIKI, vol (5), Hal 415-428, oleh Irsyad Muhammad, pada tahun 2012 dengan judul **“Analisis Clustering Data Lagu dengan Metode Agglomerative Hierarchical Clustering”** seperti abstrak dibawah ini.

Perkembangan industri musik menyebabkan penambahan koleksi musik atau lagu yang harus disiarkan semakin beragam. Lagu yang disiarkan adalah jenis musik *digital* yang memiliki beberapa lagu atribut informasi data musik/lagu digital. Atribut penyiaran lagu dicatat pada sebuah *log* penyiaran yang telah disediakan oleh *player* stasiun Radio. Pada penelitian ini dibangun sebuah aplikasi pengelompokan atribut informasi lagu memanfaatkan hasil penyiaran yang tercatat dalam *log player* berbasis *desktop* dengan menggunakan metode *Agglomerative Hierarchical*

Clustering. Metode pengelompokan ini menggabungkan tiap kata yang sama menjadi kumpulan *cluster* dan sub *cluster*. Perhitungan yang dilakukan adalah dengan mencari jarak kemiripan antar kata dengan menggunakan *euclidean distance* dan *single linkage clustering* untuk menggabungkan dua buah *cluster*. Dari hasil pengujian yang di dapatkan data lagu yang dikelompokkan berdasarkan kata yang dipisahkan dengan (spasi). Jadi setiap kata yang dipisahkan dengan (spasi) akan dikelompokkan berdasarkan kesamaan antar kata yang ada pada atribut lagu.

Kata Kunci : Agglomerative Hierarchical *Clustering*, *Clustering*, *Data Mining*, *Euclidean Distance*, Musik Digital

4. Dalam jurnal oleh Wijaya Arim, dengan judul “*Analisis Algoritma K-means untuk sistem Pendukung Keputusan Penjuruan Siswa di MAN Binong Subang*” seperti abstrak dibawah ini.

MAN Binong, merupakan salah satu sekolah madrasah aliyah negeri di kabupaten Subang yang mengadakan penjurusan siswa kelas X (sepuluh) yang akan melanjutkan study ke kelas XI (sebelas). Kendala yang sering ditemukan dalam proses penjurusan di MAN Binong yaitu sulitnya menentukan siswa mana yang memenuhi kriteria untuk menempati jurusan tertentu. Hal ini dikarenakan proses penjurusan masih dilakukan dengan manual, sehingga membutuhkan waktu yang cukup lama dan dinilai kurang efisien. Pada penelitian ini akan diangkat suatu kasus yaitu menganalisis algoritma *K-Means* untuk sistem pendukung keputusan penjurusan siswa di MAN Binong Subang. Algoritma *K-Means* dapat mengelompokkan (segmentasi) data yang mempunyai atribut dan mempunyai jumlah data yang banyak, sehingga dapat dimanfaatkan dalam sistem penentuan penjurusan siswa yang sesuai dengan kemampuan akademik siswa. Berdasarkan hasil analisis terhadap algoritma *K-Means* untuk sistem pendukung keputusan penjurusan,

maka kesimpulan yang dapat diambil adalah algoritma *K-Means* kurang tepat untuk sistem pendukung keputusan penjurusan tetapi algoritma *K-Means* lebih tepat untuk mengelompokkan data siswa berdasarkan data nilai yang bisa memberikan gambaran untuk penjurusan siswa.

Kata Kunci : Analisis, Sistem Pendukung Keputusan, *K-Means*, Penjurusan, Nilai

5. Dalam jurnal SNATI, oleh Tedy Rismawan dan Sri Kusumadewi, pada tahun 2008, dengan judul “**Aplikasi *K-Means* untuk Mengelompokkan Mahasiswa berdasarkan Nilai Body Mass Index (BMI) dan Ukuran Kerangka**” seperti abstrak dibawah ini.

Masalah kesehatan merupakan permasalahan yang sangat penting untuk diperhatikan, diantaranya adalah masalah BMI dan ukuran kerangka seseorang. Apabila seseorang telah mengetahui nilai BMInya, orang tersebut dapat mengontrol berat badan sehingga dapat mencapai berat badan normal yang sesuai dengan tinggi badan. Sedangkan apabila orang tersebut mengetahui ukuran kerangka tubuhnya maka orang tersebut dapat mengontrol berat badannya agar dapat selalu berada dalam keadaan ideal. Pada penelitian ini, penulis mencoba membangun suatu sistem untuk mengelompokkan data yang ada berdasarkan status gizi dan ukuran rangkanya dengan memasukkan parameter kondisi fisik dari orang tersebut. Pengelompokkan data dilakukan dengan menggunakan metode *clustering K-Means* yaitu dengan mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat kelas. Dari hasil penelitian terhadap 20 data sampel diperoleh 3 kelompok mahasiswa berdasarkan nilai BMI dan ukuran rangka, yaitu : BMI normal dan kerangka besar, BMI obesitas sedang dan kerangka sedang, BMI obesitas berat dan kerangka kecil.

Kata Kunci : *Clustering*, BMI, Ukuran Rangka, *K-Means*.