

TUGAS AKHIR

PEMBANGUNAN MODUL PEMEGANG *CLUSTERING* UNTUK APLIKASI *DATA MINING*

**Diajukan sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Strata-1
Teknik Industri**



Oleh

Nama : Warda Arofah Makmun

No. Mahasiswa : 07 522 224

**JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2012

PERNYATAAN

Demi Allah, saya akui karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang setiap satunya telah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak intelektual maka saya bersedia ijazah yang telah saya terima untuk ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 15 Januari 2012

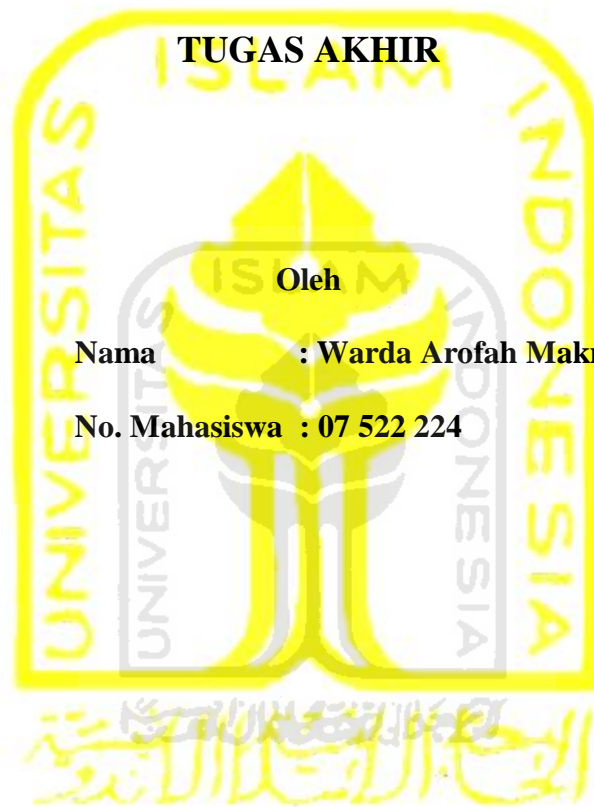


Warda Arofah Makmun

NIM 07522224

LEMBAR PENGESAHAN PEMBIMBING

**PEMBANGUNAN MODUL PEMEGANG *CLUSTERING* UNTUK
APLIKASI *DATA MINING***



Oleh
Nama : Warda Arofah Makmun
No. Mahasiswa : 07 522 224

Yogyakarta, Januari 2012

Pembimbing,

A handwritten signature in black ink, which appears to read 'Ridwan'.

Muhammad Ridwan Andi Purnomo, ST., M.Sc., Ph.D)

LEMBAR PENGESAHAN PENGUJI

PEMBANGUNAN MODUL PEMEGANG *CLUSTERING* UNTUK APLIKASI DATA MINING

TUGAS AKHIR

Oleh :

Nama : Warda Arofah Makmun

No. Mahasiswa : 07522224

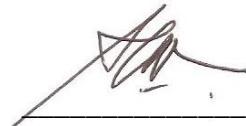
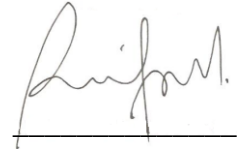
Telah Dipertahankan di Depan Sidang Penguji sebagai Salah Satu Syarat untuk
Memperoleh Gelar Sarjana Teknik Industri
Fakultas Teknologi Industri Universitas Islam Indonesia
Yogyakarta, Januari 2012

Tim Penguji

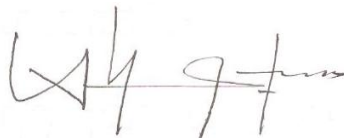
Muhammad Ridwan Andi Purnomo, ST., M.Sc., Ph.D
Ketua

Agus Mansur, ST., M.Eng, Sc
Anggota I

Sri Indrawati, ST., M.Eng
Anggota II



Mengetahui,
Ketua Jurusan Teknik Industri
Universitas Islam Indonesia



Drs. H. M. Ibnu Mastur, MSIE

HALAMAN PERSEMBAHAN

Bapak dan Ibu tercinta yang senantiasa tanpa lelah memberikan doa dan kasih sayang.

Kakak-kakak dan adikku tersayang yang selalu menemani dalam segala hal.

Bapak Muhammad Ridwan Andi Purnomo yang telah menjadi pembimbing skripsi.



MOTTO

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا ۗ

“Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya.”

(QS. Al-Baqarah 286)

وَإِذَا قِيلَ اذْشُرُوا فَاذْشُرُوا بِرَفْعِ اللَّهِ الَّذِينَ ءَامَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ ۗ وَاللَّهُ بِمَا

تَعْمَلُونَ خَبِيرٌ ۗ

“Allah pasti akan mengangkat orang – orang yang beriman dan berpengetahuan diantaramu beberapa tingkat lebih tinggi”

(QS. Al-Mujaadilah 11)

إِنَّ مَعَ الْعُسْرِ يُسْرًا ۗ فَإِذَا فَرَغْتَ فَانصَبْ ۗ

“Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain”

(QS. Al Insyirah 6 - 7)

KATA PENGANTAR

Assalamu 'alaykum Warahmatullah Wabarakatuh

Alhamdulillah, segala puji syukur hanya milik Allah Tuhan semesta alam penguasa hari pembalasan yang Maha Pengasih dan Penyayang, sholawat serta salam selalu tertuju pada Nabi Muhammad sebaik-baik ciptaan-Nya yang telah membawa kita ke jalan yang diridhai-Nya.

Dengan Rahmat dan Hidayah Allah SWT akhirnya tugas akhir yang berjudul “Pembangunan Modul Pemegang *Clustering* Untuk Aplikasi *Data mining*” dapat terselesaikan dengan baik.

Tujuan dari penyusunan tugas akhir ini adalah guna memperoleh gelar sarjana pada program studi Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.

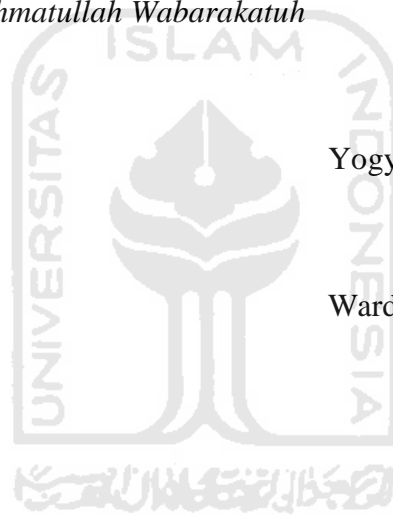
Dalam penyelesaian laporan tugas akhir ini tidak terlepas dari bantuan berbagai pihak. Untuk itu dalam kesempatan ini, dengan rasa hormat penulis ingin mengucapkan terimakasih kepada :

1. Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Kepala Prodi Teknik Industri dan seluruh staf, Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Muhammad Ridwan Andi Purnomo, ST., M.Sc., Ph.D selaku Dosen Pembimbing yang telah memberikan bantuan dan arahnya dalam penyusunan Tugas Akhir ini.
4. Kedua orang tuaku tercinta atas segala doa, bantuan, dan kasih sayang yang senantiasa mengalir untukku.

5. Kepala Laboratorium, Laboran dan seluruh Asisten Laboratorium *Data mining* yang selalu memberi dukungan dan membantu dalam jalannya penelitian.
6. Semua pihak yang tidak dapat disebutkan satu persatu.

Akhir kata penulis berharap semoga tugas akhir ini dapat memberikan manfaat dan berkah khususnya di dunia ilmu pengetahuan bagi semua pihak. Penulis menyadari bahwa dalam penyusunan tugas akhir ini masih banyak terdapat kekurangan sehingga dengan kerendahan hati penulis mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan pada masa mendatang.

Wassalamu'alaykum Warahmatullah Wabarakatuh



Yogyakarta, 15 Januari 2012

Warda Arofah Makmun

ABSTRAK

Clustering merupakan salah satu teknik data mining yang sering digunakan dalam dunia bisnis. Banyak sekali vendor-vendor yang membuat aplikasi data mining ataupun clustering namun untuk saat ini belum ada sebuah modul pemegang untuk clustering yang dibangun dalam sebuah spreadsheet. Penelitian ini mencoba membangun modul pemegang untuk clustering. Modul pemegang dibangun menggunakan bahasa pemrograman visual basic for application yang terdapat didalam aplikasi microsoft excel. Beberapa studi kasus telah diujikan untuk mengetahui performansi aplikasi yang dibuat. Hasil dari pengujian menunjukkan bahwa modul pemegang dapat berjalan dengan baik.

Kata kunci: Data mining, Spreadsheet, Clustering.



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN	ii
LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENGESAHAN PENGUJI	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
KATA PENGANTAR	vii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II KAJIAN LITERATUR	
2.1 Kajian Pustaka Deduktif	6
2.1.1 Data, Informasi, Pengetahuan	6
2.1.2 <i>Knowledge Discovery in Database</i>	7
2.1.3 <i>Data mining</i>	8
2.1.4 <i>Clustering</i>	11
2.1.5 <i>K-means dan Subtractive Clustering</i>	12
1 <i>K-means Clustering</i>	12
2 <i>Subtractive Clustering</i>	13
2.1.6 <i>Microsoft Excel dan Visual Basic for Application</i>	15
2.1.7 Perangkat Lunak <i>Data mining</i>	16
2.2 Kajian Pustaka Induktif	17

BAB III METODOLOGI PENELITIAN

3.1	Objek Penelitian	19
3.2	Diagram Alir Penelitian	20
3.3	Identifikasi Masalah	21
3.4	Kajian Literatur	21
3.5	Pembahasan Metode	21
3.6	Pembangunan <i>Handler Module</i>	22
3.7	Pengujian Sistem	22
3.8	Pembahasan Hasil	22
3.9	Kesimpulan dan Saran	22

BAB IV PENGUMPULAN DAN PENNGOLAHAN DATA

4.1	<i>Input</i> Sistem	23
4.2	<i>Handler Module</i>	24
4.2.1	<i>Handler Module</i> untuk algoritma <i>K-means</i>	24
4.2.2	<i>Handler Module</i> untuk algoritma <i>Subtractive</i>	25
4.3	<i>Output</i> Sistem	29
4.4	Pengujian Aplikasi	30
4.5	Pengujian <i>Handler Module K-means</i>	31
4.6	Pengujian <i>Handler Module Subtractive</i>	32

BAB V PEMBAHASAN

5.1	<i>Handler Module</i> untuk <i>K-means</i> dan <i>Subtractive Clustering</i>	34
5.2	Kelebihan dari Sistem yang Dibuat	36
5.3	Kekurangan dari Sistem yang Dibuat	37

BAB VI KESIMPULAN DAN SARAN

5.1	Kesimpulan	38
5.2	Saran	38

Daftar Pustaka

Lampiran

DAFTAR TABEL

Table 4.1	Tabel Data Uji Modul <i>K-means</i>	32
Table 4.2	Tabel Data Uji Modul <i>Subtractive</i>	33
Tabel 5.1	Tabel Nilai Rata-Rata	35



DAFTAR GAMBAR

Gambar 2.1	Tahapan-tahapan KDD	7
Gambar 3.1	Diagram Alir Penelitian	20
Gambar 4.1	Contoh Data <i>Input</i>	25
Gambar 4.2	Kode Bayangan Pembagian <i>Cluster</i>	26
Gambar 4.3	Kode Bayangan Proses Iterasi <i>K-means</i>	26
Gambar 4.4	Kode Bayangan Normalisasi	27
Gambar 4.5	Kode Bayangan Perhitungan Potensi Dan Pemilihan Potensi	27
Gambar 4.6	Kode Bayangan Proses Iterasi <i>Subtractive</i>	28
Gambar 4.7	Kode Bayangan <i>Denormalisasi</i>	29
Gambar 4.8	Kode Bayangan Pemilihan Anggota <i>Cluster</i>	29
Gambar 4.9	<i>Output</i> Sistem	30
Gambar 4.10	Alur Pengujian Aplikasi	31



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengaruh teknologi komputer terhadap organisasi dan masyarakat terus meningkat saat teknologi baru berkembang, dan teknologi saat ini makin luas. Makin banyak aspek aktivitas organisasional ditandai dengan interaksi antara mesin dan orang. Aplikasi komputer telah beralih dari aktivitas pemrosesan transaksi ke analisis masalah dan aplikasi solusi. Dengan berkembangnya teknologi komputasi akan meningkatkan kecepatan untuk mendapatkan informasi dan kecepatan dimana seorang pengambil keputusan membuat keputusan untuk merespons perubahan dalam lingkungan (Turban *et al.*, 2005).

Untuk memahami sebuah situasi dan perubahan dalam lingkungan, seorang pengambil keputusan membutuhkan data, informasi, dan pengetahuan. Hal-hal tersebut kemudian harus diintegrasikan dan diorganisasi dalam sebuah cara yang membuat mereka menjadi berguna. Pengambil keputusan kemudian harus mampu menerapkan peranti analisis sehingga data, informasi, dan pengetahuan dapat digunakan untuk mendapatkan manfaat (Turban *et al.*, 2005).

Metodologi tradisional untuk menganalisis data, tidak dapat menangani data dalam jumlah besar. Sementara para pelaku bisnis memiliki kebutuhan-kebutuhan untuk memanfaatkan gudang data yang sudah dimiliki, para peneliti melihat peluang untuk melahirkan sebuah teknologi baru yang menjawab kebutuhan ini, yaitu *data*

mining. Teknologi ini sekarang sudah ada dan diaplikasikan oleh perusahaan-perusahaan untuk memecahkan berbagai permasalahan bisnis (Moertini, 2002).

Data mining berarti mengekstrak atau menggali pengetahuan dari jumlah data yang sangat besar (Han & Kamber, 2000). Ada beberapa teknik dalam *data mining* diantaranya yaitu *Clustering*, Klasifikasi, Regresi, *Market Basket Analysis* (MBA). *Clustering* salah satunya adalah pembagian populasi yang heterogen ke dalam satu atau lebih kelompok atau *cluster* yang lebih homogen (Berry & Linoff, 2004). Tujuan dari metode *clustering* adalah untuk mengelompokkan sejumlah data atau objek ke dalam kluster sehingga setiap kluster akan terisi data yang semirip mungkin (Santosa, 2007).

Sebagaimana popularitas *data mining* meningkat, berbagai metode pun telah diberlakukan bagi bisnis terutama pemasaran. Analisis *Cluster* menjadi penting karena metode ini memungkinkan pengungkapan struktur dan asosiasi dalam data yang sebelumnya tidak diketahui, sehingga bermanfaat dan masuk akal ketika data tersebut ditemukan. Analisis *cluster* telah digunakan secara ekstensif untuk deteksi penipuan, penipuan kartu kredit dan *e-commerce*, dan segmentasi pasar pelanggan di dalam sistem CRM. Metode ini umum digunakan dalam bidang ilmu seperti ilmu biologi, kedokteran, genetika, ilmu sosial, ilmu antropologi, arkeologi, astronomi, pengenalan karakter, dan bahkan dalam pengembangan MIS (Turban *et al.*, 2005)

VBA adalah bahasa pemrograman yang dibuat oleh *microsoft* yang dapat membangun sebuah aplikasi. Sebagai contohnya, VBA untuk *excel* adalah bahasa pemrograman yang terkandung dalam *microsoft excel*. Dengan adanya bahasa pemrograman VBA dapat meningkatkan dan mengotomatisasikan aplikasi dalam *microsoft excel* (Darlington, 2004). Untuk membuat aplikasi dalam *microsoft excel*,

kode pemrograman dapat ditulis dalam sebuah *module* yang dapat dilihat, ditulis serta disunting melalui *Visual Basic Editor* (VBE).

Dengan memanfaatkan fasilitas *module* yang ada dalam bahasa pemrograman VBA, algoritma *data mining* khususnya *clustering* dapat dibangun didalamnya. Penelitian ini akan membahas bagaimana *module* tersebut dibangun. Dengan dibangunnya *module* ini, kasus yang diselesaikan dengan metode *clustering* dapat teratasi dengan cepat dan keputusan dapat diambil dengan cepat dan baik.

1.2 Rumusan Masalah

Dari uraian latar belakang diatas maka dapat dirumuskan pokok permasalahan dari penelitian yang akan dilakukan, bagaimana membangun sebuah *handler module* dalam lembar kerja *spreadsheets* seperti *microsoft excel* untuk memudahkan *user* dalam melakukan proses *clustering* dalam *data mining*.

1.3 Batasan Masalah

Agar penelitian ini lebih terarah, mudah dipahami dan tujuan penelitian dapat tercapai secara tepat dan baik, maka perlu dilakukan pembatasan lingkup penelitian. Adapun pembatasan lingkup penelitian ini adalah:

1. *Handler Module* yang akan dibangun dalam penelitian ini adalah salah satu dari algoritma dalam *data mining* yaitu algoritma *clustering*
2. Algoritma yang dipilih adalah algoritma *k-means* dan *subtractive clustering*, dan tidak membangun algoritma *clustering* yang lain.
3. Pembangunan *handler module* tidak sampai pada tahap transformasi dan profilisasi.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk membangun *Handler Module* untuk penyelesaian *data mining* khususnya untuk algoritma *k-means* dan *subtractive clustering*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk :

1. Membangun *Handler Module* untuk penyelesaian *clustering* (penggugusan) dalam *data mining*
2. Membantu pengguna dalam melakukan proses *data mining* khususnya dalam penyelesaian dengan menggunakan metode *k-means* dan *subtractive clustering*.
3. Menambah wawasan bagi pihak umum mengenai pembangunan *Handler Module* untuk algoritma *k-means* dan *subtractive clustering*.
4. Dapat digunakan sebagai bahan rujukan penelitian berikutnya.

1.6 Sistematika Penulisan

Laporan tugas akhir ini disusun secara sistematis ke dalam beberapa bab, dan masing-masing bab akan diuraikan sebagai berikut :

BAB II KAJIAN LITERATUR

Bab ini berisi tentang penjelasan teori penunjang yang digunakan sebagai landasan dalam menyelesaikan permasalahan yang ada dan

menjawab rumusan masalah, literatur-literatur yang berhubungan dengan *data mining* khususnya untuk metode *clustering*.

BAB III METODE PENELITIAN

Bab ini berisi tentang uraian metode pengumpulan data, diagram alur penelitian, dan cara pengolahan serta analisis data.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Pada bab ini berisikan pengumpulan data-data yang akan diolah sesuai dengan penelitian yang dilakukan, serta pengolahan data untuk menyelesaikan kasus yang di angkat.

BAB V PEMBAHASAN

Bab ini berisikan pembahasan terhadap data-data yang telah di olah menggunakan landasan berupa teori-teori penunjang penelitian yang di sajikan pada bab-2.

BAB VI PENUTUP

Bab ini berisikan kesimpulan dari pembahasan yang telah dilakukan dengan menjawab rumusan masalah yang telah diajukan serta saran – saran dan rekomendasi untuk penelitian selanjutnya.

DAFTAR PUSTAKA

LAMPIRAN

BAB II

KAJIAN PUSTAKA

2.1 Kajian Deduktif

Bab ini berisi tentang penjelasan teori penunjang yang digunakan sebagai landasan dalam menyelesaikan permasalahan yang ada dan menjawab rumusan masalah yang diajukan.

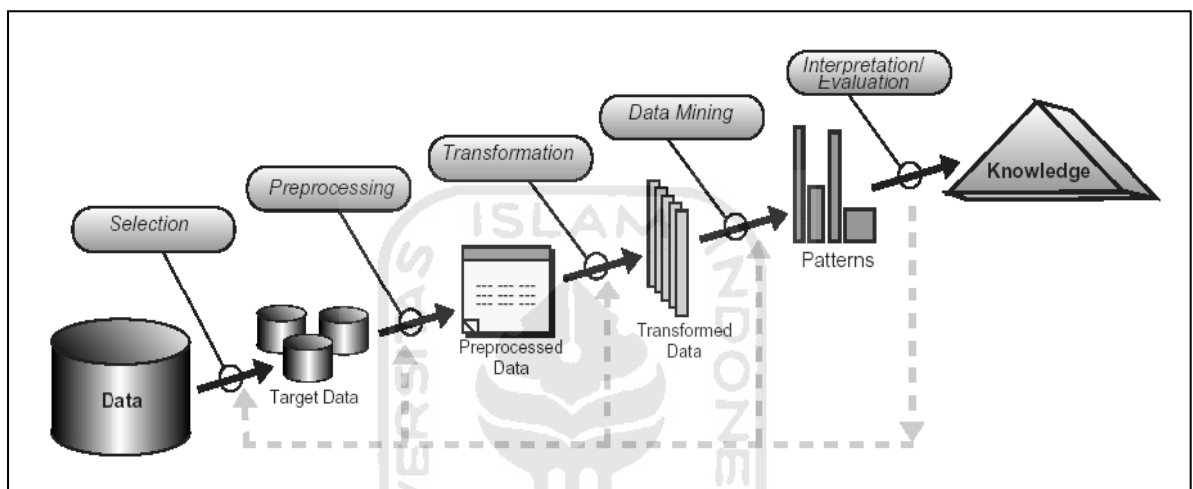
2.1.1. Data, Informasi, Pengetahuan

Dalam konteks teknologi informasi pengertian dari data, informasi, dan pengetahuan sangatlah berbeda satu sama lain. Pengertian data adalah item – item mengenai sesuatu, kejadian, aktivitas dan transaksi yang direkam, diklasifikasikan dan disimpan namun tidak diorganisasi untuk menyampaikan semua makna spesifik. Item data dapat berupa *numeric*, *alphanumeric*, gambar, suara, atau *image* (Turban *et al.*, 2005).

Pengertian informasi adalah data yang telah diorganisasi dalam sebuah cara yang membuat mereka bermakna bagi penerima (Turban *et al.*, 2005). Sebuah aplikasi sistem pendukung manajemen memproses item-item data sehingga hasilnya berguna bagi suatu tindakan atau keputusan yang diinginkan. Sedangkan pengertian pengetahuan adalah informasi yang kontekstual, relevan dan dapat dilakukan (Turban *et al.*, 2005).

2.1.2. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) adalah keseluruhan proses untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat baru, sah dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola – pola sejumlah kumpulan data (Zaine, 1999).



Gambar 2.1 Tahapan – Tahapan KDD

Adapun langkah-langkah utama KDD adalah sebagai berikut (Fayyad et al., 1996):

- a. Seleksi. Dalam tahap ini dilakukan pemahaman terhadap masalah yang akan dicari solusinya melalui penemuan pengetahuan, pengetahuan sebelumnya yang relevan, penentuan tujuan dan pemilihan data yang akan dianalisis.
- b. Pra-pengolahan. Dalam tahap ini dilakukan proses pra-pengolahan data, termasuk pembersihan data.
- c. Transformasi. Dalam tahap ini dilakukan pengurangan dan/atau perubahan tipe data, sehingga data siap untuk dipresentasikan ke teknik-teknik *data mining*.
- d. *Data mining*. Dalam tahap ini dilakukan pemilihan tugas *data mining* yang sesuai dengan tujuan dan algoritma *data mining* untuk pencarian pola.

- e. Interpretasi/evaluasi. Dalam tahap ini dilakukan visualisasi dan interpretasi terhadap pola yang ditemukan untuk dijadikan pengetahuan.

2.1.3. *Data mining*

Ada banyak definisi dari *Data mining*, menurut Gartner Group, *data mining* merupakan proses dari pola-pola rancang bangun matematika dari sekumpulan besar data.

Selain itu ada beberapa definisi lain dari *data mining*, yaitu :

- a. *Data mining* adalah proses dalam menemukan pengetahuan menarik dari sejumlah data besar yang tersimpan dalam tiap *database*, gudang data, atau tempat penyimpanan informasi lain (Larose, 2005).
- b. *Data mining* adalah salah satu tahapan dari proses KDD yang menyajikan analisis data dan algoritma untuk menemukan pola dari fakta satu persatu atau model dari suatu data (Fayyad, 1996).
- c. *Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database*. *Data mining* adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban *et al.*,2005)

Data mining meliputi tugas-tugas yang dikenal sebagai ekstraksi pengetahuan, arkeologi data, eksplorasi data dan memanen informasi. Berikut adalah beberapa karakteristik utama dan sasaran *data mining* (Turban *et al.*,2005) :

1. Data sering dikubur pada sebuah database yang sangat besar, yang kadang berisi data-data dari beberapa tahun. Sebagian ada yang dihapus dan sebagian hanya disimpan dalam *data warehouse*.
2. Peranti *data mining* sudah digabung dengan *spreadsheet* dan peranti pengembangan perangkat lunak lainnya. Jadi data yang sudah di-*mining* dapat dianalisis dan diproses dengan cepat dan mudah.

Ada beberapa teknik-teknik dalam *data mining*, Larose (2005) menyebutkan beberapa teknik *data mining* yang umum digunakan, yang akan dijelaskan pada sub bab berikut ini:

Pertama adalah deskripsi. Model *data mining* harus setransparan mungkin. Artinya hasil dari model *data mining* harus dapat menjelaskan pola yang jelas untuk untuk interpretasi dan penjelasan yang intuitif. Beberapa metode memberikan interpretasi yang transparan daripada model *data mining* yang lain. Sebagai contoh pohon keputusan memberikan hasil yang mudah dan sangat *humanfriendly*. Di sisi lain, jaringan saraf yang buram dan *nonspecialist*, karena *nonlinear* dan kompleksitas model.

Kedua adalah estimasi. Estimasi mirip sekali dengan klasifikasi namun data yang diolah berupa data numerikal bukan kategorikal. Model yang dibangun berdasarkan *record* yang lengkap, yang memberikan nilai dari variabel sasaran serta nilai prediktor. Kemudian untuk observasi yang baru, perkiraan dari nilai variabel target yang dibuat didasarkan pada nilai-nilai prediktor.

Selanjutnya adalah prediksi. Prediksi mirip sekali dengan klasifikasi dan estimasi. Namun untuk prediksi, hasilnya terletak pada masa depan. Setiap metode dan teknik yang digunakan untuk klasifikasi dan estimasi dapat juga digunakan untuk

prediksi. Termasuk model statistik sederhana yaitu regresi linear sederhana dan korelasi dan regresi berganda.

Keempat adalah klasifikasi. Dalam klasifikasi nilai data yang digunakan berupa data kategoris. Seperti golongan pendapatan misalnya dibagi menjadi tiga kelas atau kategori : pendapatan tinggi, pendapatan menengah dan pendapatan rendah. Kumpulan data untuk pengolahan klasifikasi disebut *training set*. Kemudian algoritma akan melihat *record* baru yang tidak ada informasi tentang pendapatan golongan. Berdasarkan, klasifikasi dalam *training set*, algoritma dapat menetapkan klasifikasi pada *record* baru tersebut.

Kelima adalah *association rule*. Tugas *association rule* adalah untuk menemukan hubungan variabel pada peristiwa yang terjadi bersama-sama dalam satu waktu. Istilah paling lazim dalam dunia bisnis adalah analisis *afinitas* atau *market basket analisis*. *Association rule* berusaha untuk menemukan aturan untuk mengukur hubungan antara dua atau lebih atribut. Ukuran dari aturan asosiasi terbentuk berdasarkan nilai *support* dan *confidence*.

Selanjutnya adalah *clustering*. *Clustering* mengacu pada pengelompokan *record*, observasi, atau kasus kedalam kelas objek yang sama. Satu *cluster* adalah kumpulan *record* yang sama satu sama lain dan berbeda dengan *record* dalam *cluster* yang lain. *Clustering* berbeda dengan klasifikasi, tidak ada variabel target dalam *clustering*. Tugas pengelompokan dalam *clustering* tidak bertugas untuk mengklasifikasi dan estimasi atau memprediksi nilai dari sebuah variabel target. Sebaliknya algoritma *clustering* hanya berupaya untuk membagi seluruh data set ke dalam sub kelompok atau *cluster* yang lebih homogen, dimana kesamaan dari *record* dalam satu *cluster* dimaksimalkan dan kesamaan diluar *cluster* diminimalkan.

2.1.4. *Clustering*

Clustering, salah satu metode *data mining* yang sering digunakan dalam dunia bisnis. *Clustering* berarti mengelompokkan objek, observasi, atau kasus ke dalam suatu kelas yang memiliki objek yang sama. Sebuah *cluster* adalah kumpulan dari objek yang sama satu sama lain dan berbeda dengan objek pada *cluster* yang lain (Larose, 2005).

Analisis *cluster* merupakan metode yang sangat penting untuk menggolongkan berbagai item ke dalam pengelompokan umum yang disebut *cluster*. Metode ini umum didalam biologi, kedokteran, genetika, ilmu social, ilmu antropologi, arkeologi, astronomi, pengenalan karakter, dan bahkan dalam pengembangan MIS. Sebagaimana popularitas *data mining*, berbagai metode pun telah diberlakukan bagi bisnis, terutama pemasaran. Analisis *cluster* telah digunakan secara ekstensif untuk deteksi penipuan, penipuan kartu kredit dan *e-commerce*, dan segmentasi pasar pelanggan di dalam sistem CRM. Lebih banyak aplikasi dalam bisnis terus dikembangkan ketika kekuatan analisis *cluster* dipahami dan digunakan (Turban *et al.*,2005) Berbeda dengan klasifikasi, *clustering* tidak memerlukan kelas yang telah didefinisikan sebelumnya atau kelas hasil training, dengan demikian *clustering* dinyatakan sebagai bentuk pembelajaran berdasarkan observasi dan bukan berdasarkan contoh (Han & Kamber, 2001).

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.

2.1.5. *K - means dan Subtractive Clustering*

1. *K-means Clustering*

K-means merupakan salah satu algoritma *clustering* yang paling simpel dan paling populer. Dalam teknik ini kita ingin mengelompokkan objek kedalam k kelompok atau kluster. Untuk melakukan klastering ini, nilai k harus ditentukan terlebih dahulu.

Algoritma *k-means* klastering bisa diringkas sebagai berikut (Santosa, 2007) :

- a. Pilih jumlah kluster k
- b. Inisialisasi k pusat kluster.
- c. Tempatkan setiap data/objek ke kluster terdekat. Kedekatan kedua obyek ditentukan berdasar jarak kedua obyek tersebut. Demikian juga kedekatan suatu data dengan pusat kluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat kluster. Jarak paling dekat satu data dengan satu kluster tertentu akan menentukan suatu data masuk dalam kluster mana.
- d. Hitung kembali pusat kluster dengan keanggotaan kluster yang sekarang. Pusat kluster adalah rata – rata dari semua obyek dalam kluster tertentu. Jika dikehendaki bisa juga memakai median dari kluster tersebut. Jadi rata – rata bukan satu – satunya ukuran yang bisa dipakai.
- e. Tugaskan lagi setiap obyek dengan memakai pusat kluster yang baru. Ulangi dari langkah ke- 3 sampai pusat kluster tidak berubah lagi.

2. *Subtractive Clustering*

Algoritma *Subtractive Clustering* (chiu, 1994) menggunakan data poin sebagai kandidat (calon) dari pusat *cluster*. Matriks data dengan n data poin $\{x_1, \dots, x_n\}$ didalam m dimensi yang didalamnya terdapat output dan dinormalisasi didalam *hypercube*. Metode *Subtractive Clustering* merupakan pengembangan dari metode *Mountain Clustering*. Pada metode *Mountain Clustering* bergantung pada pemecahan jaringan dan dimensi datanya, dengan cara demikian metode ini tidak akan terkomputasi dengan baik ketika diaplikasikan untuk dataset yang berdimensi tinggi dengan pemecahan jaringan yang bertambah/meningkat. Dengan menggunakan Algoritma *Subtractive Clustering*, fungsi dari metode *Mountain Clustering* akan dihitung berdasarkan data point bukan dalam titik jaringan (Yang *et al*, 2009).

Algoritma *Subtractive Clustering* dapat digambarkan dalam langkah-langkah sebagai berikut (Yang *et al*, 2009) :

Step 1 : hitung potensi P_i untuk setiap point, $1 \leq i \leq N$

untuk mengetahui titik mana yang berpotensi dapat dihitung dengan rumus :

$$P_i = \sum \exp(-\alpha \|x_i - x_l\|^2) \dots\dots\dots 1$$

Step 2 : set nomor untuk node tersembunyi $l = 1$ dan pilih data point dengan nilai potensi yang paling tinggi sebagai pusat node tersembunyi pertama. Perkiraan X_l menjadi lokasi titik dan P_l menjadi nilai potensi yang cocok.

Step 3 : hitung kembali nilai potensi untuk setiap data poin dengan menggunakan rumus :

$$P_i = P_i - P_l \exp(-\beta \|x_i - x_l\|^2) \dots\dots\dots 2$$

dimana $\beta = 4/r_b^2$ dan $r_b > 0$ mewakili radius dari lingkungan untuk potensi yang signifikan ketika perhitungan nilai potensi kembali terjadi.

Step 4 : jika $\max P_i \leq \varepsilon P_l$, akhiri perhitungan algoritma ini; sebaliknya, set $L=L+1$ dan temukan data poin dengan nilai potensi yang paling besar. Perkirakan X_L sebagai lokasi dari titik P_L menjadi nilai potensi yang cocok.

Step 5 : pilih X_L sebagai pusat node tersembunyi yang baru. Hitung kembali nilai potensi untuk setiap data poin dengan menggunakan rumus :

$$P_i = P_i - P_L \exp(-\beta \|x_i - x_L\|^2) \dots \dots \dots 3$$

Squash factor digunakan untuk mengalikan nilai radius r_a , untuk menentukan pusat *cluster* yang ada didekatnya. Nilai *default* untuk *squash factor* adalah 1.25. untuk menentukan pusat *cluster* baru dapat diterima menjadi pusat cluster atau tidak dapat digunakan dua pecahan pembanding yang disebut *accept ratio* dan *reject ratio*. Apabila nilai hasil bagi antara potensi tertinggi suatu titik data dengan nilai potensi awal lebih besar dengan nilai *accept ratio* maka titik data tersebut dapat diterima sebagai pusat *cluster* baru, apabila nilai hasil bagi antara potensi tertinggi suatu titik data dengan nilai potensi awal diantara *accept ratio* dan *reject ratio* maka titik data tersebut akan dipertimbangkan untuk menjadi pusat *cluster* baru. Dan apabila nilai hasil bagi antara potensi tertinggi suatu titik data dengan nilai potensi awal lebih kecil dari nilai *reject ratio* maka titik data tersebut tidak diterima sebagai pusat *cluster* dan tidak akan dipertimbangkan kembali untuk menjadi pusat *cluster*.

2.1.6. *Microsoft Excel dan Visual Basic for Application*

Ide untuk aplikasi perangkat lunak berbasis *spreadsheet* pertama kali dibangun didalam program yang disebut VisiCalc dan tersedia secara komersial di tahun 1970an. Aplikasi ini dibangun oleh Dan Bricklin dan Bob Frankston dan ditulis pada *macro Apple II* (Darlington, 2004)

Kemudian, sejak tahun 1980 hingga saat ini, *Microsoft* melanjutkan komitmen mereka untuk mengembangkan *Basic* mulai dalam sistem operasi DOS hingga sistem operasi *Windows*. Kini *Visual Basic* merupakan bahasa pemrograman visual yang sangat sukses. *Visual Basic Excel* telah mengalami beberapa revisi sejak tahun 1990. Sejalan dengan perubahan sistem operasi berbasis *windows*, tetapi *excel 5* yang dirilis tahun 1994 telah mengandung versi pertama VBA. Sejak *microsoft office 97* dirilis, VBA telah dibangun dalam 4 dari 5 aplikasi *microsoft office* diantaranya : *Word, Access, Excel, dan Power Point*. VBA adalah bahasa pemrograman yang dibuat oleh *microsoft* yang dapat membangun sebuah aplikasi. Sebagai contohnya, VBA untuk *excel* adalah bahasa pemrograman yang terkandung dalam *microsoft excel*. Dengan adanya bahasa pemrograman VBA dapat meningkatkan dan mengotomatisasikan aplikasi dalam *microsoft excel* (Darlington, 2004).

Dengan VBA kita bisa melakukan banyak hal seperti melakukan peramalan dan penganggaran, menganalisis data ilmiah, menciptakan faktur dan formulir lainnya, mengembangkan grafik dari data, dan lain-lain (Walkenbach, 2007). Untuk membuat aplikasi dalam *microsoft excel*, kode pemrograman dapat ditulis dalam sebuah module. Module VBA bersama-sama dengan *worksheet* dan *object* lain disimpan di dalam *Workbook Excel* dan dapat dilihat, ditulis serta disunting melalui *Visual Basic Editor* (VBE).

2.1.7. Perangkat Lunak Clustering

Selain perangkat lunak untuk *data mining* dimana analisis *cluster* diletakkan, ada beberapa paket khusus untuk analisis cluster (Turban *et al.*,2005), meliputi :

- a. ClustanGraphies (Clustan)
- b. DecisionWORKS Suite (Advanced Software Application)
- c. SPSS (SPSS)
- d. PolyAnalyst Cluster Engine (Megaputer)
- e. Sokal Code (Hand)



2.2. Kajian Pustaka Induktif

Penelitian terdahulu yang meneliti tentang pembuatan aplikasi *k-means* ataupun *subtractive clustering* adalah Sutiknyo (2008) yang melakukan penelitian dengan menggunakan algoritma *k-means* untuk penggolongan suara. Data yang digunakan adalah berupa nilai formant yang didapat dari nilai magnitude tertinggi atau puncak dari suatu frame antara rentang frekuensi tertentu yang dapat mewakili suara pengguna. Sistem dapat mendeteksi suara baru yang belum diketahui. Suara tidak diketahui (*unknown voiced*) dibandingkan dengan 2 buah *centroid* yang mewakili suara dewasa dan anak-anak dengan menghitung jarak minimum (*Euclidean Distance*) sehingga oleh sistem dapat diputuskan suara tidak diketahui tersebut sebagai suara dewasa atau anak-anak.

Penelitian yang lain dilakukan oleh Dantes (2008). Penelitian yang dilakukan adalah pembangunan sistem informasi klasifikasi tingkat prestasi akademik mahasiswa. Variabel yang dimasukkan adalah id mahasiswa, jenis tes masuk universitas, nilai tes masuk dan IPK. Dari sistem yang dibuat bisa langsung terlihat presentase pengaruh jenis tes masuk universitas terhadap nilai IPK, dan dapat terlihat jumlah data yang masuk dalam masing-masing *cluster*.

Penelitian selanjutnya dilakukan oleh Kusri (2007). Penelitian yang dilakukan adalah mengimplementasikan algoritma *k-means* untuk mengubah kelas kontinyu menjadi kelas *diskrit*. Dengan menggunakan *script* PHP dan *software* pengelola *database* My SQL, tabel kelas kontinyu berhasil dirubah menjadi tabel *diskrit* dengan menggunakan algoritma *k-means*.

Priyono *et al.*, (2005) melakukan penelitian dengan menggunakan metode *subtractive clustering* untuk pembangkitan aturan *fuzzy*. Dalam proses *subtractive*

clustering, nilai r atau jari-jari yang digunakan adalah sebesar 0.3 dan 0.5, dengan menggunakan nilai r sebesar 0.3, rule yang terbentuk adalah 8 rule tetapi dengan menggunakan nilai r sebesar 0.5, aturan yang diperoleh adalah 4 rule.

Doan et al., (2005) melakukan penelitian dengan menggabungkan dua teknik yaitu *subtractive clustering* dan Algoritma Genetik Mikro. Penelitian ini mengusulkan suatu skema yang mengekstraksi data yang paling representatif dari sekumpulan data mentah. Data yang direduksi adalah data sungai Mississippi di vicksberg (1975-1993) dan sungai wabasa di Mt. Carmel (1960-1978). Data yang direduksi selanjutnya digunakan untuk proses peramalan arus sungai, dengan pereduksian data berguna untuk mempersingkat waktu peramalan. Ditunjukkan dalam penelitian bahwa data yang direduksi hampir 40% dari data mentah, namun hasil akurasi dari peramalan data arus sungai sama bagus ketika data belum direduksi.

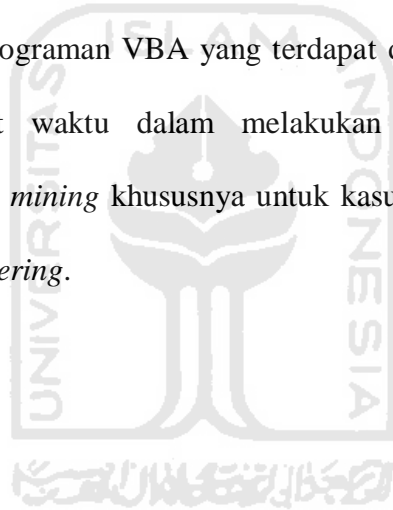
Yang et al., (2009) melakukan penelitian untuk mendeteksi *outlier* dengan algoritma *subtractive clustering* berdasarkan RBF *neural network*. Tujuan pendeteksian *outlier* adalah untuk mengarah pada jaringan saraf tiruan yang tepat dan mencapai komputasi yang efisien. Penelitian menunjukkan bahwa algoritma *subtractive clustering* berdasarkan RBF sangat efektif dan efisien untuk pendeteksian *outlier* dengan tingkat kesalahan rendah dan tingkat pendeteksian *outlier* yang tinggi.

BAB III

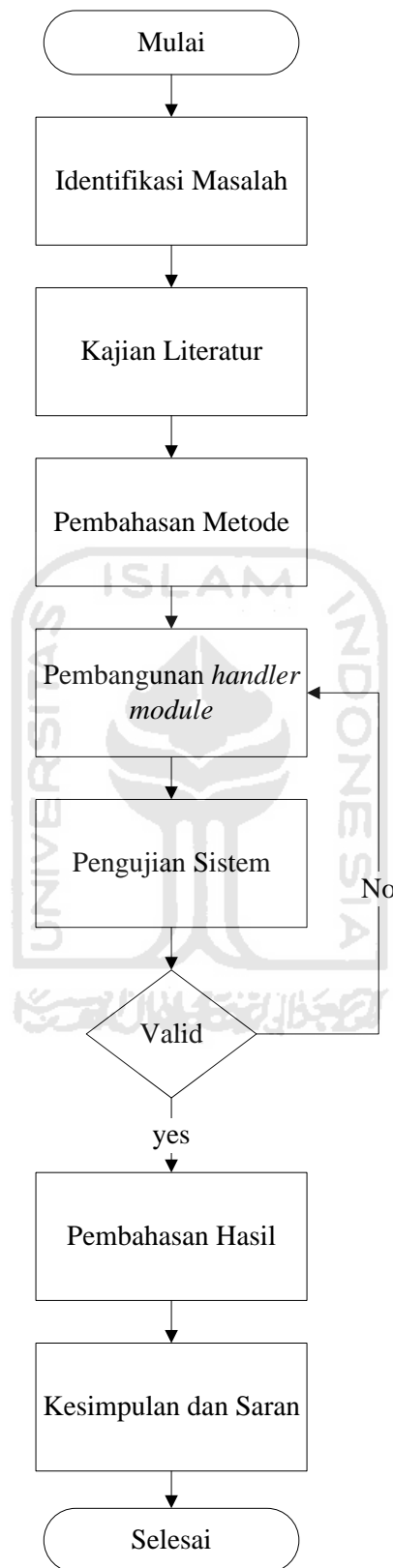
METODOLOGI PENELITIAN

3.1. Objek Penelitian

Penelitian ini difokuskan kepada pembangunan *handler module* yang dibangun menggunakan bahasa pemrograman VBA yang terdapat dalam *microsoft excel* untuk membantu mempersingkat waktu dalam melakukan pengolahan data dengan menggunakan metode *data mining* khususnya untuk kasus yang diselesaikan dengan menggunakan metode *clustering*.



3.2. Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian

3.3. Identifikasi Masalah

Tahap pertama dalam penelitian ini adalah mengidentifikasi masalah. Banyak sekali perangkat lunak untuk *data mining* yang dibangun, namun belum ditemukan modul pemegang untuk *clustering* yang dibangun didalam basisdata seperti *spreadshhet*. Dengan memanfaatkan bahasa pemrograman *visual basic for application* yang ada didalam *microsoft excel*, maka modul untuk *clustering* akan dibangun didalamnya.

3.4. Kajian Literatur

Literatur-literatur yang ada berisi mengenai teori yang berhubungan dengan permasalahan pada penelitian ini dan pada bahasan ini juga mengkaji penelitian-penelitian terdahulu yang sudah terlebih dahulu melakukan penelitian sebagai bahan acuan. Teori yang digunakan terdiri dari bahasa pemrograman VBA dalam *microsoft excel* dan juga literatur-literatur yang berhubungan dengan *data mining* khususnya untuk metode *clustering k-means* dan *subtractive clustering*.

3.5. Pembahasan Metode

Di tahap ini dilakukan perhitungan analisis *data mining* khususnya untuk algoritma *clustering* yang akan dibangun secara manual. Tahap ini dimaksudkan supaya langkah – langkah dalam algoritma yang dibangun lebih jelas sehingga membantu mempermudah dalam pembangunan sistem.

3.6. Pembangunan *Handler Module*

Dalam tahap ini *handler module* akan dibangun dalam *microsoft excel* dengan menggunakan bahasa pemrograman VBA yang sudah tersedia dalam *microsoft excel*. Dimana *handler module* yang dibangun adalah salah satu metode dalam *data mining* yaitu *clustering* dan dikhususkan untuk algoritma *k-means* dan algoritma *subtractive clustering*.

3.7. Pengujian Sistem

Dalam tahap ini, ada beberapa kumpulan data yang akan dimasukkan kedalam sistem yang dibangun guna mengetahui *handler module* yang dibangun berhasil atau tidak. Data yang digunakan berasal dari studi kasus sebelumnya yang sudah diselesaikan dengan menggunakan metode *clustering* ataupun dengan metode lainnya.

3.8. Pembahasan Hasil

Setelah tahap validasi akan dilakukan pembahasan terhadap *handler module* yang dibangun serta pembahasan tentang kelebihan dan kelemahan sistem yang dibangun.

3.9. Kesimpulan dan Saran

Tahap ini akan berisi tentang penjelasan singkat dari hasil akhir penelitian. Dalam hal ini hasil dari pembangunan *handler module* dan diberikan rekomendasi-rekomendasi untuk penelitian selanjutnya.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1. Pembangunan *Handler Module*

Modul yang akan dibangun adalah modul untuk algoritma *clustering* dalam *data mining*. Modul dibangun dengan menuliskan kode bahasa menggunakan bahasa pemrograman VBA yang ditulis di dalam VBE.

4.1.1. Input Sistem

Dalam suatu sistem terdapat input, proses dan output. Input dari sistem yang dibuat adalah data beserta atribut yang dimasukkan oleh *user* ke dalam *microsoft excel* yang berisi *handler module* yang dibuat. Data yang akan diproses telah mengalami tahap transformasi ke dalam bentuk numerikal. Berikut adalah contoh dari format data yang dimasukkan dalam *microsoft excel* :

	A	B	C	D	E	F	G
1	No	X1	X2	X3	X4		
2	1	15000000	25000000	42	5000000		
3	2	20000000	26420000	72	5230000		
4	3	17820000	22052000	35	5200000		
5	4	16205000	18500000	12	4250000		
6	5	8000000	15200000	5	3500000		
7	6	14260000	19640000	15	4023000		
8	7	7025000	15230000	19	5000000		
9	8	25032000	34000000	28	8000000		
10	9	24320100	35100000	39	12500000		
11	10	25602000	38200000	43	13250000		
12	11	19872000	28000000	27	10500000		
13	12	19000000	25000200	41	6350000		
14	13	16540200	30000200	29	7525000		
15	14	28920000	41000000	58	15620000		
16	15	15870200	26750000	19	4025000		
17	16	26840320	39000200	47	13025000		
18	17	24601200	38450000	64	11000250		
19	18	21650000	37525000	60	9850000		
20	19	18602000	30500000	74	11230000		
21	20	35024000	52000000	73	18230000		
22	21	39024300	52050000	26	15725000		
23	22	27500000	36500000	6	10560000		
24	23	32500500	45600000	10	16583000		
25	24	27963000	40250000	38	13670000		
26	25	37250020	51000000	68	18530000		
27	26	16523000	26750000	9	8500000		
28	27	25690000	39565000	48	15250000		
29	28	33450000	51065000	37	21500000		
30	29	9850000	1350000	13	2000000		
31	30	16950000	24580000	18	4500000		

Gambar 4.1 Contoh Data Masukan

Data yang akan diproses harus bertipe numerikal, apabila data yang akan dimasukkan bertipe kategorikal maka *user* berperan dalam proses transformasi ke dalam bentuk numerikal.

4.2. Handler module

Handler module dibangun dalam *microsoft excel* dan dibangun dengan menggunakan bahasa pemrograman VBA yang terdapat didalamnya. Ada dua algoritma *data mining* yang dibangun dalam penelitian ini yaitu algoritma *k-means* dan *subtractive clustering*.

4.2.1. Handler module untuk algoritma k-means

Modul yang dibangun untuk *k-means clustering* terdiri dari satu *prosedur function* dengan nama *k-means*. *Prosedur function* inilah yang akan menjalankan langkah-langkah algoritma dalam *k-means clustering*.

Dalam modul *k-means*, data yang telah masuk akan dibagi secara rata dan tanpa sisa sesuai dengan jumlah *cluster* yang dimasukkan oleh *user*. Kode bayangan untuk pembagian *cluster* tersebut adalah sebagai berikut :

```
Memulai proses pembagian cluster :
Membagi data kedalam jumlah cluster secara rata sesuai
dengan k (jumlah cluster)
```

Gambar 4.2 Kode Bayangan Pembagian *Cluster*

Setelah data dibagi rata sesuai dengan jumlah *cluster*, maka data yang telah dimasukkan oleh *user* selanjutnya akan diproses dengan algoritma *k-means*. Disini proses iterasi untuk algoritma *k-means* akan berlangsung. Rata-rata dari tiap pusat *cluster* akan dihitung dan jarak antar data terhadap pusat *cluster* juga akan dihitung, maka akan didapat keanggotaan data terhadap pusat *cluster*. Iterasi akan berhenti ketika anggota *cluster* sebelumnya dan sesudahnya adalah sama.

```
Memulai proses iterasi k-means :
Menghitung rata-rata tiap cluster.
Menghitung jarak antar data dengan menggunakan ukuran euclidean
distance.
Mencari nilai minimal kecenderungan data terhadap pusat cluster.
```

Gambar 4.3 Kode Bayangan Proses Iterasi *K-Means*

4.2.2. *Handler module* untuk algoritma *subtractive clustering*

Modul yang dibangun untuk algoritma *subtractive clustering* juga terdiri dari satu *prosedur function* dengan nama *subtractive*. *Prosedur function* inilah yang akan menjalankan langkah-langkah algoritma dalam *subtractive clustering*.

Dalam algoritma *subtractive clustering*, tahap pertama yang dilakukan adalah normalisasi. Data yang telah dimasukkan akan dinormalisasi agar diperoleh data dalam interval [0, 1]. Kode bayangan yang menunjukkan proses normalisasi adalah sebagai berikut :

```

Memulai proses normalisasi :
Menghitung nilai maximal dan nilai minimal untuk setiap
data dalam setiap variabel
Melakukan perhitungan normalisasi  $X = (X_i - X_{min}) / (X_{max} - X_{min})$ 

```

Gambar 4.4 Kode Bayangan Normalisasi

Setelah data ternormalisasi, tahap selanjutnya adalah perhitungan potensi awal untuk setiap data dan perhitungan potensi tertinggi. Kode bayangan yang menunjukkan proses ini adalah sebagai berikut :

```

Memulai proses perhitungan densitas :
Menghitung nilai potensi untuk tiap-tiap titik data dengan
menghitung jarak tiap data dengan semua data.
 $D_i = \text{jumlah dari bilangan } e^{(-4 * (X_i - X_{ij}))}$ 
Lakukan untuk semua titik data.
Mencari nilai potensi tertinggi dari nilai potensi titik data.

```

Gambar 4.5 Kode Bayangan Perhitungan Potensi Awal Dan Pemilihan Potensi Tertinggi

Dalam tahap berikutnya merupakan rangkaian tahap iterasi yang terdiri dari tahap pengujian terhadap semua titik dan pengurangan potensi pada setiap titik yang ada, pemilihan pusat *cluster* berikutnya, perhitungan rasio serta perbandingan dengan *accept ratio* dan *reject ratio*. Tahap iterasi dapat dilihat pada kode bayangan dibawah ini :

Memulai proses iterasi subtractive :

Menghitung nilai rasio dengan membagi nilai potensi tertinggi dengan nilai potensi tertinggi awal.

memperbandingkan nilai rasio dengan nilai accept ratio dan reject ratio yan telah ditetapkan.

Jika nilai rasio > accept rasio, maka data diterima sebagai pusat cluster.

Jika nilai rasio < reject rasio, maka data tidak dipertimbangkan lagi menjadi pusat cluster, dan iterasi dihentikan.

Jika nilai reject ratio < rasio < accept ratio, maka data akan dipertimbangkan dengan menghitung nilai Smd. Pertama mencari nilai G_{ij} dengan mengurangkan pusat cluster yang terpilih dengan pusat cluster sebelumnya dan dibagi dengan r atau jari-jari. Kemudian mencari nilai Sd dengan menjumlahkan nilai G_{ij} yang telah dikuadratkan.

$Md = Sd_{awal}$, jika $Sd_{selanjutnya} > Md$, maka Md tidak berubah, jika $Sd_{selanjutnya} < Md$, maka $Md = Sd_{selanjutnya}$.

$Smd = \sqrt{Md}$. Rasio yang ada dijumlahkan dengan Smd. Jika hasilnya ≥ 1 maka data diterima menjadi pusat cluster, namun jika hasilnya < 1 , data tidak akan dipertimbangkan kembali menjadi pusat cluster dan data diset menjadi 0.

Apabila titik sudah diset 0, maka mencari titik data tertinggi kedua setelah data tersebut. Lalu memperbandingkan kembali dengan nilai rasio.

Gambar 4.6 Kode Bayangan Proses Iterasi

Setelah itersi berhenti dan didapat beberapa pusat *cluster*, tahap berikutnya adalah denormalisasi. Pada tahap ini nilai-nilai pada data dikembalikan pada nilai awal. Tahap denormalisasi ini dapat dilihat pada kode bayangan dibawah ini :


```
Memulai perhitungan denormalisasi :  
Mengembalikan titik-titik data kedalam nilai awal.  
Denormalisai = nilai pusat data * (nilai maximum data -  
nilai minimum data) + nilai minimum data
```

Gambar 4.7 Kode Bayangan Denormalisasi

Tahap terakhir dari proses *subtractive clustering* adalah penentuan anggota *cluster*, tahap ini menggunakan fungsi *gaussian*, tahap ini dapat dilihat pada kode bayangan dibawah ini :

```
Memulai perhitungan fungsi gaussian :  
Mencari nilai  $\mu$  untuk setiap data terhadap pusat  
cluster yang ada.  
Kecenderungan data terhadap pusat cluster ditentukan  
dengan nilai  $\mu$  terkecil
```

Gambar 4.8 Kode Bayangan Pemilihan Anggota *Cluster*

4.3. Output Sistem

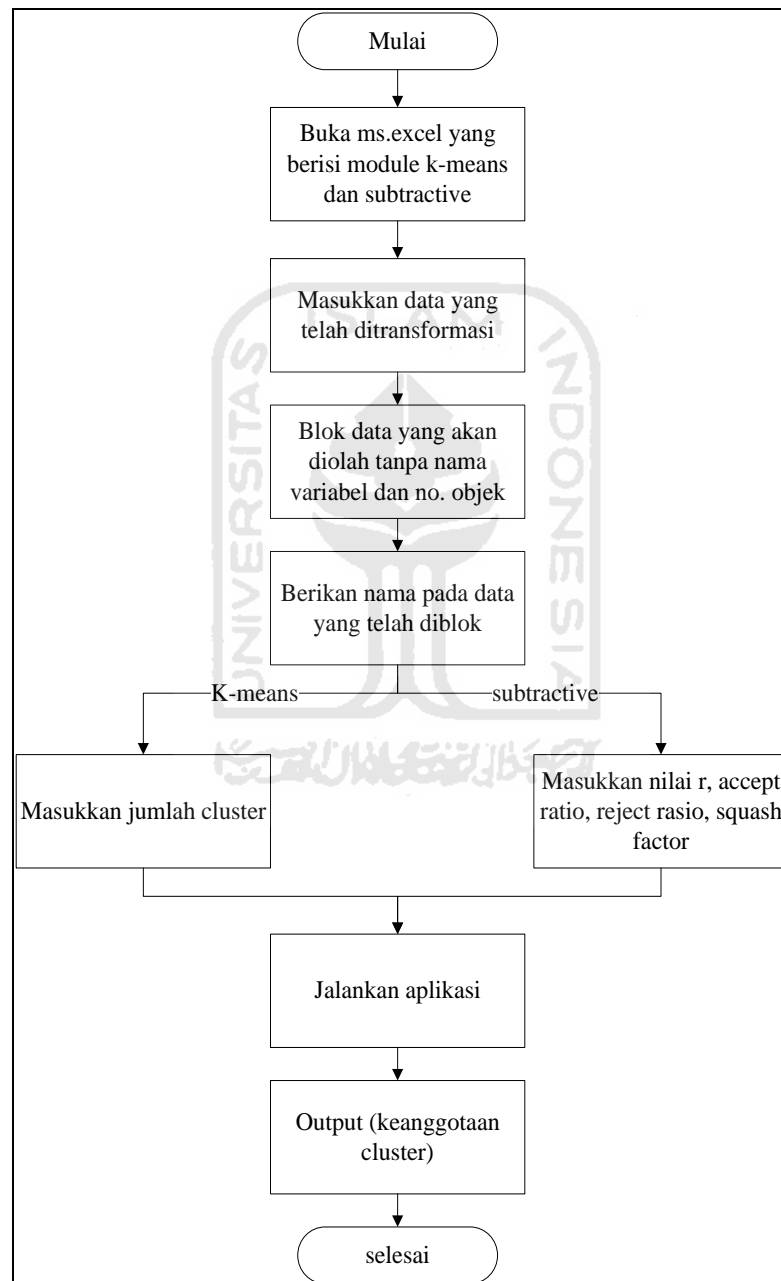
Output dari sistem yang dibuat adalah munculnya banyak *cluster* dan keanggotaan *cluster* untuk masing-masing data yang diolah. *Cluster member* akan ditampilkan disamping data yang diinputkan. Tanpa perlu mengetahui alur proses algoritma, seseorang dapat melakukan proses *clustering* ini dengan mudah. Contoh *output* sistem yaitu :

Data masukan					Output (keanggotaan cluster)		
	A	B	C	D	E	F	G
	No	X1	X2	X3	X4	Cluster member	
1	1	15000000	25000000	42	5000000	4	
2	2	20000000	26420000	72	5230000	4	
3	3	17820000	22052000	35	5200000	4	
4	4	16205000	18500000	12	4250000	1	
5	5	8000000	15200000	5	3500000	1	
6	6	14260000	19640000	15	4023000	1	
7	7	7025000	15230000	19	5000000	1	
8	8	25032000	34000000	28	8000000	1	
9	9	24320100	35100000	39	12500000	2	
10	10	25602000	38200000	43	13250000	2	
11	11	19872000	28000000	27	10500000	1	
12	12	19000000	25000200	41	6350000	4	
13	13	16540200	30000200	29	7525000	1	
14	14	28920000	41000000	58	15620000	2	
15	15	15870200	26750000	19	4025000	1	
16	16	26840320	39000200	47	13025000	2	
17	17	24601200	38450000	64	11000250	2	
18	18	21650000	37525000	60	9850000	2	
19	19	18602000	30500000	74	11230000	2	
20	20	35024000	52000000	73	18230000	3	
21	21	39024300	52050000	26	15725000	5	
22	22	27500000	36500000	6	10560000	5	
23	23	32500500	45600000	10	16583000	5	
24	24	27963000	40250000	38	13670000	2	
25	25	37250020	51000000	68	18530000	3	
26	26	16523000	26750000	9	8500000	1	
27	27	25690000	39565000	48	15250000	2	
28	28	33450000	51065000	37	21500000	5	
29	29	9850000	1350000	13	2000000	1	
30	30	16950000	24580000	18	4500000	1	
31	31						

Gambar 4.9 Output Sistem

4.4. Pengujian aplikasi

Untuk mengetahui aplikasi yang dibuat berjalan dengan baik atau tidak, perlu dilakukan pengujian terhadap aplikasi. Adapun langkah-langkah yang dilakukan adalah :



Gambar 4.10 Alur Pengujian Aplikasi

Dalam alur pengujian aplikasi dapat terlihat langkah-langkah pengujian aplikasi. Tahap pertama *user* memasukkan data yang akan diolah, kemudian memblok data tanpa nama variabel dan no objek. Setelah diblok, data tersebut dinamai. Untuk *handler module k-means user* cukup memasukkan jumlah *cluster* yang diinginkan. Dan untuk *handler module subtractive user* perlu memasukkan nilai *r* atau jari-jari, *accept ratio*, *reject ratio*, dan *squash factor*. Untuk algoritma *subtractive*, besar kecil nilai *r* akan mempengaruhi banyak *cluster* yang terbentuk. Aplikasi akan berjalan untuk melakukan perhitungan dengan masing-masing algoritma. Setelah akhir proses akan terlihat output dari *handler module* yang dibuat yaitu berupa keanggotaan *cluster* dari masing-masing data.

4.4.1. Pengujian Handler Module K-means

Untuk *handler module k-means*, Studi kasus yang akan dicobakan diperoleh dari buku *Introduction To Business Data mining* (Olson dan Shi, 2007). Data yang digunakan adalah data set pengeluaran seseorang untuk mengetahui seberapa banyak seseorang tersebut menghabiskan uangnya di toko bahan makanan, dengan 13 variabel yang disebutkan diantaranya umur, jenis kelamin, status pernikahan, tanggungan, pemasukan, lama tahun bekerja, lama tinggal, lama sekolah, lisensi mengemudi, kepemilikan rumah, kartu kredit, *churn*, *progrocesies*. Sebelum diinputkan dalam sistem data telah mengalami proses standarisasi atau transformasi untuk memudahkan pengolahan. Data tersebut adalah :

Tabel 4.1 Tabel Data Uji Modul *K-means*

Case	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
1	1	0	0.5	0	0.8	0.5	0	0.3	0	0	0.2	0	0.031
2	0.88	0	1	0	0.512	0.4	0	0	1	0	1	1	0.044
3	0.06	0	0	0	0.414	0.7	1	0	1	0	0.2	0	0

Case	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
4	0.56	0	1	0.2	0.591	1	0	0.3	1	1	0.8	0	0.073
5	0.72	1	1	0	0.574	1	0.9	0.3	1	1	1	1	0.037
6	0.14	1	1	0.2	0.292	0.6	0	0.3	1	1	0.4	0	0.14
7	0.68	1	1	0	0.415	1	0.75	1	1	0	1	1	0.052
8	0.66	1	1	0.4	0.433	0.3	0.45	0	0	0	0	0	0.059
9	0.72	0	1	0	0.311	1	0.1	0	1	0	1	1	0.069
10	0.9	0	1	0	0.59	1	0.4	0.3	1	1	0.6	0	0.036

Dari studi kasus diatas, jumlah cluster yang dimasukkan adalah 2, dan setelah diolah menggunakan *handler module k-means* hasilnya adalah *cluster* pertama beranggotakan 6 anggota yaitu 1,2,3,4,9,10 dan *cluster* kedua beranggotakan 4 anggota yaitu 5,6,7,8.

4.4.2. Pengujian *Handler Module Subtractive*

Data yang diujikan untuk *handler modul subtractive* diambil dari buku aplikasi logika *fuzzy* (Kusumadewi dan Purnomo, 2010). Disebutkan bahwa ada seorang pengusaha yang ingin membuat gedung untuk menyimpan produk dan bahan baku. Faktor – faktor yang dipertimbangkan untuk menentukan lokasi gedung adalah jarak dengan pasar dan sumber bahan baku. Lokasi gudang diharapkan dekat dengan pasar dan sumber bahan baku. Berdasarkan studi kelayakan yang dibuat, diperoleh informasi tentang posisi (dalam koordinat) pasar dan sumber bahan baku. Ada 20 pasar dan 10 sumber bahan baku dengan posisi yang terlihat seperti tabel dibawah ini :

Tabel 4.2 Tabel Data Uji modul *subtractive*

pasar/sumber bahan baku	koordinat [km dari 0.0]	
	X	Y
P1	0.9	1.2
P2	1.2	1.5
P3	6.4	0.6

pasar/sumber bahan baku	koordinat [km dari 0.0]	
	X	Y
P4	1.8	-0.2
P5	-2.5	1.8
P6	0.5	4.6
P7	2.3	1.8
P8	0.8	0.8
P9	-5.2	0.6
P10	-1.4	-4.6
P11	-0.8	0.9
P12	5.2	-5.2
P13	3.2	-6.3
P14	-5	-4.5
P15	0.6	-2.6
P16	2.5	2.3
P17	-4.2	-0.4
P18	-0.8	0.6
P19	2.6	1.4
P20	1.8	7.2
S1	4.6	1.8
S2	2.5	-8.2
S3	2.2	1.6
S4	-3.5	5.8
S5	1.5	4.9
S6	-8.1	1.6
S7	-2.6	2.1
S8	-5.4	-4.2
S9	4.6	-2.3
S10	1.3	3.5

Dari studi kasus diatas, dimasukkan nilai r sebesar 0.6, nilai *squash factor* sebesar 0.5, dan nilai *accept ratio* serta *reject ratio* masing-masing adalah 0.15 dan 0.5. Setelah diproses menggunakan *handler module subtractive* didapatkan dua *cluster*, masing-masing untuk *cluster* pertama beranggotakan data 1,2,3,4,5,6,7,8,11,12,13,15, 16,18,19,20,21,23,24,25,27,29,30. Sedangkan untuk *cluster* kedua beranggotakan data 9,10,14,17,22,26,28.

BAB V

PEMBAHASAN

5.1. *Handler Module* untuk Algoritma *K-means* dan *subtractive clustering*

Hasil dari pengolahan studi kasus yang diujicobakan untuk modul *k-means* adalah 2 *cluster* dengan keanggotaan masing-masing adalah untuk *cluster* pertama beranggotakan 6 anggota yaitu 1,2,3,4,9,10 dan untuk *cluster* kedua beranggotakan 4 anggota yaitu 5,6,7,8. Pola yang terbentuk dari pengolahan data tersebut adalah terkelompoknya data-data yang ada ke dalam masing-masing *cluster*. Data-data yang berada dalam satu *cluster* berarti memiliki kemiripan yang dekat sehingga mereka bisa bergabung menjadi satu *cluster*.

Untuk validasi *handler module k-means* maka akan dilakukan pembuktian dengan cara mengambil salah satu data secara acak, kemudian akan dicari jarak terdekat terhadap pusat *cluster*. Misalkan untuk data ke-4 akan dicari kecenderungan terhadap kedua pusat *cluster* yang ada. Pusat *cluster* pertama yaitu 1,2,3,4,9,10, pusat *cluster* kedua yaitu 5,6,7,8. Maka akan dihitung nilai rata-rata dari masing-masing pusat *cluster* dan akan dihitung juga jarak data ke-4 terhadap masing-masing *cluster*. Untuk nilai rata-rata masing-masing pusat *cluster* akan ditunjukkan pada Tabel 5.1 dibawah ini :

Tabel 5.1 Tabel Nilai Rata-Rata Pusat Cluster

Cls	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
1	0.687	0	0.75	0.033	0.536	0.767	0.25	0.15	0.833	0.333	0.633	0.333	0.042
2	0.55	1	1	0.15	0.429	0.725	0.525	0.4	0.75	0.5	0.6	0.5	0.072

Jarak data ke-4 terhadap pusat *cluster* yaitu :

$$d(4,I) = (0.56-0.687)^2 + (0-0)^2 + (1-0.75)^2 + (0.2-0.033)^2 + (0.591-0.536)^2 + (1-0.767)^2 + (0-0.25)^2 + (0.3-0.15)^2 + (1-0.833)^2 + (1-0.333)^2 + (0.8-0.6333)^2 + (0-0.333)^2 + (0.073-0.042)^2 = 0.860816917$$

$$d(4,II) = (0.56-0.55)^2 + (0-1)^2 + (1-1)^2 + (0.2-0.15)^2 + (0.591-0.429)^2 + (1-0.725)^2 + (0-0.525)^2 + (0.3-0.4)^2 + (1-0.75)^2 + (1-0.5)^2 + (0.8-0.6)^2 + (0-0.5)^2 + (0.073-0.072)^2 = 1.99275725$$

Dari perhitungan diatas dapat dilihat bahwa jarak data ke-4 terhadap *cluster* pertama adalah 0.8608816917 dan terhadap pusat *cluster* kedua adalah 1.99275725. jarak minimal data ke-4 adalah jarak data ke-4 terhadap pusat *cluster* pertama. perhitungan ini dapat membuktikan bahwa aplikasi dapat digunakan dengan baik, karena validasi juga menunjukkan bahwa data ke-4 masuk ke *cluster* pertama.

Untuk hasil yang diujicobakan kepada modul *subtractive* didapatkan dua *cluster* masing-masing untuk *cluster* pertama beranggotakan data 1,2,3,4,5,6,7,8,11, 12,13,15,16,18,19,20,21,23,24,25,27,29,30. Sedangkan untuk *cluster* kedua beranggotakan data 9,10,14,17,22,26,28. Hasil dari perhitungan ini sama dengan hasil dari perhitungan yang dilakukan Kusumadewi dan Purnomo (2010). Analisis yang didapat, dapat digunakan untuk pertimbangan perusahaan dalam mencari letak gudang yang strategis. Pola yang terbentuk dari pengolahan ini adalah tergabungnya pasar atau sumber bahan baku yang jaraknya berdekatan. Pasar atau sumber bahan baku

otomatis akan bergabung menjadi satu cluster ketika jarak posisi dari masing-masing titik berdekatan.

Untuk validasi modul *subtractive*, akan diujikan dengan memasukkan nilai r , ketika nilai r semakin besar maka jumlah *cluster* yang terbentuk akan semakin kecil, namun jika nilai r semakin kecil maka jumlah *cluster* akan semakin besar. Ketika nilai r sebesar 0.6 maka *cluster* yang terbentuk adalah dua *cluster*. Namun ketika nilai r sebesar 0.3 maka *cluster* yang terbentuk adalah 6 *cluster* dengan *cluster* pertama beranggotakan data 1,2,4,7,8,15,16,19,23,30. Untuk *cluster* kedua beranggotakan 5,9,11,17,18,24,26,27, sedangkan *cluster* ketiga beranggotakan data 10,14,28, sedangkan *cluster* keempat beranggotakan data ke 12,13,22, sedangkan *cluster* kelima beranggotakan data ke 3,21,29, sedangkan *cluster* keenam beranggotakan data ke 6,20,25.

5.2. Kelebihan dari sistem yang dibuat

Sistem yang dibangun memiliki beberapa kelebihan antara lain :

1. Aplikasi yang dibuat sangat memudahkan *user* untuk melakukan pengolahan data karena modul dibangun didalam *microsoft excel*, dimana *microsoft excel* merupakan program yang hampir semua perusahaan atau organisasi lain dapat menggunakannya.
2. Dibanding *software* komersial *data mining* yang lain, aplikasi ini lebih mudah digunakan dan tentunya perusahaan dapat menghemat biaya untuk melakukan pengolahan data.

5.3. Kekurangan dari sistem yang dibuat

Selain memiliki kelebihan, sistem juga memiliki kekurangan, antara lain :

1. Sistem bersifat semi otomatis, karena masih membutuhkan peran *user* dalam tahap transformasi data.
2. Aplikasi hanya dapat mengelompokkan data yang ada. Tidak sampai tahap profilisasi yang bertujuan untuk mengetahui profil setiap *cluster* yang sudah terkelompok menjadi satu.



BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Dari penelitian yang telah dilakukan, kesimpulan yang didapat yaitu bahwa modul pemegang untuk *clustering* dapat dibangun dengan menggunakan bahasa pemrograman VBA yang terdapat dalam *microsoft excel*. Dalam hal ini modul pemegang untuk *k-means* dan *subtractive clustering* dapat dibangun dengan baik, dengan dibuktikan bahwa modul pemegang tersebut dapat digunakan dengan baik untuk analisis *data mining*. Dengan dibangunnya modul pemegang untuk kedua algoritma tersebut maka diharapkan dapat mempermudah pengguna dalam melakukan proses *clustering* dalam *data mining*. Pengecekan terhadap modul pemegang telah dilakukan dan menunjukkan bahwa aplikasi yang dibangun dapat menyelesaikan masalah – masalah yang berhubungan dengan *clustering*.

6.2. Saran

Saran yang dapat direkomendasikan dari penelitian ini adalah

1. Untuk penelitian yang akan datang diharapkan dapat membangun aplikasi clustering berbasis *spreadsheet* dimulai dari tahap *transformasi* sampai dengan tahap *profilisasi*.

2. Diharapkan untuk dapat membangun modul pemegang untuk algoritma yang lain khususnya untuk permasalahan *Data mining*.



DAFTAR PUSTAKA

- Berry, M.J.A dan Linoff, G.S., 2004. *Data mining* Techniques For Marketing, Sales, And Costumer Relationship Management Second Edition. John Wiiley, Inc. Indianapolis, Indiana
- Chiu, S.L., 1994. Fuzzy Model Identification Based On Cluster Estimation, *Journal of intelligent and fuzzy system*, Vol.2 , 267-278.
- Chiu, S.L., 1997. extracting fuzzy rules from data for approxiamation and pattern classification. *Fuzzy Information Engginering : A Guided Tour of Application*.
- Fayyad, U., Shapiro, G.P., & Smyth, P., 1996. from *data mining* to knowledge discovery in databases. *Artificial Intelligence American Association for Artificial Intelligence*, 31-54.
- Han, Jiawei., and Kamber, Micheline., (2000). *Data mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA
- Dantes, R.G dan Agustini, K., 2008. Sistem Informasi Klasifikasi Tingkat Prestasi Akademik Mahasiswa Berdasarkan Seleksi Ujian Masuk Perguruan Tinggi Dengan Algoritma K-Means. *Jurnal Pendidikan dan Pengajaran*. Undhiksa.
- Darlington, K., 2004. *VBA for Excel Made Simple*. Burlington : Made Simple Book
- Doan, C.D., Liong, S. Y. dan Karunasinghe, D.S.K., 2005. Derivation of effective and efficient data set with subtractive clustering method and genetic algorithm. *Journal of Hydroinformatics*. 219-233.
- Junying, C., Zheng, Q., dan Jia, J., 2008. A Weighted Mean Subtractive Clustering Algorithm, *Information Technology Journal* vol.7, no 2, 356-360.
- Kusrini., 2007. Pendiskritan Kelas Kontinyu Dengan Algoritma K-means Clustering. Stimik Amikom.
- Kusumadewi, S dan Purnomo, H., 2010. *aplikasi logika fuzzy untuk pendukung keputusan*. Yogyakarta, Graha Ilmu.
- Larose, D. T. (2005). In *Discovering Knowledge In Data : An Introduction to Data mining* (p. 2). Hoboken, New Jersey: A John Wiley & Sons, Inc., Publication
- Moertini, V.S., 2002. *Data mining sebagai solusi bisnis*, INTEGRAL, vol. 7 no. 1.
- Olson, D dan Shi, Y., 2007. *Introduction to Business Data mining*. New York, America. McGraw-Hill Companies, Inc.
- Priyono, A., Muhammad Ridwan., Ahmad Jais Alias., Riza Atiq O. K Rahmat., Azmi Hassan., Mohd. Alaudin Mohd. Ali., 2005. Generation of Fuzzy Rules With Subtractive Clustering, *Jurnal Teknologi*, 43 (D), 143-153.

- Santosa, B., 2007. *Data mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta, Graha Ilmu.
- Sutiknyo, P. H. P., 2008. Penggolongan Suara berdasarkan Usia Dengan Menggunakan Metode K-means. Surabaya. Institut Teknologi Surabaya.
- Turban, E., Aronson, J.E., & Liang, T.P., 2005. *Decision Support System and Intelligent System*, Yogyakarta, ANDI
- Walkenbach, J., 2007. *Excel 2007 VBA Programming for Dummies*. John Wiley, Inc. Indianapolis, Indiana
- Yang, P., Zhu, Q., dan Zhong, X., 2009, Subtractive Clustering Based RBF Neural Network Model for Outlier Detection, *Journal of computers*, vol. 4, no 2.
- Zaiane, O.R., 1999. *Introduction to Data mining. Principles of Knowledge Discovery in Databases*.



LAMPIRAN

1. Handler Module K-means

```

Function Kmeans(sheetName As String, DataRange As String, JumlahCluster
As Integer)
Dim A As Double, B As Double
Dim Data() As Double
Dim baris As Integer, kolom As Integer
Dim i As Integer, j As Integer, k As Integer, m As Integer, n As
Integer

baris = Sheets(sheetName).Range(DataRange).Rows.Count
kolom = Sheets(sheetName).Range(DataRange).Columns.Count

ReDim Data(baris, kolom)

For A = 1 To kolom
  For B = 1 To baris
    Data(B, A) = Sheets(sheetName).Range(DataRange).Cells(B, A).Value
  Next B
Next A

Dim DataCluster() As String
Dim DataClusterSebelumnya() As String
Dim AnggotaCluster() As String
Dim RataRata() As Double
Dim Jarak() As Double
Dim jml As Double
Dim sama As Boolean
Dim minimal
Dim clusterterpilih As Double

ReDim DataCluster(1 To JumlahCluster)
ReDim DataClusterSebelumnya(1 To JumlahCluster)
Dim jmlpercluster As Integer
jmlpercluster = baris \ JumlahCluster

For i = 1 To JumlahCluster
  If (i <> JumlahCluster) Then
    For j = (((i - 1) * jmlpercluster) + 1) To (((i - 1) *
jmlpercluster) + jmlpercluster)
      If (j = (((i - 1) * jmlpercluster) + 1)) Then
        DataCluster(i) = j
      Else
        DataCluster(i) = DataCluster(i) & "," & j
      End If
    Next j
  Else
    For j = (((i - 1) * jmlpercluster) + 1) To baris
      If (j = (((i - 1) * jmlpercluster) + 1)) Then
        DataCluster(i) = j
      Else
        DataCluster(i) = DataCluster(i) & "," & j
      End If
    Next j
  End If
Next i

```

```

sama = False
n = 0
Do While (sama = False)
    n = n + 1
    ReDim RataRata(JumlahCluster, kolom)

    For i = 1 To JumlahCluster
        ReDim AnggotaCluster(1 To 1)
        AnggotaCluster = Split(DataCluster(i), ",")
        For j = 1 To kolom
            RataRata(i, j) = 0
            For k = 1 To baris
                For m = 0 To UBound(AnggotaCluster)
                    If (k = Val(AnggotaCluster(m))) Then
                        RataRata(i, j) = RataRata(i, j) + Data(k, j)
                    End If
                Next m
            Next k
            RataRata(i, j) = RataRata(i, j) / (UBound(AnggotaCluster) +
1)
        Next j
    Next i

    ReDim Jarak(baris, JumlahCluster)

    For i = 1 To baris
        For j = 1 To JumlahCluster
            Jarak(i, j) = 0
            For k = 1 To kolom
                Jarak(i, j) = Jarak(i, j) + ((Data(i, k) - RataRata(j,
k)) ^ 2)
            Next k
        Next j
    Next i

    For i = 1 To JumlahCluster
        DataClusterSebelumnya(i) = DataCluster(i)
    Next i

    For i = 1 To JumlahCluster
        DataCluster(i) = ""
    Next i

    Sheets(sheetName).Cells(Sheets(sheetName).Range(DataRange).Cells(1,
1).Row - 1, Sheets(sheetName).Range(DataRange).Cells(1, 1).Column +
kolom).Value = "Cluster Member"

    For i = 1 To baris
        minimal = 1000000 ^ 2
        For j = 1 To JumlahCluster
            If (minimal > Jarak(i, j)) Or (j = 1) Then
                clusterterpilih = j
                minimal = Jarak(i, j)
            End If
        Next j
        If DataCluster(clusterterpilih) = "" Then
DataCluster(clusterterpilih) = i Else DataCluster(clusterterpilih) =
DataCluster(clusterterpilih) & "," & i
        Next i

```



```

sama = True
For i = 1 To JumlahCluster
    If (DataClusterSebelumnya(i) <> DataCluster(i)) Then
        sama = False
    End If
Next i

If sama = True Then
    For i = 1 To JumlahCluster
        Next i
    End If
Loop

End Function

```

2. Handler Module subtractive

```

Option Explicit
Option Base 1
Function Subtractive(sheetName As String, dataRange As String, radius
As Double, acceptRatio As Double, rejectRatio As Double, squashFactor
As Double)
Dim A As Integer, B As Integer, C As Integer, D As Integer, E As
Integer, F As Integer
Dim data() As Double, normalisedData() As Double, xMin() As Double,
xMax() As Double, xGap() As Double
Dim distanceData() As Double, initPotencyData() As Double,
clusterCenters() As Double, denormal() As Double, sigma() As Double,
miu() As Double, stValue() As Double
Dim dcValue() As Double, centerCandidate() As Double
Dim noOfData As Integer, noOfDataCol As Integer
Dim minData As Double, maxData As Double, gapValue As Double
Dim distanceValue As Double
Dim potencyValue As Double, maxInitPotVal As Double
Dim mValue As Double, zValue As Double, ratioValue As Double, MdValue
As Integer, sdValue As Double, smdValue As Double
Dim centerCounter As Integer, centerPosition As Integer
Dim eval As Double
Dim clusterMember() As String

eval = 2.7182818284

noOfData = Sheets(sheetName).Range(dataRange).Rows.Count
noOfDataCol = Sheets(sheetName).Range(dataRange).Columns.Count

ReDim xMin(1, noOfDataCol)
ReDim xMax(1, noOfDataCol)
ReDim xGap(1, noOfDataCol)
ReDim data(noOfData, noOfDataCol)
ReDim normalisedData(noOfData, noOfDataCol)
ReDim distanceData(noOfData, noOfData)
ReDim initPotencyData(noOfData)
ReDim clusterCenters(noOfData, noOfDataCol)
ReDim dcValue(noOfData)
ReDim stValue(noOfData)
ReDim centerCandidate(1, noOfDataCol)
ReDim denormal(noOfData, noOfDataCol)
ReDim sigma(noOfDataCol)

```

```

For A = 1 To noOfDataCol
    For B = 1 To noOfData
        data(B, A) = Sheets(sheetName).Range(dataRange).Cells(B,
A).Value
    Next B
Next A

For A = 1 To noOfDataCol
    For B = 1 To noOfData
        If B = 1 Then
            minData = data(B, A)
            maxData = data(B, A)
        Else
            If data(B, A) >= maxData Then
                maxData = data(B, A)
            End If
            If data(B, A) <= minData Then
                minData = data(B, A)
            End If
        End If
    Next B
    xMin(1, A) = minData
    xMax(1, A) = maxData
    xGap(1, A) = maxData - minData
Next A

For A = 1 To noOfDataCol
    For B = 1 To noOfData
        normalisedData(B, A) = (data(B, A) - xMin(1, A)) / xGap(1, A)
    Next B
Next A

For A = 1 To noOfData
    potencyValue = 0
    For B = 1 To noOfData
        distanceValue = 0
        For C = 1 To noOfDataCol
            distanceValue = distanceValue + (((normalisedData(A, C) -
normalisedData(B, C)) / radius) ^ 2)
        Next C
        potencyValue = potencyValue + (eval ^ (-4 * distanceValue))
    Next B
    initPotencyData(A) = potencyValue
    If A = 1 Then
        maxInitPotVal = potencyValue
        centerPosition = A
    Else
        If potencyValue >= maxInitPotVal Then
            maxInitPotVal = potencyValue
            centerPosition = A
        End If
    End If
Next A

mValue = maxInitPotVal
zValue = maxInitPotVal

```

```

'Start iteration
centerCounter = 0
Do
    ratioValue = zValue / mValue

    If ratioValue > acceptRatio Then
acceptCenter:
        centerCounter = centerCounter + 1
        For A = 1 To noOfDataCol
            clusterCenters(centerCounter, A) =
normalisedData(centerPosition, A)
        Next A
        For A = 1 To noOfData
            potencyValue = 0
            For B = 1 To noOfDataCol
                potencyValue = potencyValue +
((clusterCenters(centerCounter, B) - normalisedData(A, B)) / (radius *
squashFactor)) ^ 2)
            Next B
            stValue(A) = potencyValue
            dcValue(A) = mValue * (eval ^ (-4 * stValue(A)))
            initPotencyData(A) = initPotencyData(A) - dcValue(A)
            If initPotencyData(A) < 0 Then initPotencyData(A) = 0
            If A = 1 Then
                maxInitPotVal = initPotencyData(A)
                centerPosition = A
            Else
                If initPotencyData(A) >= maxInitPotVal Then
                    maxInitPotVal = initPotencyData(A)
                    centerPosition = A
                End If
            End If
        Next A
        zValue = maxInitPotVal
        GoTo nextProcess
    End If

    If ratioValue < acceptRatio And ratioValue >= rejectRatio Then
reEvaluation:
        For A = 1 To noOfDataCol
            centerCandidate(1, A) = normalisedData(centerPosition, A)
        Next A
        MdValue = -1
        For A = 1 To centerCounter
            sdValue = 0
            For B = 1 To noOfDataCol
                sdValue = sdValue + (((centerCandidate(1, B) -
clusterCenters(A, B)) / radius) ^ 2)
            Next B
            If A = 1 Then
                MdValue = sdValue
            Else
                If sdValue <= MdValue Then MdValue = sdValue
            End If
        Next A
        smdValue = sdValue ^ 0.5
        If ratioValue + sdValue >= 1 Then
            GoTo acceptCenter

```

```

Else
    initPotencyData(centerPosition) = 0
    For A = 1 To noOfData
        If A = 1 Then
            maxInitPotVal = initPotencyData(A)
            centerPosition = A
        Else
            If initPotencyData(A) >= maxInitPotVal Then
                maxInitPotVal = initPotencyData(A)
                centerPosition = A
            End If
        End If
    Next A
    If maxInitPotVal = 0 Then Exit Do
    GoTo reEvaluation
End If
End If
If ratioValue < acceptRatio And ratioValue < rejectRatio Then Exit
Do
nextProcess:
Loop

For C = 1 To centerCounter
    For D = 1 To noOfDataCol
        denormal(C, D) = (clusterCenters(C, D) * (xMax(1, D) - xMin(1,
D))) + xMin(1, D)
    Next D
Next C

For D = 1 To noOfDataCol
    sigma(D) = 2 * (((radius * (xMax(1, D) - xMin(1, D)))) / (8 ^ (1 /
2))) ^ 2)
Next D

ReDim miu(noOfData, centerCounter)
ReDim clusterMember(centerCounter)
Dim maxmiu As Double
Dim choosen As String

Sheets(sheetName).Cells(Sheets(sheetName).Range(dataRange).Cells(1,
1).Row - 1, Sheets(sheetName).Range(dataRange).Cells(1, 1).Column +
noOfDataCol).Value = "Cluster Member"
For C = 1 To noOfData
    maxmiu = -1
    choosen = ""
    For D = 1 To centerCounter
        miu(C, D) = 0
        For E = 1 To noOfDataCol
            miu(C, D) = miu(C, D) + (((data(C, E) - denormal(D, E)) ^
2) / sigma(E))
        Next E
        miu(C, D) = eval ^ (-1 * miu(C, D))
        If maxmiu < miu(C, D) Then
            maxmiu = miu(C, D)
            choosen = D
        End If
    Next D
    If (clusterMember(choosen) = "") Then
        clusterMember(choosen) = C
    End If
Next C

```

```
Else
    clusterMember(choosen) = clusterMember(choosen) & "," & C
End If

    Sheets(sheetName).Cells(Sheets(sheetName).Range(dataRange).Cells(1,
1).Row + C - 1, Sheets(sheetName).Range(dataRange).Cells(1, 1).Column +
noOfDataCol).Value = choosen
Next C

For D = 1 To centerCounter
Next D

Sheets("Sheet1").Activate

End Function
```



3. Pengujian Handler Module Subtactive

Rekap Data Kepuasan Penumpang Bandara Sultan Syarif Kasim II Pekanbaru

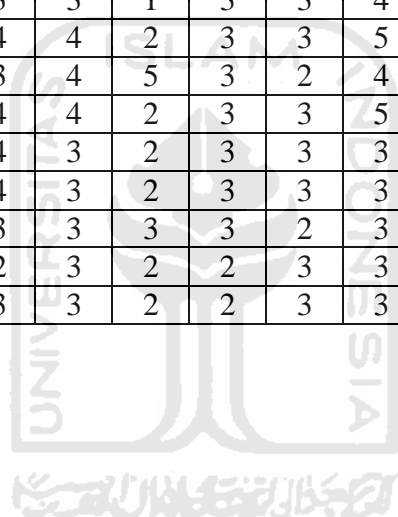
Responden	Jumlah Skor																			
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	X15	x16	x17	x18	x19	x20
1	2	3	4	2	3	2	2	2	3	1	2	2	2	1	3	1	3	3	2	1
2	3	3	3	3	1	2	2	3	1	2	3	2	3	4	4	3	3	3	2	2
3	3	4	3	3	3	2	2	4	5	1	4	4	5	3	5	5	2	3	3	2
4	2	3	3	4	3	4	2	3	4	5	3	2	4	5	3	2	4	3	3	4
5	3	4	5	5	3	4	4	4	4	3	3	4	5	4	5	4	5	5	4	3
6	2	3	3	3	3	2	2	2	3	2	3	3	2	2	3	3	3	3	3	2
7	3	3	3	4	3	4	3	3	3	2	1	3	3	3	3	2	3	3	2	2
8	3	4	5	5	5	4	4	3	3	2	4	3	4	3	5	4	4	3	3	2
9	2	3	2	4	3	4	3	2	4	3	2	3	2	3	4	3	2	3	2	1
10	2	3	3	4	3	4	2	3	3	2	2	3	3	3	3	3	2	3	2	1
11	2	3	4	3	5	2	4	3	3	2	3	3	3	3	3	4	3	3	2	2
12	3	3	3	3	2	2	2	3	4	2	3	3	3	3	4	4	2	3	3	2
13	3	3	3	3	2	2	2	3	3	1	3	3	4	4	4	4	2	3	3	3
14	2	3	3	4	3	3	2	3	4	1	2	4	3	3	5	4	4	3	4	3
15	3	3	3	3	2	2	1	3	3	2	3	4	3	3	4	3	3	4	1	1
16	1	2	3	1	2	2	1	5	4	1	4	4	4	2	3	3	3	3	2	1
17	2	3	4	3	3	3	4	3	3	2	2	3	2	2	2	2	3	4	2	2
18	3	3	4	4	4	4	4	4	4	4	3	3	4	4	4	4	3	4	4	4
19	3	4	3	4	4	4	4	3	3	3	4	3	4	3	4	4	3	4	3	3
20	2	2	3	4	3	2	3	2	2	2	2	2	2	2	4	4	3	4	3	2
21	3	4	3	5	4	4	5	4	5	4	3	5	4	5	5	5	5	5	3	4
22	3	2	3	3	4	3	2	3	4	4	2	3	3	2	2	3	3	4	2	3
23	2	2	3	2	2	2	3	3	3	2	3	3	4	3	3	4	3	3	2	3
24	1	3	4	4	2	3	2	4	3	1	4	4	5	3	5	5	3	5	3	3
Responden	Jumlah Skor																			
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20
25	2	3	2	3	1	2	3	3	1	2	3	2	3	4	4	3	3	4	2	2
26	2	3	3	4	2	2	2	3	4	2	3	4	4	3	4	4	3	3	3	2
27	2	4	3	2	4	3	4	3	3	2	2	3	3	4	3	2	3	3	2	2
28	2	4	3	4	2	2	4	4	4	2	3	3	5	3	4	4	3	3	2	3
29	2	3	2	4	3	3	4	4	3	2	3	3	3	3	4	4	3	4	3	3
30	1	3	4	2	3	2	4	2	3	1	2	2	2	1	3	1	3	4	2	1
31	3	4	3	3	4	3	3	4	4	4	3	2	3	4	4	3	4	3	3	4
32	2	5	3	4	2	3	4	2	3	4	2	2	3	3	3	3	3	3	3	2
33	3	3	3	4	3	4	2	3	4	5	3	2	4	5	3	2	4	4	3	4
34	1	2	3	4	3	2	3	2	2	2	2	2	2	2	4	4	3	5	3	2
35	2	3	2	2	3	3	3	3	3	2	2	2	3	3	3	2	4	3	3	3
36	2	4	3	2	4	4	4	3	3	2	2	3	3	4	3	2	3	4	2	2
37	1	4	5	5	3	4	4	4	4	3	3	4	5	4	5	4	5	5	4	3
38	3	3	3	1	4	4	3	5	5	2	4	3	2	4	4	5	4	4	3	3
39	2	4	4	4	4	4	4	3	3	3	4	3	4	3	4	4	3	4	3	3

40	2	2	3	3	2	2	2	3	4	2	4	4	4	3	4	4	3	4	3	2
41	2	3	3	4	2	2	2	3	4	2	3	4	4	3	4	4	3	4	3	2
42	3	3	3	3	2	2	2	3	4	2	3	3	3	3	4	4	3	3	3	2
43	1	3	3	3	2	2	2	3	3	1	3	3	4	4	4	4	3	3	3	3
44	2	3	4	2	3	2	2	2	3	1	2	2	2	1	3	1	3	3	2	1
45	2	3	3	3	1	2	2	3	1	2	3	2	3	4	4	3	3	3	2	2
46	3	4	3	3	3	2	2	4	5	1	4	4	5	3	5	5	2	3	3	2
47	3	3	3	4	3	4	2	3	4	5	3	2	4	5	3	2	4	3	3	4
48	3	4	5	5	3	4	4	4	4	3	3	4	5	4	5	4	5	3	4	3
49	1	3	3	3	3	2	2	2	3	2	3	3	2	2	3	3	3	4	3	2
50	1	3	3	4	3	4	3	3	3	2	1	3	3	3	3	2	3	5	2	2
51	2	4	5	5	5	4	4	3	3	2	4	3	4	3	5	4	4	5	3	2

Kondisi	Jumlah Skor																			
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20
52	2	3	2	4	3	4	3	2	4	3	2	3	2	3	4	3	2	4	2	1
53	3	3	3	4	3	4	2	3	3	2	2	3	3	3	3	2	3	2	1	1
54	2	3	4	3	5	2	4	3	3	2	3	3	3	3	3	4	3	3	2	2
55	1	3	3	3	2	2	2	3	4	2	3	3	3	3	4	4	2	3	3	2
56	3	3	3	3	2	2	2	3	3	1	3	3	4	4	4	4	2	3	3	3
57	3	3	3	4	3	3	2	3	4	1	2	4	3	3	5	4	4	3	4	3
58	2	3	3	3	2	2	1	3	3	2	3	4	3	3	4	3	3	3	1	1
59	3	2	3	1	2	2	1	5	4	1	4	4	4	2	3	3	3	3	2	1
60	2	3	4	3	3	3	4	3	3	2	2	3	2	2	2	2	3	4	2	2
61	1	3	4	4	4	4	4	4	4	4	3	3	4	4	4	4	3	4	4	4
62	2	4	3	4	4	4	4	3	3	3	4	3	4	3	4	4	3	4	3	3
63	1	2	3	4	3	2	3	2	2	2	2	2	2	2	4	4	3	4	3	2
64	1	4	3	5	4	4	5	4	5	4	3	5	4	5	5	5	5	5	3	4
65	2	2	3	3	4	3	2	3	4	4	2	3	3	2	2	3	3	4	2	3
66	2	2	3	2	2	2	3	3	3	2	3	3	4	3	3	4	3	3	2	3
67	3	3	4	4	2	3	2	4	3	1	4	4	5	3	5	5	3	5	3	3
68	2	3	2	3	1	2	3	3	1	2	3	2	3	4	4	3	3	4	2	2
69	2	3	3	4	2	2	2	3	4	2	3	4	4	3	4	4	3	3	3	2
70	2	4	3	2	4	3	4	3	3	2	2	3	3	4	3	2	3	3	2	2
71	3	4	3	4	2	2	4	4	4	2	3	3	5	3	4	4	3	3	2	3
72	2	3	2	4	3	3	4	4	3	2	3	3	3	3	4	4	3	4	3	3
73	2	3	4	2	3	2	4	2	3	1	2	2	2	1	3	1	3	4	2	1
74	3	4	3	3	4	3	4	4	4	4	3	2	3	4	4	3	4	4	3	4
75	2	3	3	4	2	3	4	3	3	3	3	4	3	4	3	3	3	3	3	3
76	2	3	3	2	3	2	4	2	4	3	2	2	2	2	4	4	3	4	3	3
77	2	2	3	3	2	2	4	3	4	2	4	4	4	3	4	4	3	5	3	2
78	3	3	3	1	4	4	4	5	5	2	4	3	2	4	4	5	4	3	3	3

Kondisi	Jumlah Skor																			
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20
79	2	3	2	2	3	3	4	3	3	2	2	2	3	3	3	2	4	4	3	3
80	2	4	4	4	2	2	4	3	5	2	3	3	5	2	4	5	3	3	3	2

81	3	3	3	4	4	4	4	4	3	2	3	3	3	3	3	4	3	3	3	3
82	2	3	2	2	3	2	4	3	4	3	3	2	3	3	4	4	3	3	3	2
83	2	5	3	4	2	3	3	2	3	4	2	2	3	3	3	3	3	3	2	
84	2	3	3	2	3	2	4	3	2	3	2	3	2	3	2	3	3	4	2	
85	2	4	3	4	4	4	3	2	4	3	3	3	3	4	2	2	3	4	2	
86	2	2	3	3	2	2	4	3	3	2	2	2	2	2	2	2	3	3	2	
87	3	3	4	5	3	3	3	4	4	1	2	4	3	3	3	3	5	4	3	
88	3	4	4	3	4	4	4	3	3	2	2	3	2	3	3	3	3	4	3	
89	2	4	3	4	4	4	3	2	4	3	3	3	3	4	2	2	3	4	2	
90	2	3	4	1	5	4	4	3	5	4	1	3	3	2	2	4	4	4	2	
91	2	3	4	2	3	3	4	3	3	2	2	2	3	3	3	2	4	3	3	
92	3	3	3	3	2	2	3	2	3	1	3	3	4	4	4	4	2	3	3	
93	2	3	3	4	3	3	4	3	4	1	2	4	3	3	5	4	4	3	3	
94	2	3	3	3	1	2	3	2	1	2	3	2	3	4	4	3	3	3	2	
95	2	3	3	3	2	2	4	3	3	1	3	3	4	4	4	4	3	2	3	
96	3	4	3	4	2	2	3	4	4	2	3	3	5	3	4	4	3	3	2	
97	3	3	3	4	3	4	4	3	4	5	3	2	4	5	3	2	4	3	3	
98	3	4	3	4	2	2	4	4	4	2	3	3	5	3	4	4	2	4	2	
99	2	3	3	4	3	3	4	4	3	2	3	3	3	3	4	4	3	4	3	
00	2	3	3	4	4	4	3	4	3	2	3	3	3	3	3	4	3	4	3	
01	2	2	2	3	2	3	3	3	3	3	3	2	3	3	3	3	3	3	3	
02	2	2	2	3	4	3	4	2	3	2	2	3	3	4	2	3	3	3	2	
03	2	2	3	3	3	4	3	3	3	2	2	3	3	3	3	3	3	3	2	



Dari studi kasus diatas, dimasukkan nilai r sebesar 1.5, nilai *squash factor* sebesar 1.25, dan nilai *accept ratio* serta *reject ratio* masing-masing adalah 0.15 dan 0.5. Setelah diproses menggunakan *handler module subtractive*, hasilnya adalah terdapat 5 *cluster*, *cluster* pertama beranggotakan 36 anggota yaitu 5, 8, 18, 19, 20, 21, 22, 24, 25, 29, 34, 37, 38, 39, 48, 51, 61, 62, 63, 64, 65, 67, 68, 72, 74, 75, 76, 78, 81, 82, 87, 93, 99, 100, 101, 103, *cluster* kedua beranggotakan 31 anggota yaitu 2, 3, 12, 13, 14, 1, 16, 23, 26, 28, 40, 41, 42, 43, 45, 46, 55, 56, 57, 58, 59, 66, 69, 71, 77, 80, 92, 94, 95, 96, 98 , untuk *cluster* ketiga beranggotakan 24 anggota yaitu 7, 9, 10, 11, 17, 27, 32, 35, 36, 50, 52, 53, 54, 60, 70, 79, 83, 84, 85, 88, 89, 90, 91, 102, untuk *cluster* keempat beranggotakan 5 anggota yaitu 4,31,33,47,97 dan untuk *cluster* kelima beranggotakan 7 anggota yaitu 1,6,30,44,49,73,86.



4. Pengujian *Handler Module K-means*

Studi kasus yang pertama diambil dari buku analisis multivariat (Supranto J, 2004). Dalam buku ini terdapat studi kasus untuk mengelompokkan pelanggan berdasarkan sikap mereka terhadap berbelanja (*shopping*). Disebutkan terdapat 6 variabel sikap yang diidentifikasi. Keenam variabel tersebut adalah : berbelanja itu menyenangkan, berbelanja kurang baik atau tak bagus bagi anggaran, saya biasanya berbelanja sekaligus dengan makan diluar, saya mencoba mendapatkan harga termurah sewaktu berbelanja, saya tidak begitu peduli dengan berbelanja, saya bisa menghemat uang banyak dengan cara melakukan perbandingan harga. Pelanggan diminta untuk mengekspresikan tingkat persetujuan untuk masing pernyataan, dengan pernyataan 7 butir dimana 7 sangat setuju dan 1 sangat tidak setuju. Didapat 20 data siap pakai yang akan diuji cobakan kedalam sistem yang telah dibuat. Data tersebut adalah :

Tabel 5.1 Rekap Data Uji *handler module k-means*

Case No	X1	X2	X3	X4	X5	X6
1.	6	4	7	3	2	3
2.	2	3	1	4	5	4
3.	7	2	6	4	1	3
4.	4	6	4	5	3	6
5.	1	3	2	2	6	4
6.	6	4	6	3	3	4
7.	5	3	6	3	3	4
8.	7	3	7	4	1	4
9.	2	4	3	3	6	3
10.	3	5	3	6	4	6
11.	1	3	2	3	5	3
12.	5	4	5	4	2	4
13.	2	2	1	5	4	4
14.	4	6	4	6	4	7
15.	6	5	4	2	1	4
16.	3	5	4	6	4	7
17.	4	4	7	2	2	5
18.	3	7	2	6	4	3
19.	4	6	3	7	2	7
20.	2	3	2	4	7	2

Dari studi kasus diatas, jumlah cluster yang dimasukkan adalah 3, dan setelah diolah menggunakan *handler module k-means* yang telah dibuat, hasilnya adalah *cluster* pertama beranggotakan 8 anggota yaitu 1,3,6,7,8,12,15,17, *cluster* kedua beranggotakan 6 anggota yaitu 2,5,9,11,13,20 dan *cluster* ketiga beranggotakan 6 anggota yaitu 4,10,14,16,18,19.

