

**PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK
MENENTUKAN STRATEGI PROMOSI UNIVERSITAS ISLAM NEGERI
SUNAN KALIJAGA YOGYAKARTA**

Skripsi

Untuk Memenuhi Sebagian Persyaratan

Mencapai Derajat Sarjana S-1

Program Studi Teknik Informatika



Disusun Oleh

Bayu Resi Indrawan

12651052

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2018



PENGESAHAN SKRIPSI/TUGAS AKHIR

Nomor : B-292/Un.02/DST/PP.05.3/06/2018

Skripsi/Tugas Akhir dengan judul : Penerapan Algoritma K-Means Clustering untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta

Yang dipersiapkan dan disusun oleh :
Nama : Bayu Resi Indrawan
NIM : 12651052
Telah dimunaqasyahkan pada : 11 Mei 2018
Nilai Munaqasyah : A/B
Dan dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga

TIM MUNAQASYAH :

Ketua Sidang

Agus Mulyanto, M.Kom
NIP. 19710823 199903 1 003

Penguji I

Maria Ulfah Siregar, M.IT. Ph.D
NIP.19780106 200212 2 001

Penguji II

Sumarsono, M. Kom
NIP.19710209 200501 1 003

Yogyakarta, 7 Juni 2018
UIN Sunan Kalijaga
Fakultas Sains dan Teknologi
Dekan



Dr. Murtono, M.Si
NIP. 19691212 200003 1 001



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp : 1 Bendel Skripsi

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Bayu Resi Indrawan

NIM : 12651052

Judul Skripsi : Penerapan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta

Sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 30 April 2018

Pembimbing

Agus Mulyanto, S.Si, M.Kom

NIP. 19710823 199903 1 003

SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Bayu Resi Indrawan

NIM : 12651052

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi dengan judul **“Penerapan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta”** tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 30 April 2018

Yang Menyatakan



Bayu Resi Indrawan
NIM. 12651052

KATA PENGANTAR

Segala puji bagi Allah SWT tuhan semesta alam yang selalu memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Penerapan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Islam Negeri Yogyakarta”. Tak lupa pula penulis haturkan salawat serta salam kepada Nabi junjungan kita Nabi Muhammad SAW yang telah berjuang demi berdiri tegaknya daulah islamiyah di muka bumi ini.

Penulis juga mengucapkan terimakasih kepada semua pihak yang telah membantu dalam proses pelaksanaan penelitian tugas akhir ini sehingga laporan tugas akhir ini dapat terselesaikan.

Selanjutnya penulis mengucapkan terimakasih kepada :

1. Bapak Prof. Drs. K.H. Yudian Wahyudi, M.A., Ph.D. selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Bapak Dr. Murtono, M.Si, selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Dr. Bambang Sugiantoro, MT, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
4. Bapak Agus Mulyanto, S.SI, M.Kom selaku Dosen Pembimbing skripsi yang telah menasehati, dan memberi pengarahkan dengan segala kesabarannya.

5. Seluruh Dosen Program Studi Teknik Informatika yang telah memberi bekal ilmu pengetahuan kepada penulis, semoga ilmunya menjadi amal jariyah.
6. Semua Pegawai dan Staff di bagian akademik kemahasiswaan Universitas Islam Yogyakarta.
7. Bapak Amanu Budi Santoso, ibu Sulastri serta seluruh anggota keluarga besar yang tidak bisa penulis tulis satu per satu atas doa, perhatian, kasih sayang dan dukungan moril maupun materilnya.
8. Teman-teman Program Studi Teknik Informatika, khususnya angkatan 2012 Mandiri Kelas K (Katak'12) tercinta yang telah banyak memberi dukungan dan semangat yang sangat besar.
9. Serta semua pihak yang tidak dapat disebutkan satu per satu, yang telah banyak memberikan dukungan, motivasi, inspirasi dan membantu dalam proses penyelesaian skripsi ini.

Penulis menyadari masih banyak sekali kekurangan dalam penelitian ini, oleh karena itu kritik dan saran senantiasa penulis harapkan. Akhir kata semoga penelitian ini dapat menjadi panduan serta referensi yang sangat berguna bagi pembaca dan dapat dimanfaatkan dalam pengembangan ilmu pengetahuan.

Yogyakarta, 26 April 2018

Bayu Resi Indrawan
12651052

HALAMAN PERSEMBAHAN

Dengan mengucap segala rasa syukur penulis mempersembahkan tugas akhir ini untuk :

- ❖ Ayahku yang telah berjuang sejauh ini buatku, Ibuku yang tak mengenal lelah untuk selalu mendoakan ku dan tetap menjadi motivasi terbesar dalam perjalanan hidupku. Kedua malaikat tanpa sayapku yang tak pernah bosan mendoakan dan menyayangiku, yang terus mendukungku sampai sejauh ini. Semoga Ayah dan ibuku panjang umur dan bisa melihatku menjadi anak yang membanggakan keluarga suatu hari nanti, amin.
- ❖ Adik-adikku yang selalu ada disisiku disetiap aku membutuhkan pertolongan, Gandhang Resi arta yang senantiasa selalu menjaga kebugaranku dengan rutin mengajak berolahraga setiap minggu nya di telaga futsal dan Galit Resi Ardhana yang selalu menjaga kesehatanku sehingga aku tidak pernah kebingungan apabila sakit sedang menghampiriku, kakak mu ini akan selalu sayang kalian.
- ❖ Dosen dan keluarga besar Teknik Informatika, Pak Sumarsono ketua program studi yang selalu sedia dan terbuka menerima keluhan kesah para mahasiswanya. Pak Didik yang selalu mengarahkan dan selalu peduli kepada anak bimbingnya, Ibu Ade, Pak Mustakim, Pak Agung, Ibu Uyun, Pak Bambang, Pak Rahmat, Pak Agus Mulyanto dan Pak Aulia yang selalu sabar memberikan ilmu-ilmunya. semoga Bapak dan Ibu

dosen panjang umur, diberikan kesehatan, dan selalu bahagia sampai tua kelak, amin.

- ❖ Teman - teman seperjuangan dan keluarga besar Teknik Informatika Mandiri /Khusus 2012 (Katak'12) yang sangat aku cintai, Nur Kukuh Wicaksana, Juhdan Aja, Firmansyah Gustap Hikmatyar, Gatot Cakra Sutradana, Ikhzan Zuhriyanto, Edi Gunawan yang entah bagaimana kabarnya, Eri Tulang (Eri Kurniawan), Dwi Kumala Mursid, Iwan Peyek (Iwan Nasrullah), Mas Mandrok (Irfan Andriyanto), Nuge Jamp (Hilyas Riza), Pakdhe (Nur Rohman), Ripah Gemah Nuripah, Andek (Andi Wijanarko), Muhammad Edi Iswanto, Teh Lusi Anggarini, Dwi Indah Permatasari, Rizky Fahrezi, Fajar Septian Nugraha, Indra Faisol Alim, Ainul Yaqin Muntaqo, Luqyana Khalda Kiki, Ahmad Kliwir (Ahmad Choirudin),Ahmad Deviyanto, Krisna Adi Yogantoro,Rahulloh Ayyatulloh Komaeni Noor Bintang,Malika Dhini, Ulvi Ismaya, Fahrieza Rahmadziba, Gumeta Sari Maharani,Afin Hilman,M Berlian Rafsanjani, Nanang Paminto Aji, Valdi Adrian Abrar, Kharizma Azdhi, Erin Septiana, Novie Astutie, Nurul Zuni Rahmawati,Tri Wiji Hastuti, Taufik, Edita Rizki, Asep Baeri, Irham, Ulfa Septi Muslimah, Perwira Putra Agni, Surahmat Laguni dan Abdul Mukhrim, terimakasih atas semua kenangan indah yang telah kalian berikan. Semoga persahabatan ini tidak akan terpisahkan sampai hari akhir nanti.

- ❖ Pihak - pihak yang selalu memberikan bantuannya, semangat, dan doanya baik secara langsung maupun tidak yang tidak dapat penulis sebutkan namanya satu per satu.



MOTTO

*Barang siapa yang keluar dalam menuntut ilmu maka
ia adalah seperti berperang di jalan Allah
hingga ia pulang*

(H.R.Tirmidzi)

*Belajar dari masa lalu, hidup untuk masa kini, dan
berharap untuk masa yang akan datang*

(Albert Einstein)

*Semakin sulit tantangannya, semakin besar kebahagiaan yang
anda rasakan saat mampu melewatinya.*

(Gianluigi Buffon)

*Ayo segera bangun mimpimu atau orang lain
akan memperkejakan kamu untuk
membangun mimpi mereka*

(Farrah Grey)

**PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK
MENENTUKAN STRATEGI PROMOSI UNIVERSITAS ISLAM NEGERI
YOGYAKARTA**

Bayu Resi Indrawan

NIM. 12651052

INTISARI

Pada proses penerimaan mahasiswa baru Universitas UIN Yogyakarta akan menghasilkan data-data yang baru yang berlimpah berupa profil dari mahasiswa baru. Berdasarkan hal di atas informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data mahasiswa sehingga menghasilkan informasi yang bermanfaat bagi pihak universitas. Informasi yang didapatkan dapat membantu pihak universitas sebagai bahan pertimbangan dalam menentukan strategi promosi penerimaan mahasiswa baru di tahun depan.

Penelitian ini bertujuan untuk mengelompokkan data mahasiswa kedalam sebuah cluster dengan menggunakan metode Data Mining algoritma K-Means Clustering. Data mahasiswa dikelompokkan berdasarkan kemiripan data tersebut sehingga data dengan karakteristik yang sama akan berada dalam satu cluster. Atribut yang digunakan adalah kota asal, asal sekolah, dan IPK mahasiswa selama 2 semester awal dengan nilai di atas 2,75.

Cluster yang terbentuk setelah proses K-Means Clustering adalah tiga cluster dengan cluster pertama berjumlah 508 data mahasiswa, cluster kedua 139 data mahasiswa, dan cluster ketiga 62 data mahasiswa. Cluster dengan nilai rata – rata IPK paling tinggi adalah cluster satu. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan cluster yang terbentuk oleh pihak Universitas UIN Yogyakarta.

Kata Kunci: Asal sekolah, kota asal, IPK, *K-Means Clustering*, Data Mining, Universitas UIN Yogyakarta.

**APPLICATION OF DATA MINING FOR DETERMINING PROMOTION
STRATEGY UIN UNIVERSITY YOGYAKARTA USING
K-MEANS CLUSTERING ALGORITHM**

Bayu Resi Indrawan

NIM. 12651052

ABSTRACT

New admissions process of UIN Yogyakarta University students generate data that are highly abundant in the form of student profile data. Based on it some hidden information could be known by doing data processing using that's profile's data and also obtaining useful information for university. Information that we get could be contribute for university as a consideration to decide new admissions promotion in the next year.

This research aims to classify student data's into a cluster by utilizing Data Mining processing clustering techniques. The algorithm used for the cluster techniques is K-Means algorithm. K-Means is one method of non-hierarchical clustering of data that can group student data into several clusters based on the similarity of the data, so the data of students who have similar characteristics are grouped into one cluster and that have different characteristics grouped in another cluster. Attributes that used in this study is student's country, student's high school, and GPA for two semesters with a value above 2,75.

Cluster that formed after K-Means Algorithm process is three cluster with the first cluster amounted to 508 student data, second cluster amounted to 139 student data, and the third cluster amounted to 62 student data. Cluster with the highest average GPA is the first cluster. The results of this research are used for a making decision to determine promotion strategy based on clusters formed.

Keywords: student's high school, country, GPA, *K-Means Clustering*, Data Mining, Universitas UIN Yogyakarta.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN	iv
KATA PENGANTAR	iv
PERSEMBAHAN	vi
MOTTO	ix
INTISARI	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xxii
DAFTAR RUMUS	xix
DAFTAR LAMPIRAN	xx
DAFTAR SINGKATAN	xxi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3

1.6 Keaslian Penelitian	4
1.7 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	6
2.1 Tinjauan Pustaka	6
2.2 Landasan Teori	10
2.2.1 Data Mining	10
2.2.2 Teknik – Teknik Data Mining	16
2.2.3 Algoritma K-Means Clustering	22
2.2.4 Konsep Aplikasi Berbasis Web	25
2.2.5 Definisi Sistem	25
2.2.6 World Wide Web	27
2.2.7 Data Flow Diagram (DFD)	27
2.2.8 Entity-Relationship Diagram (ERD)	28
2.2.9 Pengolahan Data Microsoft Excel	30
BAB III METODE PENELITIAN	32
3.1 Metode Penelitian	32
3.2 Peralatan Penelitian	32
3.2.1 Perangkat Keras	32
3.2.2 Perangkat Lunak	32
3.3 Tahap – Tahap Penelitian	33
3.3.1 Pengumpulan Data	33
3.3.2 Pre-processing Data	34
3.3.3 K-Means Clustering	35

BAB IV HASIL DAN PEMBAHASAN	39
4.1 Objek Penelitian	39
4.2 Lokasi Penelitian	39
4.3 Pengumpulan Data	40
4.4 Pre-processingData	42
4.4.1 DataReduction	42
4.4.2 DataCleaning	44
4.4.3 Data Transformation	45
4.4.4 Data Integration	50
4.5 Penerapan K-Means Clustering	52
4.5.1 Penerapan Algoritma Menggunakan Microsoft Excel 2010	53
4.5.2 Hasil Penerapan K-Means menggunakan Microsoft Excel 2010	60
4.5.3 Penerapan Algoritma Menggunakan Aplikasi PHP.....	67
4.5.4 Perancangan Sistem Aplikasi K-Means Clustering	68
4.5.5 Implementasi Sistem Aplikasi K-Means Clustering	84
4.5.6 Hasil Implementasi Sistem Aplikasi K-Means Clustering	96
BAB V PENUTUP	102
5.1 Kesimpulan	102
5.2 Saran	103
DAFTAR PUSTAKA	104

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terdahulu	8
Tabel 4.1 Tabel Frekuensi Kota Asal Mahasiswa	46
Tabel 4.2 Tabel Frekuensi Asal Sekolah Mahasiswa	50
Tabel 4.3 Tabel Titik Pusat Cluster	54
Tabel 4.4 Tabel Titik Pusat Cluster Baru (Cluster 1)	58
Tabel 4.5 Tabel Titik Pusat Cluster Baru (Cluster 2)	58
Tabel 4.6 Tabel Titik Pusat Cluster Baru (Cluster 3)	58
Tabel 4.7 Tabel Hasil Cluster 1 Penerapan Algoritma K-Means Clustering (Microsoft Excel)	61
Tabel 4.8 Tabel Hasil Cluster 2 Penerapan Algoritma K-Means Clustering (Microsoft Excel)	62
Tabel 4.9 Tabel Hasil Cluster 3 Penerapan Algoritma K-Means Clustering (Microsoft Excel)	64
Tabel 4.10 Tabel Database Dataset	77
Tabel 4.11 Tabel Database Frek_asek	78
Tabel 4.12 Tabel Database Frek_Kota	79

Tabel 4.13 Tabel Database Centroid_lama80

Tabel 4.14 Tabel Database Nomor80

Tabel 4.15 Tabel Database Proses81



DAFTAR GAMBAR

Gambar 2.1 Tahapan Proses KDD.....	12
Gambar 2.2 Classification	16
Gambar 2.3 Clustering	18
Gambar 2.4 Grafik 3-year price history	20
Gambar 2.5 Sequence	21
Gambar 2.6 Model Sistem Menurut Scott	26
Gambar 2.7 Notasi Dasar Data Flow Diagram (DFD)	28
Gambar 2.8 Simbol-simbol Entity Relationship Diagram (ERD)	29
Gambar 2.9 Tampilan Microsoft Excel	31
Gambar 4.1 Verifikasi DataCentroid (Microsoft Excel)	60
Gambar 4.2 DFD Level 0 (Diagram Konteks)	69
Gambar 4.3 DFD Level 1	70
Gambar 4.4 ERD (Entity Relationship Diagram)	75
Gambar 4.5 Relasi Database (sistem K-Means clustering)	83
Gambar 4.6 Contoh File excel Yang Diimport Kedalam Sistem	84

Gambar 4.7 Halaman Depan Sistem K-Means Clustering	85
Gambar 4.8 Halaman Bantuan Sistem K-Means Clustering	86
Gambar 4.9 Halaman Aplikasi K-Means Clustering	87
Gambar 4.10 Halaman Berhasil Import File excel	88
Gambar 4.11 Halaman Transformation Data (1)	89
Gambar 4.13 Halaman Dataset (1)	90
Gambar 4.15 Menentukan CentroidRandom.....	92
Gambar 4.16 Halaman Proses K-Means Clustering (1)	93
Gambar 4.17 Halaman Proses K-Means Clustering (2)	94
Gambar 4.18 Halaman Proses K-Means Clustering (3)	95
Gambar 4.19 Halaman Hasil Proses K-Means (cluster satu)	96
Gambar 4.20 Halaman Hasil Proses K-Means (cluster dua)	98
Gambar 4.21 Halaman Hasil Proses K-Means (cluster tiga)	100

DAFTAR RUMUS

Rumus 2.1 Rumus Posisi Centroid.....	24
Rumus 3.1 Rumus Mencari Jarak Data ke Setiap Cluster	36



DAFTAR LAMPIRAN

LAMPIRAN B Tabel Dataset Mahasiswa	
LAMPIRAN C Tabel Perhitungan Jarak Data ke Setiap Cluster (M.Excel)	
LAMPIRAN D Tabel Clustering Data (M.Excel)	
LAMPIRAN E Source Code PHP (index.PHP)	
LAMPIRAN F Source Code PHP (import.PHP)	
LAMPIRAN G Source Code PHP (frek.PHP)	
LAMPIRAN H Source Code PHP (dataset.PHP)	
LAMPIRAN I Source Code PHP (proses.PHP)	
LAMPIRAN J Source Code PHP (hasil.PHP)	
LAMPIRAN K Source Code PHP (function.PHP)	



DAFTAR SINGKATAN

AHC	: Algorithm Hierarchical Clustering
NIM	: Nomer Induk Mahasiswa
KDD	: Knowledge Discovery in Database
IQ	: Intelligence Quotient
SQL	: Structured Query Language
NASDAQ	: National Association of Securities Dealers Automated Quotations
URL	: Uniform Resource Locator
HTML	: Hyper Text Markup Language
HTTP	: Hyper Text Transfer Protocol
PHP	: Hypertext Preprocessor
DFD	: Data Flow Diagram
ERD	: Entity Relationship Diagram
UIN	: Universitas Islam Negeri

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining merupakan proses untuk mendapatkan informasi dari basis data yang besar dengan cara penggalian data dengan memanfaatkan kumpulan data. Dengan kemajuan teknologi informasi sekarang sudah semakin berkembang pesat dan hampir mencakup di segala bidang. Kemajuan teknologi tersebut menghasilkan data yang besar dan banyak mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai data dari siswa yang mengikuti pendidikan

Misalnya dalam proses penerimaan mahasiswa baru sebuah perguruan tinggi akan menghasilkan data baru berupa profil dari mahasiswa baru. Hal seperti ini akan terjadi secara berulang pada sebuah perguruan tinggi. Jika dilakukan pengolahan data pada sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu menentukan strategi promosi penerimaan mahasiswa baru tahun berikutnya.

Pengolahan data tersebut dapat dilakukan menggunakan metode Data Mining dengan algoritma K-Means Clustering. Hasil dari pengolahan data tersebut akan sangat membantu pihak prodi atau fakultas. Dari penelitian ini diharapkan dapat diketahui apakah dari data masuk setiap angkatan mahasiswa

didapatkan informasi sebaran mahasiswa. Sehingga diketahui calon promosi kedepannya akan dilakukan di wilayah mana yang menjadi prioritas wilayah dari mahasiswa terbanyak.

Atribut yang dibutuhkan dalam pengolahan adalah kota asal mahasiswa, asal sekolah mahasiswa, dan IPK mahasiswa. Berdasarkan atribut tersebut selanjutnya data akan dikelompokkan menjadi mahasiswa berdasarkan asal kota, asal sekolah, dan IPK mahasiswa menggunakan algoritma K-Means Clustering.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini adalah:

1. Bagaimana menerapkan metode Data Mining dengan algoritma K-Means Clustering ke dalam aplikasi untuk menentukan wilayah yang tepat untuk promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta?
2. Bagaimana sebaran wilayah berdasarkan data mahasiswa baru sesuai dengan indeks prestasi akademik mahasiswa?
3. Bagaimana strategi promosi yang tepat untuk cluster yang terbentuk?

1.3 Batasan Masalah

Berdasarkan identifikasi masalah yang telah diuraikan di atas, maka permasalahan dibatasi pada :

1. Data yang digunakan adalah data mahasiswa Universitas Islam Negeri Sunan Kalijaga Yogyakarta Fakultas Sain dan Teknologi Prodi Teknik Informatika angkatan 2009 sampai dengan angkatan 2016.
2. Algoritma yang digunakan adalah Algoritma K-Means Clustering.
3. Atribut atau parameter data yang digunakan adalah asal kota dan IPK semester pertama pada Prodi Teknik Informatika.
4. Menggunakan bahasa pemrograman PHP dan database MySQL.

1.4 Tujuan Penelitian

Sesuai dengan masalah yang telah dirumuskan, maka tujuan dari penelitian ini untuk mengetahui :

1. Menerapkan metode Data Mining dengan algoritma K-Means Clustering ke dalam aplikasi untuk menentukan wilayah yang tepat untuk promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Mengetahui sebaran wilayah berdasarkan data mahasiswa angkatan 2009 sampai angkatan 2016.
3. Menentukan strategi promosi yang tepat untuk cluster yang terbentuk.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut :

1. Membantu pihak Prodi Teknik Informatika dengan aplikasi tersebut untuk memudahkan proses menentukan strategi promosi supaya tepat sasaran.

2. Membantu pihak Prodi Teknik Informatika untuk memutuskan langkah selanjutnya untuk menentukan wilayah mana yang di prioritaskan untuk dilakukannya promosi.
3. Membantu pihak Prodi Teknik Informatika sebagai bahan pertimbangan untuk menentukan langkah selanjutnya dalam penerimaan mahasiswa baru.

1.6 Keaslian Penelitian

Penelitian tentang penerapan algoritma K-Means clustering untuk menentukan strategi promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta ini sejauh pengetahuan penulis belum pernah dilakukan sebelumnya. Model penelitian ini menerapkan metode Data Mining algoritma K-Means Clustering untuk menganalisis persebaran kota asal dan asal sekolah mahasiswa berdasarkan IPK mahasiswa.

1.7 Sistematika Penulisan

Penyusunan laporan ini mengkaji beberapa bab bahasan dalam pengerjaannya. Sistematika penulisan dalam penyusunan laporan ini dimulai dari BAB I sampai BAB V.

BAB I PENDAHULUAN

Pada bab ini dijelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat penelitian, keaslian penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Pada bab ini dijelaskan tentang tinjauan pustaka dan landasan teori yang berhubungan dengan penelitian ini.

BAB III METODE PENELITIAN

Pada bab ini dijelaskan tentang objek, rincian perangkat yang digunakan serta metode penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini dimuat hasil analisis penelitian dan pembahasan mengenai penerapan metode yang digunakan.

BAB V KESIMPULAN

Bab ini berisikan kesimpulan dari hasil analisis serta saran untuk perbaikan penelitian selanjutnya.



BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Dalam penelitian yang dilakukan sebelumnya (Johan Oscar Ong, 2013), dengan judul Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University, penelitian ini bertujuan untuk mengetahui strategi promosi yang dapat dilakukan oleh pihak marketing President University. Sehingga diketahui persebaran daerah yang akan dilakukan promosi untuk meningkatkan strategi dari tahun ke tahun.

Penelitian selanjutnya (Asroni, 2015), dengan judul Penerapan Metode K-Means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang, bertujuan untuk mengelompokkan mahasiswa berdasarkan IPK dan beberapa atribut mata kuliah. Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan cluster yang terbentuk adalah dengan mengirim tim admisi UMM yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi berdasarkan potensi akademik mahasiswa dengan melakukan penyelarasan menggunakan promotion mix dan dengan melihat rata-rata IPK pada setiap cluster..

Penelitian selanjutnya (Fina Nasari, 2015), dengan judul Penerapan K-Means Clustering pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama), bertujuan untuk mengetahui jurusan yang

diminati jika asal sekolah adalah SMA atau Sekolah Menengah Pertama maka rata-rata jurusan yang diambil adalah Sistem Informasi dan jika asal Sekolahnya adalah SMK, rata-rata jurusan yang diambil adalah Teknik Informatika. Dari data tersebut maka diketahui jurusan apa yang nantinya akan diambil oleh mahasiswa baru di tahun depan.

Dalam penelitian yang dilakukan sebelumnya (Andi Wijanarko, 2016) dengan judul Penerapan Data Mining Untuk Menentukan Strategi Promosi Universitas PGRI Yogyakarta Menggunakan Algoritma K-means Clustering. Penelitian ini melakukan strategi untuk promosi Universitas PGRI Yogyakarta dengan menggunakan K-means clustering agar mengetahui persebaran wilayah berdasarkan indeks prestasi mahasiswa.

Penelitian selanjutnya (Nopal Febrian , 2017) dengan judul “Penerapan Data Mining Untuk Menentukan Pengelompokan Uang Kuliah Tunggal (UKT) UIN Sunan Kalijaga Menggunakan Algoritma K-Means Clustering.” Penelitian ini bertujuan untuk melakukan pola penentuan kelompok uang kuliah tunggal (UKT) dengan menggunakan algoritma k-means clustering di Universitas Islam Negeri (UIN) Sunan Kalijaga, untuk memahami tahapan tahapan pengclusteran data dengan menggunakan algoritma k-means clustering dan di implementasikan untuk pengelompokan uang kuliah tunggal.

Dalam penelitian yang dilakukan sebelumnya (Akhmad Choerudin Wakhid, 2017) dengan judul Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Pemilihan Program

Studi IAIN Salatiga. Penelitian ini bertujuan untuk mencari pola pemilihan program srtudi yang akan dilakukan pada mahasiswa IAIN Salatiga.

Tabel 2.1. Tabel Penelitian Terdahulu

No	Penelitian	Objek	Hasil
1	Johan Oscar Ong Program Studi Teknik Industri dari President University	Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing PRESIDENT UNIVERSITY	Dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa ada tiga strategi promosi yang dapat dilakukan oleh pihak marketing Preisdent Univeristy, yaitu: 1. Melakukan promosi dengan mengirim tim marketing yang sesuai dengan jurusan yang paling banyak diminati. 2. Melakukan promosi pada kota-kota di Indonesia yang didasarkan pada tingkat kemampuan akademik dari calon mahasiswa.
2	Rima Dias Ramadhani dari Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro	Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi UNIVERSITAS DIAN NUSWANTORO	Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan cluster yang terbentuk adalah dengan mengirim tim admisi UDINUS yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi

			berdasarkan potensi akademik mahasiswa dengan melakukan penyesuaian menggunakan promotion mix dan dengan melihat rata-rata IPK pada setiap cluster.
3	Fina Nasari dari Sistem Informasi Universitas Potensi Utama	PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA)	Berdasarkan dari hasil Cluster kesimpulan yang dapat diambil adalah bahwa jika asal sekolah adalah SMA atau Sekolah Menengah Pertama maka rata-rata jurusan yang diambil adalah Sistem Informasi dan jika asal Sekolahnya adalah SMK rata-rata jurusan yang diambil adalah Teknik Informatika.
4	Andi wijanarko dari UIN Sunan Kalijaga	Penerapan Data Mining Untuk Menentukan Strategi Promosi Universitas PGRI Yogyakarta Menggunakan Algoritma K-Means Clustering	Melakukan strategi untuk promosi Universitas PGRI Yogyakarta dengan Menggunakan K-Means clustering untuk mengetahui persebaran wilayah berdasarkan indeks prestasi mahasiswa.
5	Nopal Febriyan	Penerapan Data Mining Untuk	melakukan pola pemilihan terhadap

		Mengetahui Pengelompokan Uang Kulilah Tunggal UIN Sunan Kalijaga Menggunakan Algoritma K-Means clustering.	data mahasiswa Universitas Islam Negeri (UIN) Sunan Kalijaga.
6	Akhmad Choerudin Wakhid	Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Pemilihan Program Studi IAIN Salatiga.	Penelitian ini bertujuan untuk mencari pola pemilihan program srtudi yang akan dilakukan pada mahasiswa IAIN Salatiga.

2.2 Landasan Teori

Dalam landasan teori ini akan dijelaskan secara garis besar teori-teori yang menjadi dasar atau acuan dalam penelitian ini :

2.2.1 Data Mining

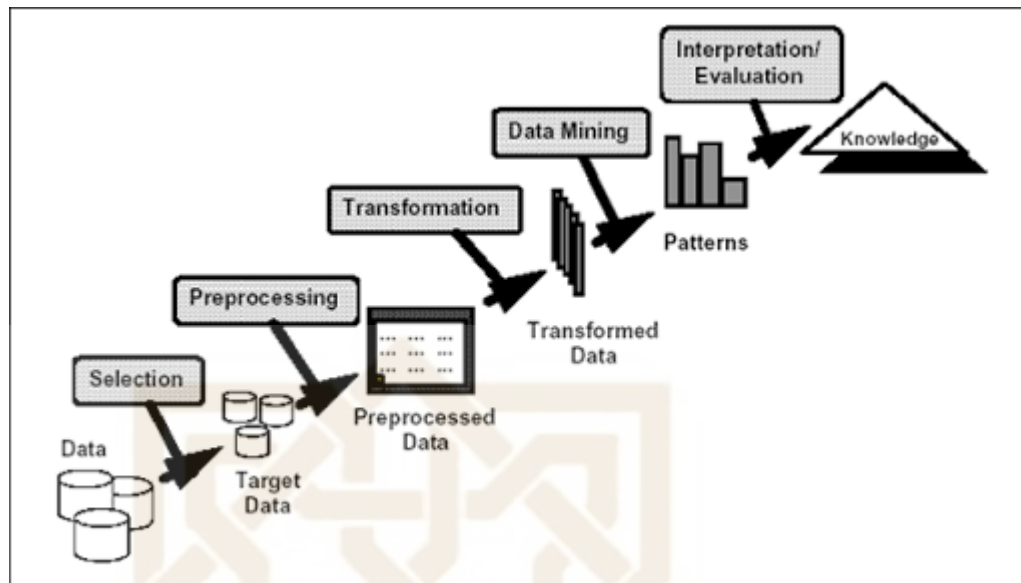
Alasan utama Data Mining sangat menarik perhatian industri informasi dalam beberapa tahun belakangan ini karena tersedianya data dalam jumlah yang sangat besar dan semakin besarnya kebutuhan untuk mengubah data tersebut menjadi informasi dan pengetahuan yang berguna dan bermanfaat. Data Mining adalah kegiatan mengekstrasi atau menambang pengetahuan dari data yang berukuran dan berjumlah cukup

besar, informasi-informasi inilah yang nantinya sangat berguna untuk pengembangan (Fadli, 2011).

Data Mining berisi pencarian *trend* atau pola yang diinginkan dalam database yang besar untuk membantu pengambilan keputusan di waktu yang akan datang (Herawati, 2013). Data Mining bisa digunakan oleh perusahaan besar untuk mendapatkan informasi yang dapat menunjang dan meningkatkan proses bisnis perusahaan tersebut. Data Mining mengeksplorasi basis data untuk menemukan pola-pola tersembunyi untuk mencari informasi dan prediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka (Moertini, 2002).

Data Mining terkadang dianggap sama dengan istilah KDD (Knowledge Discovery in Database). Namun, Data Mining adalah salah satu bagian dari KDD. Dibandingkan dengan KDD, Data Mining lebih dikenal di kalangan pelaku bisnis. Sebagai komponen dalam KDD, Data Mining berkaitan dengan ekstraksi dan perhitungan pola-pola yang telah dianalisis.

Tahapan dalam proses KDD dijelaskan di bawah ini (Luthfi, 2009) :



Gambar 2.1 Tahapan Proses KDD

1. *Data Selection*

- Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (*discovery*) akan dilakukan.
- Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses Data Mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/ Cleaning*

- Pemrosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan *noise*.
- Sebelum proses Data Mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD.

- Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).
- Dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. Transformation

- Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.
- Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses Data Mining.
- Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data

4. Data Mining

- Pemilihan tugas Data Mining; pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll.
- Pemilihan algoritma Data Mining untuk pencarian (searching).
- Proses Data Mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam Data Mining sangat bervariasi.
- Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/ Evaluation*

- Penerjemahan pola-pola yang dihasilkan dari Data Mining.
- Pola informasi yang dihasilkan dari proses Data Mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.
- Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Definisi Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu Data Mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan database. Data Mining adalah proses menerapkan metode ini untuk data dengan maksud untuk mengungkap pola-pola tersembunyi. Dengan arti lain Data Mining adalah proses untuk penggalian pola-pola dari data. Data Mining menjadi alat yang semakin penting untuk mengubah data tersebut menjadi informasi. Hal ini sering digunakan dalam berbagai praktek profil, seperti pemasaran,

pengawasan, penipuan deteksi dan penemuan ilmiah. Telah digunakan selama bertahun-tahun oleh bisnis, ilmuwan dan pemerintah untuk menyaring volume data seperti catatan perjalanan penumpang penerbangan, data sensus dan supermarket *scannerdata* untuk menghasilkan laporan riset pasar.

Alasan utama untuk menggunakan Data Mining adalah untuk membantu dalam analisis koleksi pengamatan perilaku. Data tersebut rentan terhadap *collinearity* karena diketahui keterkaitan. Fakta yang tak terelakkan pada Data Mining adalah bahwa subset/set data yang dianalisis mungkin tidak mewakili seluruh domain, dan karenanya tidak boleh berisi contoh-contoh hubungan kritis tertentu dan perilaku yang ada di bagian lain dari domain . Untuk mengatasi masalah semacam ini, analisis dapat ditambah menggunakan berbasis percobaan dan pendekatan lain, seperti *Choice Modelling* untuk data yang dihasilkan manusia.

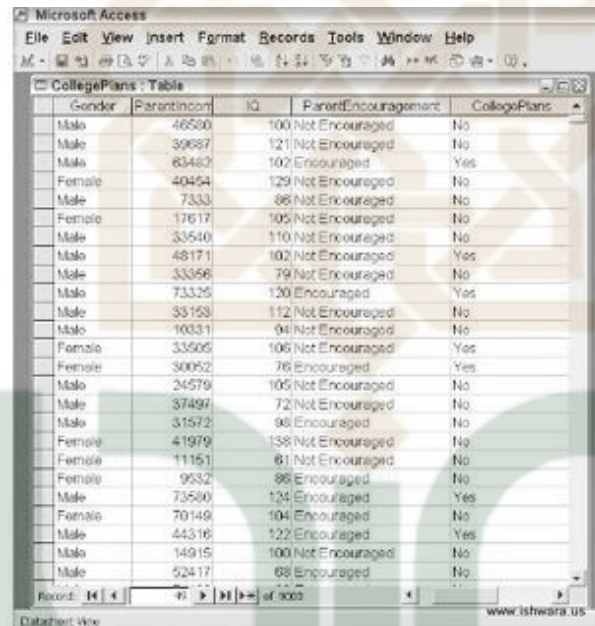
Beberapa teknik yang sering disebut-sebut dalam literatur Data Mining dalam penerapannya antara lain: *clustering*, *classification*, *association rule mining*, *neural network*, *genetic algorithm*, dan lain-lain. Yang membedakan persepsi terhadap Data Mining adalah perkembangan teknik-teknik Data Mining untuk aplikasi pada database skala besar. Sebelum populernya Data Mining, teknik-teknik tersebut hanya dapat dipakai untuk data skala kecil saja.

2.2.2 Teknik – Teknik Data Mining

1. Classification

Classification adalah tindakan untuk memberikan kelompok pada setiap keadaan. Setiap keadaan berisi sekelompok atribut, salah satunya adalah *class attribute*. Metode ini untuk menemukan sebuah model yang dapat menjelaskan *class attribute* itu sebagai fungsi dari input attribute.

Contohnya kita lihat pada contoh kasus College Plan



Gender	ParentIncome	IQ	ParentEncouragement	CollegePlans
Male	46500	100	Not Encouraged	No
Male	39687	121	Not Encouraged	No
Male	63462	102	Encouraged	Yes
Female	40454	129	Not Encouraged	No
Male	7333	86	Not Encouraged	No
Female	17617	105	Not Encouraged	No
Male	33540	110	Not Encouraged	No
Male	48171	102	Not Encouraged	Yes
Male	33366	79	Not Encouraged	No
Male	73326	120	Encouraged	Yes
Male	33153	112	Not Encouraged	No
Male	10331	94	Not Encouraged	No
Female	33505	106	Not Encouraged	Yes
Female	30052	76	Encouraged	Yes
Male	24579	105	Not Encouraged	No
Male	37497	72	Not Encouraged	No
Male	31572	98	Encouraged	No
Female	41979	138	Not Encouraged	No
Female	11151	81	Not Encouraged	No
Female	9532	86	Encouraged	No
Male	73500	124	Encouraged	Yes
Female	70149	104	Encouraged	No
Male	44316	122	Encouraged	Yes
Male	14915	100	Not Encouraged	No
Male	52417	68	Encouraged	No

Gambar 2.2 Classification

Class adalah *attribute* *CollegePlans* yang berisi dua pernyataan, Yes dan No, perhatikan ini. Sebuah *Classification Model* akan menggunakan atribut lain dari kasus tersebut (input attribute; yaitu kolom IQ, Gender, ParentIncome, dan ParentEncouragement) untuk dapat menentukan pola (pattern) class (Output Attribute; yaitu Kolom

CollegePlans yang berisi Yes atau No). Algoritma Data Mining yang membutuhkan variabel target untuk belajar (sampai mendapatkan rule / pola yang berlaku pada data tersebut) kita standarkan dengan sebutan dengan *Supervised Algorithm*. Yang termasuk kepada *Classification Algorithm* adalah *Decision Trees*, *Neural Network* dan *Naives Bayes*.

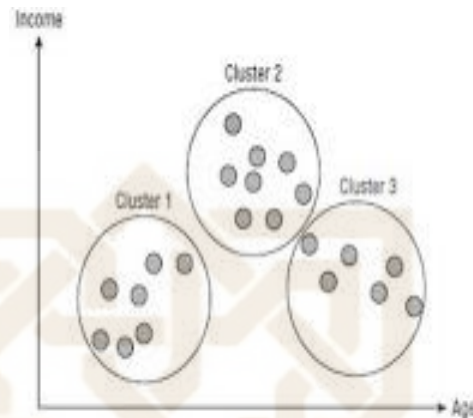
2. *Clustering*

Clustering juga disebut sebagai segmentation. Metode ini digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut, mengelompokkan data yang memiliki kemiripan atribut. Gambar ini menunjukkan kelompok data pelanggan sederhana yang berisi dua atribut, yaitu *Age* (Umur) dan *Income* (Pendapatan). Algoritma Clustering mengelompokkan kelompok data kedalam tiga segment berdasarkan kedua atribut ini.

- Cluster 1 berisi populasi berusia muda dengan pendapatan rendah
- Cluster 2 berisi populasi berusia menengah dengan pendapatan yang lebih tinggi
- Cluster 3 berisi populasi berusia tua dengan pendapatan yang relatif rendah.

Clustering adalah metode Data Mining yang *Unsupervised*, karena tidak ada satu atribut yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama. Kebanyakan Algoritma Clustering membangun sebuah model melalui

serangkaian pengulangan dan berhenti ketika model tersebut telah memusat atau berkumpul (batasan dari segmentasi ini telah stabil).



Gambar 2.3 Clustering

3. Association

Association juga disebut sebagai *Market Basket Analysis*. Sebuah problem bisnis yang khas adalah menganalisa tabel transaksi penjualan dan mengidentifikasi produk-produk yang seringkali dibeli bersamaan oleh customer, misalnya apabila orang membeli sambal, biasanya juga dia membeli kecap. Kesamaan yang ada dari data pembelian digunakan untuk mengidentifikasi kelompok kesamaan dari produk dan kebiasaan apa yang terjadi guna kepentingan *cross-selling*.

Beberapa hal dapat kita baca, misalnya :

- Ketika orang membeli susu, dia biasanya membeli keju
- Ketika orang membeli pepsi atau coke, biasanya dia membeli juice
- Di dalam istilah association, setiap item dipertimbangkan sebagai informasi.

Metode *association* memiliki dua tujuan:

- Untuk mencari produk apa yang biasanya terjual bersamaan
- Untuk mencari tahu apa aturan yang menyebabkan kesamaan tersebut.

4. *Regression*

Metode *Regression* mirip dengan metode *Classification*, yang membedakannya adalah metode *regression* tidak bisa mencari pola yang dijabarkan sebagai *class* (kelas). Metoda *regression* bertujuan untuk mencari pola dan menentukan sebuah nilai numerik.

Sebuah Teknik *Linear Line-fitting* sederhana adalah sebuah contoh dari *Regression*, dimana hasilnya adalah sebuah fungsi untuk menentukan hasil yang berdasarkan nilai dari input. Bentuk yang lebih canggih dari *regression* sudah mendukung input berupa kategori, jadi tidak hanya input berupa numerik. Teknik paling populer yang digunakan untuk *regression* adalah *linear regression* dan *logistic regression*. Teknik lain yang didukung oleh SQL Server Data Mining adalah *Regression Trees* (bagian dari dari algoritma *Microsoft Decission Trees*) dan *Neural Network*.

Regression digunakan untuk memecahkan banyak problem bisnis, contohnya untuk memperkirakan metode distribusi, kapasitas distribusi, musim dan untuk memperkirakan kecepatan angin berdasarkan temperatur, tekanan udara, dan kelembaban.

5. *Forecasting*

Forecasting adalah metode Data Mining yang sangat penting.

Contohnya digunakan untuk menjawab pertanyaan seperti berikut:

- Seperti apa jadinya nilai saham dari Microsoft Corporation (pada NASDAQ, disimbolkan sebagai MSFT) pada keesokan hari?
- Sebanyak apa penjualan produk tertentu pada bulan depan?



Gambar 2.4 Grafik 3-year price history

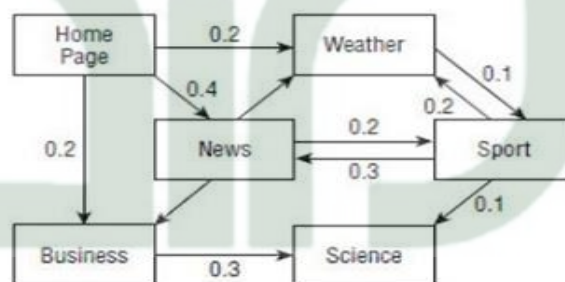
Teknik *Forecasting* dapat membantu menjawab pertanyaan-pertanyaan di atas. Sebagai inputnya teknik *Forecasting* akan mengambil sederetan angka yang menunjukkan nilai yang berjalan seiring waktu dan kemudian Teknik *Forecasting* ini akan menghubungkan nilai masa depan dengan menggunakan bermacam-macam teknik *machine-learning* dan teknik statistik yang berhubungan dengan musim, *trend*, dan *noise* pada data.

Gambar 2.4 menunjukkan dua kurva, garis yang tegas adalah *time-series* data sebenarnya dari nilai saham Microsoft, dan garis putus-putus adalah time series model yang memprediksi nilai saham berdasarkan nilai saham pada masa lalu.

6. *Sequence Analysis*

Sequence Analysis digunakan untuk mencari pola pada serangkaian kejadian yang disebut dengan *Sequence*. Sebagai contoh sebuah DNA terdiri dari rangkaian bagian: A, G, C, dan T. dan rangkaian klik pada sebuah website berisi rentetan URL. Pada kejadian nyata anda mungkin memodelkan pembelian oleh pelanggan sebagai sebuah *sequence* (rangkaiannya) data, rangkaian tersebut adalah:

- Pertama-tama Seorang pelanggan membeli komputer
- kemudian membeli speaker
- dan akhirnya membeli sebuah webcam.



Gambar 2.5 *Sequence*

Baik *Sequence* maupun time-series data mempunyai kemiripan, mereka sama sama berisi tinjauan berdekatan yang urutannya

bergantung. Bedanya adalah sebuah time-series berisi data bertipe numerik, dan sebuah *sequence series* berisi bagian yang khas.

Gambar 2.5 menunjukkan rangkaian klik pada sebuah website berita. Setiap node adalah sebuah kategori URL. Dan garis melambangkan transisi antar kategori URL tersebut. Setiap transisi di kelompokkan dengan sebuah bobot yang menggambarkan kemungkinan transisi antara satu URL dan URL yang lain.

2.2.3 Algoritma K-Means Clustering

K-Means merupakan salah satu algoritma *clustering*. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan *supervised learning* yang menerima masukan berupa vektor $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$, di mana x_i merupakan data dari suatu data pelatihan dan y_i merupakan label kelas untuk x_i . Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut.

K-Means ditemukan oleh beberapa orang yaitu Lloyd (1957, 1982), Forgey (1965), Friedman and Rubin (1967), and McQueen (1967). Ide

dari *clustering* pertama kali ditemukan oleh Lloyd pada tahun 1957, namun hal tersebut baru dipublikasi pada tahun 1982. Pada tahun 1965, Forgey juga mempublikasi teknik yang sama sehingga terkadang dikenal sebagai Lloyd-Forgey pada beberapa sumber.

K-Means merupakan metode klusterisasi yang paling terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengkluster data yang besar, mampu menangani *data outlier*, dan kompleksitas waktunya linear $O(nKT)$ dengan n adalah jumlah dokumen, K adalah jumlah kluster, dan T adalah jumlah iterasi. K-Means merupakan metode pengklusteran secara *partitioning* yang memisahkan data ke dalam kelompok yang berbeda. Dengan *partitioning* secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke klusternya. Metode ini dikembangkan oleh Mac Queen pada tahun 1967.

Dasar algoritma K-Means adalah sebagai berikut :

- Tentukan nilai k sebagai jumlah kluster yang ingin dibentuk
- Bangkitkan k *centroid* (titik pusat kluster) awal secara random
- Hitung jarak setiap data ke masing-masing *centroid* menggunakan Rumus korelasi antar dua objek yaitu *Euclidean Distance* dan kesamaan *Cosine*
- Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya*

- Tentukan posisi *centroid* baru (C_k) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

$$c_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Rumus 2.1 Rumus posisi *centroid*

Dimana n_k adalah jumlah dokumen dalam cluster k dan d_i adalah dokumen dalam cluster k. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Adapun karakteristik dari algoritma K-Means salah satunya adalah sangat sensitif dalam penentuan titik pusat awal kluster karena K-Means membangkitkan titik pusat kluster awal secara random. Pada saat pembangkitan awal titik pusat yang random tersebut mendekati solusi akhir pusat kluster, K-Means mempunyai kemungkinan yang tinggi untuk menemukan titik pusat kluster yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat kluster, maka besar kemungkinan ini menyebabkan hasil pengklasteran yang tidak tepat. Akibatnya K-Means tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode K-Means sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma K-Means hanya bisa digunakan untuk data yang atributnya bernilai numerik.

2.2.4 Konsep Aplikasi Berbasis Web

Penjelasan tentang teori dari bahasa pemrograman yang nantinya digunakan dalam penelitian ini.

1. HTML (*Hyper Text Markup Language*)

Merupakan standar dipakai pada halaman web. HTML bekerja dengan menggunakan HTTP (*Hyper Text Transfer Protocol*), yaitu *protocol* komunikasi yang memungkinkan *web server* berkomunikasi dengan *web browser* (Kadir, 2005).

2. PHP (*Hypertext Preprocessor*)

Singkatan dari PHP *Hypertext Preprocessor* yang digunakan sebagai bahasa script *server side* dalam pengembangan Web yang disisipkan pada dokumen HTML. Berbeda dengan HTML yang hanya bisa menampilkan konten statis, PHP bisa berinteraksi dengan database, file dan folder, sehingga membuat PHP bisa menampilkan konten yang dinamis dari sebuah website (Linggar, 2004).

2.2.5 Definisi Sistem

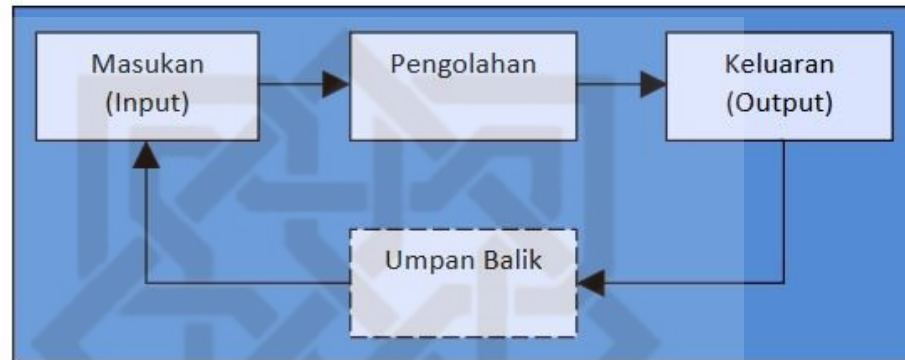
Pendefinisian sistem dapat dilakukan dalam dua pendekatan, yaitu pendekatan atas fisik dan pendekatan atas fungsi.

1. Ditinjau dari Aspek Fisik

Sistem sebagai seperangkat elemen yang digabungkan satu dengan lainnya untuk suatu tujuan bersama. Dalam kamus *Webster's*

Unbrigid, sistem adalah elemen-elemen yang saling berhubungan membentuk satu kesatuan atau organisasi (Murdick dan Ross,1993).

Sistem terdiri dari unsur-unsur seperti masukan (input), pengolahan (processing), serta keluaran (output),(Scott,1996).



Gambar 2.6 Model Sistem Menurut Scott

2. Ditinjau dari Aspek Fungsi

Sistem adalah jaringan kerja dari prosedur-prosedur yang saling berhubungan dan dikembangkan sesuai dengan skematis yang terintegrasi dalam melaksanakan suatu aktivitas utama di dalam bisnis. Sementara prosedur diartikan suatu urutan operasi klerikal (tulis-menulis), yang melibatkan beberapa orang di dalam satu atau lebih departemen, dan diterapkan untuk menjamin penanganan yang seragam dari transaksi-transaksi bisnis yang terjadi.

Sistem adalah jaringan kerja dari prosedur-prosedur yang saling berhubungan dan berkumpul bersama-sama dalam melakukan aktivitas atau untuk menyelesaikan suatu sasaran tertentu.

Sistem sebagai sekelompok elemen-elemen yang terintegrasi dengan maksud yang sama untuk mencapai suatu tujuan. Sumber daya mengalir dari elemen output dan untuk menjamin prosesnya berjalan dengan baik maka dihubungkan mekanisme kontrol (McLeod.1995).

2.2.6 World Wide Web

Untuk menerjemahkan dokumen *hypertext* ke dalam bentuk dokumen yang dapat dipahami oleh manusia, maka *web browser* melalui *web client* akan membaca halaman web yang tersimpan di sebuah *web server* melalui protokol yang sering disebut HTTP atau *Hypertext Transfer Protocol*





Perangkat lunak *web browser* saat ini tersedia dalam berbagai produk dengan kelebihan dan kekurangannya masing-masing. Beberapa *web browser* yang paling terkenal saat ini adalah *Mozilla Firefox, Internet Explorer, Opera, Google Chrome, Safari*, dan lain sebagainya.

Dokumen *hypertext* yang tersimpan di dalam sebuah *web server* memiliki banyak link atau sambungan antara satu dokumen *hypertext* dengan dokumen *hypertext* lainnya. Dengan adanya *link* antar dokumen *hypertext*, maka hal tersebut dapat memudahkan pengakses untuk mengunjungi atau mendapatkan dokumen-dokumen terkait yang diinginkannya.

2.2.7 Data Flow Diagram (DFD)

Data Flow Diagram (DFD) adalah sebuah teknik grafis untuk menggambarkan aliran informasi dan transformasi yang diaplikasikan pada saat data bergerak dari input menjadi output (Pressman, 2002).

Berikut adalah notasi – notasi pada DFD:

NOTASI	KETERANGAN
	Himpunan Entitas Luar Himpunan entitas luar yaitu sebuah elemen sistem (misalnya perangkat keras, seseorang, program yang lain) yang menghasilkan informasi bagi transformasi oleh perangkat lunak atau menerima informasi yang dihasilkan oleh perangkat lunak.
	Aliran Data Menggambarkan aliran data dari satu proses ke proses lainnya. Anak panah menunjukkan arah aliran data.
	Proses Proses atau fungsi yang mentransformasikan data secara umum.
	Penyimpanan Data Penyimpanan data berfungsi untuk menyimpan informasi yang digunakan oleh perangkat lunak.

Gambar 2.7 Notasi Dasar Data *Flow Diagram* (DFD)

2.2.8 Entity-Relationship Diagram (ERD)

Entity-Relationship Diagram (ERD) merupakan suatu model jaringan yang menggunakan susunan data yang disimpan pada sistem secara abstrak. ERD juga menggambarkan hubungan antara satu himpunan entitas yang memiliki atribut dengan himpunan entitas yang lain dalam suatu sistem yang terintegrasi (Yakub, 2008).


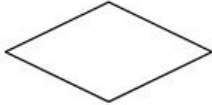


Entity-Relationship Diagram terdiri dari beberapa notasi yaitu Entitas, Atribut, dan Relasi. Pengertian dan penjelasan dari ketiga notasi tersebut adalah sebagai berikut :

Suatu entitas merupakan individu yang mewakili sesuatu yang nyata eksistensinya dan dapat dibedakan dari objek-objek yang lain. Suatu entitas memiliki sekumpulan sifat, dan nilai dari beberapa sifat tersebut dapat mengidentifikasi entitas tersebut. Sekumpulan entitas yang mempunyai tipe yang sama (sejenis) dan berada dalam lingkup yang sama membentuk suatu himpunan entitas (Fathansyah, 1999).

Atribut menentukan properti suatu objek data. Atribut dapat digunakan untuk menamai sebuah contoh dari entitas, menggambarkan contoh, dan membuat referensi ke contoh yang lain pada tabel yang lain. Satu atribut atau lebih harus didefinisikan sebagai sebuah pengidentifikasi, yang mana atribut pengidentifikasi akan menjadi sebuah kunci untuk menemukan sebuah contoh dari entitas (Pressman, 2002).

Relasi menunjukkan adanya hubungan di antara sejumlah entitas yang berasal dari sejumlah himpunan entitas yang berbeda. Kumpulan semua relasi di antara entitas-entitas yang terdapat pada himpunan entitas membentuk suatu himpunan relasi (Pressman, 2002).

Berikut adalah notasi-notasi pada *Entity-Relationship Diagram* (ERD) :

NOTASI	KETERANGAN
	Himpunan entitas , yaitu kumpulan dari objek yang dapat diidentifikasi secara unik.
	Relasi, yaitu hubungan yang terjadi antara satu atau lebih entitas. Jenis hubungan antara lain; satu ke satu, satu ke banyak, dan banyak ke banyak.
	Atribut, yaitu karakteristik dari <i>entity</i> atau relasi yang merupakan penjelasan detail tentang entitas.
	Garis, hubungan antara <i>entity</i> dengan atributnya dan himpunan entitas dengan himpunan relasinya.

Gambar 2.8 Simbol-simbol *Entity Relationship Diagram* (ERD)

2.2.9 Pengolahan Data *Microsoft Excel*

Microsoft Excel adalah sebuah aplikasi (perangkat lunak) yang merupakan bagian dari paket *Software Microsoft Office*. Perangkat lunak ini berjenis *spreadsheet*. Pada awal dibuatnya *Microsoft Excel* dapat berdiri sendiri dan bukan menjadi bagian dari *Microsoft Office*. Namun sekarang ini *Microsoft Excel* sudah merupakan satu paket dengan *Microsoft Office* yang terdiri dari berbagai perangkat lunak yang merupakan kebutuhan kantor dewasa ini. Aplikasi yang berbentuk lembar kerja ini dibuat dan didistribusikan oleh *Microsoft Corporation* untuk pengguna *Apple Mac OS* dan *Microsoft Windows*. Dalam aplikasi ini tersedia fitur pembuatan grafik dan fitur kalkulasi yang sifatnya agresif dan progresif.

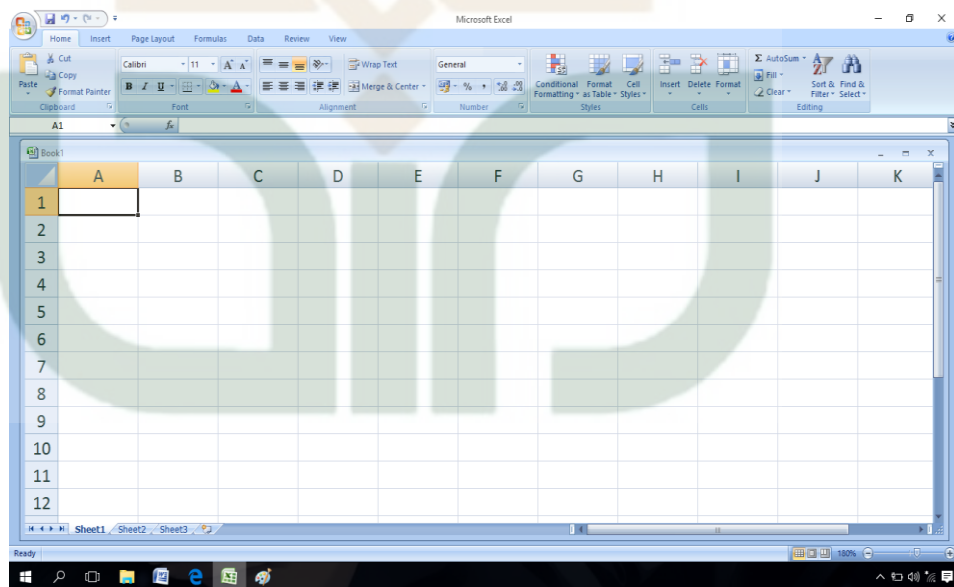
Microsoft Excel merupakan aplikasi untuk mengolah data secara otomatis yang, rumus, pemakaian fungsi-fungsi, pengolahan data dan

tabel, pembuatan grafik dan manajemen data. dapat berupa perhitungan dasar

Pemakaian rumus sendiri dapat berupa penambahan, pengurangan, perkalian, dan lain sebagainya. Sedangkan pemakaian fungsi-fungsi dapat berupa pemakaian rumus yang bertujuan untuk menghitung dalam bentuk rumus matematika maupun non matematika.

Microsoft Excel dapat juga digunakan untuk menyelesaikan berbagai keperluan administrasi, dari yang sederhana sampai dengan yang rumit. Pada pemakaian keperluan yang sederhana tersebut misalkan untuk membuat perencanaan kebutuhan suatu perusahaan, berupa perencanaan barang kebutuhan, jumlah maupun harganya.

Berikut jendela tampilan *Microsoft Excel* 2010 seperti pada gambar 2.8:



Gambar 2.9 Tampilan *Microsoft Excel*

BAB III

METODE PENELITIAN

3.1 Metode Penelitian

Metode yang digunakan dalam penelitian tugas akhir ini adalah secara terapan, yaitu peneliti menerapkan metode Data Mining dengan algoritma K-Means Clustering pada data mahasiswa Teknik Informatika angkatan 2009 sampai dengan angkatan 2016 di Universitas Islam Negeri Sunan Kalijaga Yogyakarta untuk dianalisa dan dikelompokkan sesuai dengan persebaran wilayah dan sesuai asal sekolah berdasarkan indeks prestasi kumulatif selama dua semester awal yaitu pada semester 1 dan semester 2.

3.2 Perangkat Penelitian

Dalam melakukan penelitian ini, dibutuhkan peralatan perangkat keras (*Hardware*) dan perangkat lunak (*Software*). Untuk lebih lengkapnya di jelaskan berikut ini :

3.2.1 Perangkat Keras

Perangkat keras (*hardware*) yang digunakan dalam penelitian ini yaitu menggunakan 1 buah laptop yang digunakan selama penelitian. Spesifikasi laptop sebagai berikut : *ASUZ A46C* , *Intel CORE i3* CPU 1.60GHz, Ram 4gb, *Hardisk Storage* 320 gb, *Windows* 10 64bit.

3.2.2 Perangkat Lunak

Perangkat lunak memiliki peran penting pada peneilitian ini karena hasil dari analisis data dapat diketahui dari pengolahan menggunakan

perangkat lunak dalam mengetahui hasilnya. Pada penelitian ini peneliti menggunakan *Microsoft Access 2010* untuk menggabungkan atau mengintegrasikan semua data yang didapatkan, database MySQL, Sublime Text 2 dengan bahasa pemrograman PHP dan sebagai pembanding sistem yang akan dibuat yaitu *Microsoft Excel 2010*.

3.3 Tahap – Tahap Penelitian

Dalam penelitian ini, peneliti menggunakan salah satu metode dalam Data Mining yaitu metode K-Means Clustering. Dalam metode ini memiliki beberapa tahapan. Tahapan yang dilakukan adalah:

3.3.1 Pengumpulan Data

Untuk melakukan proses K-Means Clustering tentu membutuhkan sebuah data yang cukup banyak dan sesuai dengan yang dibutuhkan, di dalam penelitian ini peneliti menggunakan data mahasiswa jurusan Teknik Informatika angkatan 2009 sampai dengan angkatan 2016 di UIN Sunan Kalijaga Yogyakarta. Pengambilan data di UIN Sunan Kalijaga Yogyakarta cukup dengan melampirkan surat izin penelitian dari pihak UIN dan melampirkan proposal penelitian ke bagian akademik UIN Sunan Kalijaga Yogyakarta. Setelah mendapat balasan dari pihak UIN Sunan Kalijaga Yogyakarta, data bisa diambil ke bagian data kemahasiswaan dengan cara mengkopidata sesuai dengan kebutuhan yang sudah tertulis di dalam proposal.

3.3.2 *Pre-processing Data*

Tahap *pre-processing data* adalah tahap dimana data yang sudah didapatkan, dipilah, dan dipisahkan agar mendapatkan data yang dibutuhkan untuk proses selanjutnya. Tahapan ini mempunyai beberapa proses dimana setiap proses tersebut saling berhubungan satu sama lainnya. Proses dalam tahapan *preprocessing* adalah sebagai berikut:

1. *Data Reduction*

Data reduction adalah proses untuk mereduksi atau mengurangi dimensi, atribut, ataupun sejumlah data yang tidak dibutuhkan dalam suatu file data. *Data reduction* sangat berguna untuk mendapatkan atribut dan sejumlah data yang akan digunakan di dalam penelitian ini.

2. *Data Cleaning*

Data cleaning adalah proses dalam tahapan *preprocessing* untuk mengisi data kosong atau blank apabila memungkinkan, duplikasi data, memperbaiki data yang tidak sesuai dengan ketentuan atau salah ketik seperti kurang huruf dan kelebihan huruf, mengubah dan memodifikasi data agar data yang akan diolah adalah data yang konsisten, mengatur data yang kurang rapi dalam penulisan huruf besar dan kecil, dan mengganti format penulisan angka dan huruf sesuai dengan yang dibutuhkan.

3. *Data Transformation*

Karena metode K-Means Clustering adalah metode yang bisa dilakukan apabila data yang dipakai adalah data berupa angka, maka

proses transformation ini sangat dibutuhkan. Proses transformation adalah tahap untuk mengubah data atribut yang selain angka ke dalam nilai angka agar data tersebut dapat diolah menggunakan algoritma K-Means Clustering.

4. Data Integration

Data integration adalah suatu proses untuk menggabungkan atau mengintegrasikan data dari beberapa file sumber. Data integration hanya dilakukan jika data berasal dari tempat yang berbeda-beda (sumber data tidak hanya dari 1 tempat). Langkah yang dilakukan antara lain mengintegrasikan skema, mengidentifikasi masalah entitas, dan mendeteksi sekaligus menyelesaikan konflik pada nilai data.

3.3.3 K-Means Clustering

Setelah data melalui tahap *pre-processing* dan tahap *transformation* maka data sudah siap untuk diolah menggunakan metode K-Means Clustering. Metode K-Means Clustering adalah proses untuk mengelompokkan data ke dalam sebuah cluster dengan titik pusat yang berbeda beda setiap cluster. Dalam penelitian ini peneliti sudah menentukan untuk menggunakan 3 cluster. Proses K-Means Clustering tersebut meliputi 5 proses, yaitu :

1. Menentukan Titik Pusat Cluster

Menentukan titik pusat cluster adalah langkah awal untuk proses K-Means Clustering. Fungsi proses ini adalah untuk menentukan titik awal sebagai patokan untuk mencari jarak antara data ke cluster yang

sudah ditentukan. Titik awal pusat cluster disebut juga dengan *centroid*. Untuk menentukan titik pusat setiap cluster bisa dilakukan dengan mencari rata-rata dari data yang akan diolah ataupun sesuai keinginan.

2. Menghitung Jarak Data ke Setiap Cluster

Setelah menentukan titik pusat di setiap cluster proses selanjutnya adalah menghitung jarak antara data ke setiap cluster yang sudah dibentuk. Rumus untuk mencari jarak (*distance*) dari satu cluster adalah

$$\sqrt{(x_i - x_{avg})^2 + (y_i - y_{avg})^2 + (z_i - z_{avg})^2}$$

Rumus 3.1 Rumus Mencari Jarak Data ke Setiap Cluster

X_i : Data pertama (diambil dari atribut pertama)

X_{avg} : Titik pusat cluster / *centroid* untuk atribut pertama

Y_i : Data kedua (diambil dari atribut kedua)

Y_{avg} : Titik pusat cluster / *centroid* untuk atribut kedua

Z_i : Data ketiga (diambil dari atribut ketiga)

Z_{avg} : Titik pusat cluster / *centroid* untuk atribut pertama

Rumus tersebut adalah rumus untuk menentukan jarak dari satu baris data ke satu cluster tertentu. Dalam penelitian ini peneliti menggunakan 3 cluster, maka perhitungannya pun menjadi 3 kali dengan titik pusat cluster yang berbeda di setiap cluster.

3. Mengalokasikan Data kedalam Cluster

Setelah mendapatkan jarak antara setiap data ke setiap cluster yang terbentuk maka proses selanjutnya adalah clustering atau mengelompokkan dan mengalokasikan data ke dalam cluster. Untuk mengelompokkan data ke dalam cluster ini cukup dilihat dari jarak terdekat dari setiap cluster. Apabila jarak yang didapatkan dari suatu data adalah dengan urutan paling kecil di setiap cluster, maka data termasuk kedalam cluster pertama.

4. Menentukan Titik Pusat Cluster Baru

Setelah mengalokasikan data ke dalam cluster yang terbentuk maka proses selanjutnya adalah menentukan titik pusat cluster baru dengan cara yang sama yaitu mencari rata-rata di setiap atribut data. Tetapi dalam perhitungan kali ini sedikit berbeda dari yang pertama, apabila proses pertama adalah mencari rata-rata dari semua atribut data maka untuk menentukan titik pusat cluster baru ini data yang digunakan sesuai dengan clusternya masing-masing. Apabila data yang masuk ke cluster satu adalah data pertama, ketiga, kelima, dan keenam, maka rata-rata yang dicari hanya menggunakan data pertama, ketiga, kelima, dan keenam.

5. Memverifikasi Titik Pusat Cluster

Setelah mendapatkan titik pusat cluster baru maka proses selanjutnya adalah memverifikasi titik pusat cluster baru tersebut dengan titik pusat cluster yang lama. Apabila hasil titik pusat cluster

baru yang didapat sama dengan titik pusat cluster yang lama, maka proses K-Means sudah selesai dan hasil dari proses K-Means Clustering sudah didapatkan dan data yang diklasifikasi sudah tidak bisa berubah-ubah lagi. Tetapi jika hasil titik pusat cluster baru yang didapat berbeda dari titik pusat cluster yang lama, maka proses K-Means tetap dilanjutkan dan mulai lagi dari proses kedua atau menghitung jarak data ke setiap cluster.



BAB IV

HASIL DAN PEMBAHASAN

4.1 Objek Penelitian

Penelitian ini mengambil objek yaitu data mahasiswa jurusan Teknik Informatika angkatan 2009 sampai dengan angkatan 2016 beserta indeks prestasi kumulatif (IPK) mahasiswa tersebut selama dua semester awal (semester 1 dan semester 2) di UIN Sunan Kalijaga Yogyakarta. Peneliti mengambil objek penelitian pada mahasiswa jurusan Teknik Informatika adalah karena jurusan Teknik Informatika di kampus tersebut baru terbentuk mulai tahun 2004.

4.2 Lokasi Penelitian

Lokasi penelitian ini dilakukan di kampus UIN Sunan Kalijaga Yogyakarta. UIN Sunan Kalijaga Yogyakarta adalah satuan lembaga Pendidikan Tinggi yang diselenggarakan di bawah pengawasan kemenag.

- Visi UIN adalah Unggul dan Terkemuka dalam Pemanduan dan Pengembangan Keislaman dan Keilmuan bagi Peradaban.
- Misi UIN ada empat yaitu:
 - Memadukan dan mengembangkan studi keislaman, keilmuan, dan keindonesiaan dalam pendidikan dan pengajaran.
 - Mengembangkan budaya ijtihad dalam penelitian multidisipliner yang bermanfaat bagi kepentingan akademik dan masyarakat.

- Meningkatkan peran-serta institusi dalam menyelesaikan persoalan bangsa berdasarkan pada wawasan keislaman dan keilmuan bagi terwujudnya masyarakat madani.
- Membangun kepercayaan dan mengembangkan kerjasama dengan berbagai pihak untuk meningkatkan kualitas pelaksanaan Tridharma Perguruan Tinggi.
- Tujuan UIN ada lima yaitu:
 - Menghasilkan sarjana yang mempunyai kemampuan akademis dan profesional yang integratif-interkoneksi.
 - Menghasilkan sarjana yang beriman, berakhlak mulia, memiliki kecakapan sosial, manajerial, dan berjiwa kewirausahaan serta rasa tanggung jawab sosial kemasyarakatan.
 - Menghasilkan sarjana yang menghargai dan menjiwai nilai-nilai keilmuan dan kemanusiaan.
 - Menjadikan Universitas sebagai pusat studi yang unggul dalam bidang kajian dan penelitian yang integratif-interkoneksi.
 - Membangun jaringan yang kokoh dan fungsional dengan para alumni.

4.3 Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengcopy file berupa *file excel* dari bagian akademik di kampus UIN. File yang didapatkan berupa data mentah yang belum diolah sama sekali. File yang didapatkan sejumlah 2 file yang terdiri dari file mahasiswa jurusan Teknik Informatika angkatan selama

delapan angkatan terakhir, dan file data Indeks Prestasi Kumulatif (IPK) mahasiswa.

File pertama adalah file data mahasiswa jurusan Teknik Informatika angkatan 2009 sampai angkatan 2016. Jumlah data yang didapatkan berjumlah 709 data mahasiswa yang terdiri dari beberapa kolom atau beberapa atribut yaitu:

1. Tahun Akademik
2. Program Studi
3. Nomor Induk Mahasiswa (NIM)
4. Nama mahasiswa
5. Tempat lahir
6. Tanggal lahir
7. Alamat mahasiswa
8. Kota mahasiswa
9. Provinsi mahasiswa
10. Asal sekolah

File kedua adalah file data Indeks Prestasi Kumulatif (IPK) mahasiswa dari angkatan 2009 sampai angkatan 2016. Jumlah data mahasiswa yang didapat dalam file ini berjumlah 709 data mahasiswa. File kedua ini berisikan beberapa kolom yaitu kolom jurusan, NIM, nama mahasiswa, dan IPK.

4.4 Pre – Processing Data

Preprocessingdata adalah tahapan dari Data Mining yaitu suatu proses atau tahapan yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas atau inputan yang baik untuk dilanjutkan ke proses selanjutnya. Tahap *preprocessingdata* ini adalah tahap yang sangat krusial atau sangat penting dan harus dilakukan dengan teliti karena proses Data Mining membutuhkan data yang konsisten dalam penulisan, benar dalam format nya, tidak ada data yang kosong, duplikasi data, dan lain lain. Data yang tidak berkualitas maka hasil dari proses Data Mining ini akan menghasilkan hasil yang tidak berkualitas juga.

Dalam data mentah yang sudah didapatkan masih terdapat banyak data yang kurang berkualitas seperti kata kata yang salah dalam penulisan atau kehilangan salah satu huruf atau sering disebut *typo*, nilai angka yang eror karena format yang salah, duplikasi data, data yang dibutuhkan berada di file yang berbeda beda, dan singkatan yang tidak konsisten. Tahapan *preprocessing* memiliki beberapa proses yaitu :

4.4.1 DataReduction

Datareduction adalah proses untuk mereduksi atau mengurangi dimensi, atribut, ataupun sejumlah data yang tidak dibutuhkan. *Datareduction* sangat berguna untuk mendapatkan atribut dan sejumlah data yang akan digunakan di dalam penelitian ini. Dalam penelitian ini atribut yang dibutuhkan adalah kota asal mahasiswa, asal sekolah

mahasiswa, dan IPK mahasiswa selama dua semester awal yaitu Semester 1 dan Semester 2.

Pada file pertama yaitu file data mahasiswa semua angkatan dari tahun 2009 sampai tahun 2016. Setelah file tersebut dianalisa masih terdapat data atribut yang tidak dibutuhkan dalam penelitian ini. Peneliti mereduksi atribut tahun akademik, nama mahasiswa, tempat lahir, tanggal lahir, alamat mahasiswa, provinsi mahasiswa karena atribut tersebut tidak begitu berpengaruh terhadap proses pengelompokan data mahasiswa dalam penelitian ini dan data yang tidak direduksi peneliti adalah NIM mahasiswa, kota asal mahasiswa, dan asal sekolah mahasiswa. Selain mereduksi atribut yang tidak dibutuhkan peneliti juga mereduksi data mahasiswa angkatan 2009 sampai dengan angkatan 2016. Sehingga hasil akhir dari proses reduksi mendapatkan data mahasiswa dengan angkatan 2009 sampai dengan angkatan 2016 berjumlah 709 data mahasiswa.

Pada file kedua yaitu file tentang Indeks Prestasi Kumulatif (IPK) mahasiswa semua angkatan dari tahun 2009 sampai tahun 2016. Setelah peneliti menganalisis file data tersebut peneliti mereduksi sejumlah data mahasiswa angkatan tahun 2009 sampai dengan angkatan 2016 karena data mahasiswa angkatan tersebut tidak dibutuhkan dalam penelitian ini. Setelah itu peneliti juga mereduksi data mahasiswa dengan nilai IPK.

4.4.2 Data Cleaning

Datacleaning adalah proses untuk mengisi *missing value* atau isian yang hilang atau kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi redundansi data yaitu penumpukan data atau duplikasi data. Peneliti melakukan proses *datacleaning* dengan aplikasi *Microsoft Excel* dengan cara memfilter setiap kolom dan mencari data yang kosong atau *missing*. Pada kolom asal kota dan asal sekolah terdapat data yang tidak konsisten dalam penulisannya sehingga data harus dikoreksi dan diubah agar data tetap konsisten.

Pada file data pertama yang sudah melewati proses *datareduction* berisikan 709 data mahasiswa angkatan 2009 sampai dengan angkatan 2016. Setelah data tersebut difilter dan dianalisa terdapat data yang kosong atau blank pada kolom asal kota mahasiswa dan asal sekolah mahasiswa. Pada kolom asal kota terdapat data yang kosong sejumlah satu data dan peneliti menambahi data tersebut dengan berdasarkan kolom tempat lahir, asal sekolah, provinsi, dan alamat rumah. Selain terdapat data yang kosong atau blank terdapat juga data yang tidak konsisten dalam penulisan pada atribut kota asal mahasiswa dan asal sekolah mahasiswa, maka peneliti harus mencari dan mengubah data yang tidak konsisten tersebut menjadi data yang konsisten dan menghasilkan data yang berkualitas.

Pada file kedua yang sudah melewati proses *datareduction* tentang indeks prestasi kumulatif berisikan sejumlah 178 data. Setelah dianalisa terdapat data angka yaitu IPK mahasiswa dengan format yang kurang pas

yaitu dengan format general maka peneliti mengganti format tersebut ke dalam format angka desimal. Selain data IPK, peneliti juga mengganti format pada data NIM mahasiswa kedalam format number karena data mentah yang didapatkan menggunakan format general.

4.4.3 Data Transformation

Data transformation adalah proses mengubah suatu data supaya mendapatkan data yang lebih berkualitas atau sesuai dengan kebutuhan. Dalam proses ini peneliti hanya akan melakukan proses transformation data pada file pertama dan file kedua.

File yang pertama adalah file tentang data mahasiswa angkatan 2009 sampai dengan angkatan 2016. Setelah file pertama melewati proses *data reduction* dan *data cleaning* peneliti sudah mendapatkan atribut sesuai dengan kebutuhan dalam penelitian ini tetapi isi data dari atribut tersebut masih belum memenuhi kondisi atau syarat yang cocok untuk dilanjutkan ke dalam proses algoritma K-Means Clustering. Dalam proses algoritma K-Means Clustering data yang bisa diolah hanya data yang berupa angka sedangkan data yang ada pada atribut kota asal mahasiswa dan asal sekolah mahasiswa adalah data yang berupa text sehingga peneliti harus mengubah data tersebut kedalam angka agar bisa dilanjutkan ke dalam proses K-Means Clustering. Langkah pertama yang dilakukan peneliti adalah mencari frekuensi atau jumlah data yang sama di setiap atribut lalu mengurutkannya dari frekuensi tersebar sampai ke terkecil dan memisahkannya kedalam tabel yang berbeda. Setelah itu peneliti memberi

inisial pada setiap data berdasarkan frekuensi terbesar ke terkecil tujuannya adalah frekuensi yang paling besar nilainya merupakan data yang paling banyak muncul maka inisial dimulai dari frekuensi terbanyak. Pada atribut kota asal mahasiswa peneliti mendapatkan 112 data asal kota dengan frekuensi dan inisial sebagai berikut:

Tabel 4.1 Tabel Frekuensi Kota Asal Mahasiswa

KOTA ASAL	FREKUANSI	INISIAL
BANTUL	35	1
SLEMAN	23	2
YOGYAKARTA	12	3
MAGELANG	11	4
BANJAR NEGARA	7	5
BREBES	7	6
CILACAP	7	7
GUNUNG KIDUL	7	8
KEBUMEN	7	9
KULON PROGO	7	10
KLATEN	6	11
NGAWI	6	12
TEMANGGUNG	6	13
TEGAL	5	14
WONOGIRI	5	15
BIMA	4	16
GRESIK	4	17

Tabel 4.1 Tabel Frekuensi Kota Asal Mahasiswa (lanjutan)

KOTA ASAL	FREKUANSI	INISIAL
MAGETAN	4	18
PATI	4	19
PONOROGO	4	20
BANYUMAS	3	21
CIREBON	3	22
LAMPUNG BARAT	3	23
PURWOREJO	3	24
SEMARANG	3	25
SUKOHARJO	3	26
TASIKMALAYA	3	27
AMBON	2	28
BALIKPAPAN	2	29
BANYUWANGI	2	30
BEKASI	2	31
BLITAR	2	32
BLORA	2	33
DEPOK	2	34
GROBOGAN	2	35
HULU SUNGAI UTARA	2	36
INDRAMAYU	2	37
KUNINGAN	2	38
MUARA ENIM	2	39
MUSI BANYUASIN	2	40
PACITAN	2	41
PEKANBARU	2	42
PURBALINGGA	2	43
WONOSOBO	2	44
ASAHAN	1	45
BATAM	1	46
BATANG	1	47
BAWANG SAKTI JAYA	1	48

Tabel 4.1 Tabel Frekuensi Kota Asal Mahasiswa (lanjutan)

KOTA ASAL	FREKUANSI	INISIAL
BENGKULU UTARA	1	49
BOGOR	1	50
BONDOWOSO	1	51
BOYOLALI	1	52
CIAMIS	1	53
DENPASAR	1	54
DHARMASRAYA	1	55
JAKARTA UTARA	1	56
JEMBER	1	57
JEPARA	1	58
JOHOR BAHRU	1	59
JOMBANG	1	60
KARANGANYAR	1	61
KARANGASEM	1	62
KARAWANG	1	63
KEDIRI	1	64
KENDAL	1	65
KENDARI	1	66
KERTAK HANYAR	1	67
KUDUS	1	68
LABUHANBATU UTARA	1	69
LAMONGAN	1	70
LAMPUNG TENGAH	1	71
LANGKAT	1	72
LOMBOK TENGAH	1	73
LOMBOK TIMUR	1	74
LUAR NEGRI	1	75
LUBUKLINGGAU	1	76
LUWU UTARA	1	77
MADIUN	1	78
MAJALENGKA	1	79
MAKASSAR	1	80
MALANG	1	81
MANDAILING NATAL	1	82
MATARAM	1	83
MEDAN	1	84

Tabel 4.1 Tabel Frekuensi Kota Asal Mahasiswa (lanjutan)

KOTA ASAL	FREKUANSI	INISIAL
METRO	1	85
MUSI RAWAS	1	86
NGANJUK	1	87
PALANGKARAYA	1	88
PANDEGLANG	1	89
PASURUAN	1	90
PEGAGAN	1	91
PEKALONGAN	1	92
PEMALANG	1	93
PONTIANAK	1	94
PRINGSEWU	1	95
PROBOLINGGO	1	96
RABANGODU	1	97
SALATIGA	1	98
SAROLANGUN	1	99
SIAK	1	100
SIMALUNGUN	1	101
SITUBONDO	1	102
SRAGEN	1	103
SURAKARTA	1	104
TAMBAN	1	105
TANJUNG BALAI	1	106
TANJUNG PINANG	1	107
TRENGGALEK	1	108
TRIMULYA	1	109
TUBAN	1	110
TULANG BAWANG BARAT	1	111
TULUNGAGUNG	1	112

Pada atribut asal sekolah mahasiswa peneliti mendapatkan 5 data asal sekolah dengan frekuensi dan inisial sebagai berikut :

Tabel 4.2 Tabel Frekuensi Asal Sekolah Mahasiswa

ASAL SEKOLAH	FREKUENSI	INISIAL
SMA	389	1
MA	148	2
SMK	147	3
PON	21	4
SMTI	4	5

File kedua adalah data indeks prestasi sementara mahasiswa karena pada file data ini peneliti belum mendapatkan data yang diperlukan untuk proses selanjutnya yaitu IPK selama semester 1 dan semester 2. File data kedua ini sudah melewati proses *reduction* dan proses *datacleaning*, sehingga data yang didapatkan adalah data mahasiswa angkatan 2009 sampai dengan angkatan 2016 dengan indeks prestasi sementara (IPS) selama 2 semester. Peneliti akan mengubah data tersebut menjadi indeks prestasi kumulatif selama 2 semester dengan menggunakan rumus :

$$\text{Rumus IPK Semester 2} = (\text{IPS smt 1} + \text{IPS smt 2}) / 2$$

Setelah peneliti mengubah data indeks prestasi sementara tersebut kedalam Indeks Prestasi Kumulatif selama 2 semester dengan rumus di atas peneliti mendapatkan sejumlah data yang bisa dilihat pada (Lampiran Tabel IP Semester 2 angkatan 2009 sampai dengan angkatan 2016).

4.4.4 Data Integration

Data integration atau integrasi data adalah suatu proses untuk menggabungkan data dari beberapa file sumber. Data *integration* ini hanya

dilakukan apabila data yang akan diolah bersumber dari beberapa file sumber. Pada pengambilan data penelitian ini peneliti mendapatkan 3 *file excel* dan semua file data tersebut harus diintegrasikan atau digabungkan sesuai dengan kebutuhan penelitian menjadi satu tabel dan satu file agar mudah untuk diolah kedalam metode K-Means clustering.

Setelah ketiga data yang didapat melewati proses *datareduction, datacleaning*, dan *datatransformation* maka langkah terakhir adalah menggabungkan atau mengintegrasikan semua data tersebut menjadi satu dataset ke dalam suatu tabel. Peneliti menggunakan aplikasi *Microsoft Access Database 2010* untuk melakukan proses *dataintegration* ini. Proses integrasi data ini membutuhkan beberapa langkah atau tahapan yaitu:

1. Langkah pertama adalah peneliti membuat database baru dengan format *Microsoft Access 97-2003* menggunakan aplikasi *Microsoft Access 2010*.
2. Langkah kedua adalah peneliti mengubah format pada ketiga *file excel* yang sudah diproses sebelumnya menjadi format *Microsoft Excel 97-2003*.
3. Kemudian peneliti mengimportkan ketiga *file excel* tersebut ke dalam database baru yang sudah dibuat sebelumnya di *Microsoft Access* dengan menentukan kolom NIM sebagai *primary key*.

4. Setelah tahapan mengimport *file excel* ke dalam database *Microsoft Access* berhasil langkah selanjutnya adalah membuat relasi antar tabel dengan primary key atau data kunci yaitu NIM mahasiswa.
5. Langkah selanjutnya adalah membuat *query design* yang akan menampilkan dataset mahasiswa berdasarkan atribut yang dibutuhkan dalam penelitian ini yaitu NIM mahasiswa, kota asal mahasiswa, asal sekolah mahasiswa, dan IPK mahasiswa selama 2 semester.
6. Setelah proses menampilkan dataset mahasiswa dengan *design query* berhasil lalu hasil tersebut di ekspor kedalam *file excel* dengan format yang sama dan mendapatkan dataset yang siap untuk diolah menggunakan algoritma K-Means Clustering. Dataset yang di dapatkan dari proses ini adalah berjumlah 709 data mahasiswa dengan memiliki 4 atribut yaitu NIM mahasiswa, kota asal mahasiswa, asal sekolah mahasiswa, dan IPK mahasiswa selama 2 semester. Dataset ini dapat dilihat dalam lembar (Lampiran Tabel Dataset Mahasiswa).

4.5 Penerapan Algoritma K-Means Clustering

Setelah data mentah yang didapat dari pihak universitas diolah pada tahapan *pre-processing* dan sudah mendapatkan dataset yang siap untuk diolah maka tahap berikutnya adalah menerapkan algoritma K-Means Clustering pada dataset tersebut. Dalam penelitian ini peneliti menerapkan algoritma K-Means Clustering menggunakan dua aplikasi yaitu aplikasi *Microsoft Excel* dan aplikasi K-Means Clustering yang akan dibuat dengan bahasa pemrograman PHP. Peneliti menggunakan dua aplikasi tersebut

bertujuan untuk membandingkan dan membuktikan bahwa aplikasi K-Means menggunakan PHP yang dibuat oleh peneliti mempunyai hasil yang sama dengan algoritma K-Means Clustering yang diterapkan menggunakan aplikasi *Microsoft Excel*. Urutan penerapan yang dilakukan peneliti adalah yang pertama menerapkan algoritma tersebut menggunakan aplikasi *Microsoft Excel 2010* kemudian menerapkannya menggunakan sistem aplikasi K-Means menggunakan bahasa pemrograman PHP.

4.5.1 Penerapan Algoritma menggunakan *Microsoft Excel 2010*

Penerapan algoritma K-Means Clustering menggunakan *Microsoft Excel 2010* membutuhkan beberapa tahapan proses sebagai berikut :

1. Menentukan Titik Pusat Cluster

Titik pusat cluster atau bisa disebut *centroid* digunakan sebagai nilai pengurang untuk perhitungan jarak antara data ke setiap cluster atau disebut *distance*. Dalam proses ini penentuan nilai titik pusat cluster bisa ditentukan dengan metode acak sesuai keinginan peneliti dengan syarat nilai *centroid* masih termasuk dalam range nilai data pada setiap atribut. Selain itu penentuan nilai titik pusat cluster bisa ditentukan dengan menggunakan nilai rata-rata disetiap atribut.

Dalam proses ini peneliti menentukan nilai titik pusat cluster menggunakan nilai rata – rata dari setiap atribut dataset yaitu kota asal mahasiswa, asal sekolah mahasiswa, dan IPK mahasiswa. Nilai rata – rata pada setiap atribut tersebut dipakai sebagai nilai *centroid* pada

cluster 2. Perhitungan rata – rata tersebut berdasarkan data mahasiswa Prodi Teknik Informatika yang dapat dilihat pada (Lampiran B Tabel Dataset Mahasiswa).

Nilai rata – rata didapatkan dengan menjumlahkan nilai data pada setiap atribut kemudian dibagi dengan jumlah data yang ada. Nilai yang didapatkan adalah pada atribut kota asal mahasiswa adalah 5,73061, asal sekolah mahasiswa 4,50816, dan IPK mahasiswa adalah 3,38. Lalu untuk mengisi nilai *centroid* pada cluster satu dan cluster tiga peneliti menentukan nilai *centroid* tersebut berdasarkan nilai *centroid* pada cluster dua yaitu nilai cluster satu kurang dari nilai cluster dua sedangkan nilai cluster tiga lebih tinggi dari nilai cluster dua. Sehingga peneliti mendapatkan nilai titik pusat setiap cluster / nilai *centroid* adalah sebagai berikut:

Tabel 4.3 Tabel Titik Pusat Cluster

<i>TITIK AWAL PUSAT</i>	<i>KOTA ASAL</i>	<i>ASAL SEKOLAH</i>	<i>IPK</i>
CLUSTER 1	2	1	1,63
CLUSTER 2	28	2	2,12
CLUSTER 3	112	5	3,83

Berdasarkan tabel 4.3 dapat dijelaskan bahwa cluster 1 mempunyai nilai *centroid* pada atribut kota asal 2, asal sekolah 1, dan IPK 1,63 sehingga cluster 1 digunakan sebagai tempat atau wadah untuk

menampung data yang mempunyai nilai terdekat dengan nilai *centroid* tersebut.

2. Menghitung Jarak Data ke Setiap Cluster

Setelah mendapatkan nilai titik pusat setiap cluster proses selanjutnya adalah menghitung jarak data ke setiap cluster yang tersedia atau bisa disebut *distance*. Cara untuk menghitung nilai jarak atau *distance* tersebut peneliti menggunakan sebuah rumus 3.1

Berdasarkan rumus 3.1 peneliti membuat sebuah fungsi dengan cara perhitungan yang sama dan menyesuaikan format yang berjalan pada aplikasi *Microsoft Excel* 2010. Perubahan rumus tersebut adalah sebagai berikut :

$$=SQRT(((data\ pertama\ kota\ asal - centroid\ kota\ asal)^2) + ((data\ pertama\ asal\ sekolah - centroid\ asal\ sekolah)^2) + ((data\ pertama\ IPK - centroid\ IPK)^2))$$

Keterangan :

- SQRT adalah fungsi akar pada *Microsoft Excel*
- Data pertama pada setiap atribut adalah kolom pertama data pada setiap atribut
- *Centroid* bisa berubah disesuaikan dengan cluster masing-masing, setelah menentukan kolom *centroid* tekan tombol F4 pada *keyboard* untuk membuat nilai tersebut tidak berubah – ubah jika didrag kebawah

- Simbol 2 adalah pangkat 2

Dengan fungsi di atas peneliti melakukan perhitungan sekali dan jika sudah berhasil *drag* ke bawah sampai dataset paling bawah. Hasil perhitungan jarak pada proses ini dapat dilihat dalam (Lampiran Tabel Perhitungan Jarak Data ke Setiap Cluster(M.Excel)).

3. Mengalokasikan Data ke dalam Cluster

Setelah proses perhitungan jarak data ke setiap cluster selesai maka proses berikutnya adalah mengalokasikan data ke dalam setiap cluster yang terbentuk. Pengalokasian data tersebut berdasarkan hasil jarak antara data ke setiap cluster, apabila nilai jarak antara data pertama ke cluster 1 lebih kecil daripada nilai jarak antara data pertama ke cluster 2 ataupun cluster 3 maka data pertama masuk ke dalam cluster 1. Pengalokasian data ini dilakukan bertujuan agar bisa menentukan titik pusat cluster baru pada proses selanjutnya. Dalam proses ini peneliti mengalokasikan data menggunakan fungsi IF pada *Microsoft Excel*, fungsi tersebut adalah sebagai berikut :

=IF(nilai hasil jarak data pertama ke cluster satu < nilai dari hasil jarak data pertama ke cluster dua; IF(nilai dari hasil jarak data pertama ke cluster satu < nilai hasil jarak data pertama ke cluster tiga; "ok"; ""); "")

Keterangan :

- Fungsi tersebut diposisikan pada kolom setelah kolom perhitungan jarak data ke cluster dan baris data pertama.

- Fungsi IF adalah fungsi dalam *Microsoft Excel* untuk menentukan nilai dengan suatu kondisi tertentu. Maksud dari fungsi diatas adalah jika kolom G5 (nilai dari hasil jarak data pertama ke cluster 1) lebih kecil dari kolom J5 (nilai dari hasil jarak data pertama ke cluster 2) dan lebih kolom G5 lebih kecil dari kolom nilai hasil jarak data pertama ke cluster 3 maka *ok* (data pertama masuk ke cluster 1).

Setelah proses pengalokasian data atau clustering data ini selesai, peneliti mendapatkan hasil yaitu data yang masuk ke cluster 1 adalah sebanyak 234 data sedangkan data yang masuk ke cluster 2 adalah sebanyak 158 data dan data yang masuk ke cluster 3 adalah sebanyak 91 data. Jika data tersebut dijumlahkan data yang didapatkan adalah 483. Terdapat 7 data yang belum masuk ke cluster manapun, itu terjadi karena nilai jarak atau *distance* pada 7 data tersebut ada yang bernilai sama dan kondisi tersebut tidak memenuhi syarat dari fungsi yang telah dibuat sebelumnya. Hasil dari proses pengalokasian data ini dapat dilihat pada (Lampiran Tabel Clustering Data (M.Excel)).

4. Menentukan Titik Pusat Cluster Baru

Setelah proses pengalokasian data sudah selesai maka tahapan selanjutnya adalah menentukan titik cluster baru. Penentuan *centroid* baru ini menggunakan metode yang hampir sama dengan menentukan titik pusat cluster pada tahapan sebelumnya, yang membedakan dari kedua tahapan tersebut adalah dari jumlah data yang digunakan. Dalam tahapan ini *centroid* dapat ditentukan dengan cara menghitung nilai rata

– rata sesuai dengan data yang sudah dikelompokkan kedalam setiap cluster masing - masing (tahapan pengalokasian data). Dalam proses ini peneliti mencari nilai rata – rata berdasarkan data yang sudah dikelompokkan ke setiap cluster masing – masing dan mendapatkan hasil sebagai berikut :

Pada cluster 1 dengan jumlah data 508 data

Tabel 4.4 Tabel Titik Pusat Cluster Baru (Cluster 1)

<i>TITIK PUSAT CLUSTER BARU</i>					
ASAL KOTA	<i>5,18</i>	ASAL SEKOLAH	<i>1,78</i>	IPK	<i>3,325</i>

- Pada cluster 2 dengan jumlah data 139 data

Tabel 4.5 Tabel Titik Pusat Cluster Baru (Cluster 2)

<i>TITIK PUSAT CLUSTER BARU</i>					
ASAL KOTA	<i>34,32</i>	ASAL SEKOLAH	<i>1,71</i>	IPK	<i>3,336</i>

- Pada cluster 3 dengan jumlah data 62 data

Tabel 4.6 Tabel Titik Pusat Cluster Baru (Cluster 3)

<i>TITIK PUSAT CLUSTER BARU</i>					
ASAL KOTA	<i>91,5</i>	ASAL SEKOLAH	<i>1,71</i>	IPK	<i>3,23</i>

5. Memverifikasi Titik Pusat Cluster

Proses ini adalah tahap untuk memverifikasi titik pusat cluster. Proses tersebut adalah proses untuk memverifikasi antara titik pusat cluster baru dengan titik pusat cluster yang lama, apabila kedua titik pusat cluster tersebut mempunyai nilai *centroid* yang berbeda maka proses K-Means Clustering masih berlanjut dan akan dimulai lagi pada proses ke-2 yaitu (menghitung jarak data ke setiap cluster) dengan menggunakan nilai titik pusat cluster yang baru. Sedangkan jika nilai kedua *centroid* tersebut bernilai sama maka proses K-Means Clustering berhenti sampai di tahap tersebut. Proses verifikasi ini dilakukan dengan tujuan untuk menentukan apakah proses K-Means Clustering sudah selesai atau masih perlu dilakukan pengulangan proses kembali.

Dalam proses ini peneliti melalui 18 tahapan proses pengulangan untuk mendapatkan nilai titik pusat cluster atau *centroid* yang sudah tidak berubah – ubah lagi.

CLUSTER 1			CLUSTER 2			CLUSTER 3			CLUSTER 1	CLUSTER 2	CLUSTER 3
8,576772	1,69685	3,195599	50,65468	1,884892	3,21338	112,5	1,709677	3,216764			
5,621716627			47,66313007			109,5024084			ok		
4,659374581			46,66584159			108,5034982			ok		
7,609134593			49,66265002			111,5023013			ok		
2,042781721			40,68074154			102,5107355			ok		
27,42605387			14,65759573			76,50103558				ok	
4,630399526			46,66319256			108,5023782			ok		
8,460041254			33,66852562			95,50343249			ok		
1,430625616			41,67145989			103,508779			ok		
1,73653636			43,66406326			105,5025544			ok		
7,455864865			34,66597726			96,50261209			ok		
7,702681521			49,66964047			111,5085697			ok		
3,675671124			45,66597851			107,5035168			ok		
126,4304299			84,35349804			22,54003025					ok
3,65250874			45,66383568			107,5025836			ok		
0,913063116			42,66409018			104,5025108			ok		
4,62996959			46,66314011			108,5023548			ok		
30,43183666			11,69016868			73,50374568				ok	
0,820089724			41,66420976			103,5024911			ok		
1,773262477			43,66545526			105,5031239			ok		
1,724751812			43,6636592			105,5023921			ok		
4,630879266			46,66324853			108,5024029			ok		
1,775544775			40,67254495			102,5057529			ok		
5,621295355			47,66298823			109,502339			ok		

Gambar 4.1 Verifikasi DataCentroid (*Microsoft Excel*)

4.5.2 Hasil Penerapan K-Means menggunakan *Microsoft Excel* 2010

Setelah melewati tahapan proses K-Means Clustering menggunakan aplikasi *Microsoft Excel* 2010 maka proses berikutnya adalah menampilkan hasil dari penerapan tersebut. Untuk mendapatkan hasil yang diolah menggunakan algoritma K-Means Clustering peneliti melewati perulangan tahapan proses sebanyak 18 kali proses. Hasil dari penerapan ini adalah sebagai berikut:

Tabel 4.7 Tabel Hasil Cluster 1 Penerapan Algoritma K-MeansClustering (*Microsoft Excel*)

HASIL CLUSTER 1			
CLUSTER 1 TERDIRI DARI 508 MAHASISWA YANG BERASAL DARI			
KOTA ASAL		ASAL SEKOLAH	
YOGYAKARTA	49	SMA	298
WONOGIRI	11	MA	86
TEMANGGUNG	13	SMK	108
TEGAL	12	SMTI	4
SUKOHARJO	5	PONDOK	12
SRAGEN	8		
SLEMAN	63		
PURWOREJO	19		
PURBALINGGA	5		
PATI	12		
PACITAN	6		
NGAWI	14		
MAGETAN	8		
MAGELANG	30		
LAMONGAN	5		
KULON PROGO	19		
KLATEN	25		
KEBUMEN	22		
JEPARA	6		
GUNUNG KIDUL	19		
GRESIK	5		
CIREBON	5		
CILACAP	17		
BREBES	17		
BOYOLALI	7		
BIMA	6		
BANYUMAS	8		
BANTUL	75		
BANJARNEGARA	17		
TOTAL	508		508

Berdasarkan hasil cluster 1 dapat diketahui terdapat 508 data mahasiswa yang termasuk ke dalam kelompok cluster 1 dengan nilai rata – rata IPK 3,32. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kota asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah mahasiswa dari kota dan asal masing – masing tersebut. Jumlah data mahasiswa pada cluster 1 adalah jumlah data yang paling banyak diantara cluster yang lain sehingga dapat dikatakan bahwa

mayoritas mahasiswa prodi Teknik Informatika paling banyak masuk ke dalam kelompok cluster 1. Berdasarkan hal tersebut pihak prodi dapat mempertimbangkan strategi promosi di tahun depan menggunakan hasil persebaran wilayah dan asal sekolah pada cluster 1.

Tabel 4.8 Tabel Hasil Cluster 2 Penerapan Algoritma K-Means Clustering

(*Microsoft Excel*)

HASIL CLUSTER 2			
CLUSTER 2 TERDIRI DARI 139 MAHASISWA YANG BERASAL DARI			
KOTA ASAL		ASAL SEKOLAH	
AMBON	2	SMA	58
BALIKPAPAN	2	MA	46
BANGKA	2	SMK	28
BANYUWANGI	3	PONDOK	7
BATANG	4		
BEKASI	5		
BLITAR	2		
BLORA	4		
CIAMIS	5		
DEMAK	4		
DEPOK	2		
GROBOGAN	3		
HULU SUNGAI UTARA	2		
INDRAMAYU	2		
JAKARTA TIMUR	2		
JAKARTA UTARA	2		
JEMBER	2		
JOMBANG	3		
KARAWANG	3		
KEDIRI	2		

Tabel 4.8 Tabel Hasil Cluster 2 Penerapan Algoritma K-Means (lanjutan)

HASIL CLUSTER 2			
CLUSTER 2 TERDIRI DARI 139 MAHASISWA YANG BERASAL DARI			
KOTA ASAL		ASAL SEKOLAH	
KENDAL	2	SMA	58
KUNINGAN	2	MA	46
LABUHAN BATU UTARA	3	SMK	28
LAMPUNG BARAT	3	PONDOK	7
LAMPUNG TIMUR	2		
LOMBOK TENGAH	3		
MADIUN	4		
MAJALENGKA	2		
MANDAILING NATAL	3		
METRO	2		
MUARA ENIM	2		
MUSI BANYUASIN	2		
OGAN KOMERING	4		
PALANGKARAYA	2		
PASURUAN	2		
PEKALONGAN	4		
PEKANBARU	2		
PONOROGO	4		
PONTIANAK	2		
REMBANG	3		
ROKAN HULU	3		
SALATIGA	2		
SEMARANG	4		
SUBANG	2		
SURAKARTA	2		
TANGERANG	3		
TASIKMALAYA	4		
TRIMULYA	1		
TUBAN	2		
TULANG BAWANG BARAT	1		
TULUNGAGUNG	2		
WONOSOBO	4		
HASIL	139		139

Berdasarkan hasil cluster 2 dapat diketahui bahwa terdapat 139 data mahasiswa yang termasuk ke dalam kelompok cluster 2 dengan nilai rata – rata IPK 3,321. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kota asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah mahasiswa dari kota dan asal masing – masing. Jumlah data mahasiswa pada cluster 2 adalah jumlah data yang kurang dari cluster 1 tetapi lebih dari cluster 3. Informasi yang menonjol dari cluster 2 adalah informasi mengenai persebaran asal sekolah mahasiswa yaitu banyaknya jumlah jurusan asal sekolah yang berbeda beda dengan frekuensi yang sangat sedikit. Berdasarkan hal tersebut dapat diketahui bahwa jurusan asal sekolah dengan frekuensi paling sedikit adalah asal sekolah yang kurang meminati untuk masuk di jurusan / prodi Teknik Informatika.

Tabel 4.9 Tabel Hasil Cluster 3 Penerapan Algoritma K-Means Clustering
(*Microsoft Excel*)

HASIL CLUSTER 3			
CLUSTER 3 TERDIRI DARI 62 MAHASISWA YANG BERASAL DARI			
KOTA ASAL		ASAL SEKOLAH	
ASAHAN	1	SMA	33
BATAM	1	SMK	11
BAWANG SAKTI JAYA	1	PONDOK	2
BENGKALIS	1	MA	16
BENGKULU UTARA	1		
BINTAN	1		
BOGOR	1		
BONDOWOSO	1		

Tabel 4.9 Tabel Hasil Cluster 3 Penerapan Algoritma K-Means Clustering

(lanjutan)

HASIL CLUSTER 3			
CLUSTER 3 TERDIRI DARI 62 MAHASISWA YANG BERASAL DARI			
KOTA ASAL		ASAL SEKOLAH	
BULELENG	1	SMA	33
BULUKUMBA	1	SMK	11
CIANJUR	1	PONDOK	2
DENPASAR	1	MA	16
DHARMASRAYA	1		
DOMPU	1		
GARUT	1		
INDRAGIRI HILIR	1		
JAMBI	1		
JOHOR BAHRU	1		
KAMPAR	1		
KARANGANYAR	1		
KARANGASEM	1		
KENDARI	1		
KERTAK HANYAR	1		
KUDUS	1		
KUTAI	1		
LAHAT	1		
LAMPUNG SELATAN	1		
LAMPUNG TENGAH	1		
LANGKAT	1		
LHOKSEUMAWE	1		
LOMBOK TIMUR	1		
LUAR NEGERI	1		
LUBUKLINGGAU	1		
LUMAJANG	1		
LUWU UTARA	1		
MAJENE	1		
MAKASAR	1		
MALANG	1		
MATARAM	1		

Tabel 4.9 Tabel Hasil Cluster 3 Penerapan Algoritma K-Means Clustering

(lanjutan)

HASIL CLUSTER 3			
CLUSTER 3 TERDIRI DARI 62 MAHASISWA YANG BERASAL DARI			
KOTA ASAL		ASAL SEKOLAH	
MEDAN	1	SMA	33
MUSI RAWAS	1	SMK	11
NGANJUK	1	PONDOK	2
PANDEGLANG	1	MA	16
PEGAGAN	1		
PEMALANG	1		
PINRANG	1		
PRINGSEWU	1		
PROBOLINGGO	1		
PURWAKARTA	1		
RABANGODU	1		
SAROLANGUN	1		
SIAK	1		
SIDOARJO	1		
SIMALUNGUN	1		
SITUBONDO	1		
SUKABUMI	1		
SUMENEP	1		
TAMBAN	1		
TANJ. JABUNG BARAT	1		
TANJUNG BALAI	1		
TANJUNG PINANG	1		
TRENGGALEK	1		
TOTAL	62		62

Berdasarkan hasil cluster 3 dapat diketahui bahwa terdapat 62 data mahasiswa yang termasuk ke dalam kelompok cluster 3 dengan nilai rata – rata IPK 3,323. Dalam tabel tersebut terdapat informasi tentang persebaran wilayah kota asal mahasiswa maupun asal sekolah mahasiswa beserta

dengan jumlah mahasiswa dari kota dan asal masing – masing. Jumlah data mahasiswa pada cluster 3 adalah jumlah data yang paling sedikit dibandingkan dengan cluster yang lain. Informasi yang menonjol dari cluster 3 adalah informasi mengenai persebaran wilayah kota asal mahasiswa yaitu banyaknya jumlah kota asal mahasiswa yang berbeda beda dengan frekuensi yang sangat sedikit. Berdasarkan hal tersebut dapat diketahui bahwa kota asal dengan frekuensi paling sedikit adalah kota yang kurang meminati untuk masuk di jurusan / prodi Teknik Informatika.

4.5.3 Penerapan Algoritma Menggunakan Aplikasi PHP

Setelah dilakukan proses manual menggunakan Aplikasi *microsoft excel*, peneliti melakukan pengujian ulang dengan menggunakan aplikasi lain. Bertujuan sebagai perbandingan antara proses manual dengan menggunakan aplikasi dan juga untuk pembuktian apakah algoritma *K-Means Clustering* telah diterapkan dengan benar pada proses manual. Dengan demikian akan terlihat apakah ada perbedaan antara proses manual dengan proses menggunakan aplikasi. Jika terjadi perbedaan, hanya selisih sedikit dan tidak mempengaruhi hasil.

Penerapan algoritma K-Means Clustering yang kedua ini peneliti menggunakan aplikasi yang dibuat dengan bahasa pemrograman PHP. Aplikasi ini adalah berbasis *website* dengan menggunakan database MySQL. Sistem yang digunakan dalam penerapan ini menggunakan metode yang sama dengan metode algoritma K-Means Clustering.

4.5.4 Perancangan Sistem Aplikasi K-Means Clustering

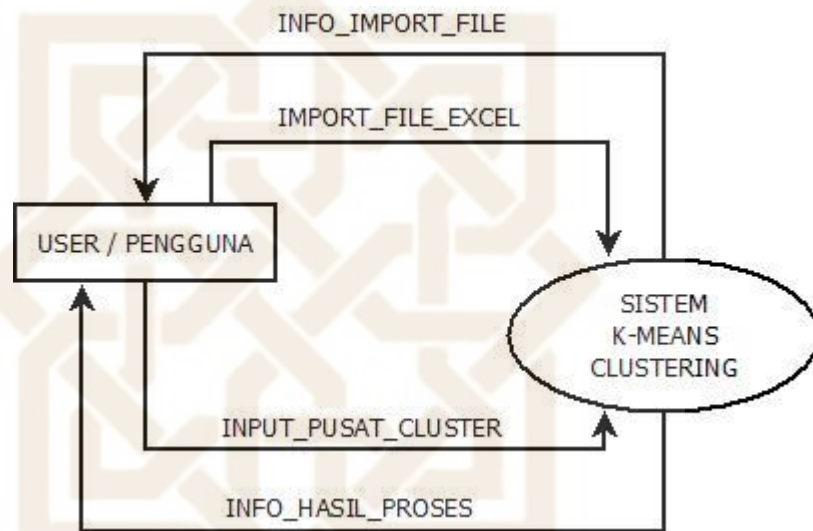
Untuk membuat sebuah sistem aplikasi berbasis web peneliti menggunakan aplikasi database MySQL sebagai tempat untuk menampung dan menyimpan sebuah data, aplikasi *Sublime Text 2* untuk proses pengcodangan berbagai perintah dan fungsi yang dibutuhkan untuk membangun sistem dengan algoritma K-Means Clustering, dan sebuah browser untuk menampilkan hasil pengcodangan. Tahapan pertama yang dilakukan peneliti adalah membuat perancangan sistem aplikasi web dengan menggunakan metode DFD (*Data Flow Diagram*) dan ERD (*Entity Relationship Diagram*). Metode DFD adalah metode perancangan sistem yang menjelaskan informasi alur dan proses pada sistem yang dibangun, sedangkan metode ERD adalah metode perancangan sistem yang menggambarkan model rancangan database dari sistem yang dibangun. Berikut adalah perancangan sistem tersebut.

1. Perancangan DFD

Perancangan sistem menggunakan metode DFD memiliki beberapa tahapan proses dengan urutan paling awal adalah penjelasan umum mengenai cara kerja sistem dan dilanjutkan dengan penjelasan yang lebih spesifik.

a. DFD Level 0 (Diagram Konteks)

DFD level 0 merupakan gambaran dari suatu informasi proses yang berjalan pada sistem aplikasi K-Means Clustering yang dirancang untuk mengetahui bagaimana sistem berinteraksi dengan beberapa *entity* untuk mepresentasikan keseluruhan sistem.



Gambar 4.2 DFD Level 0 (Diagram Konteks)

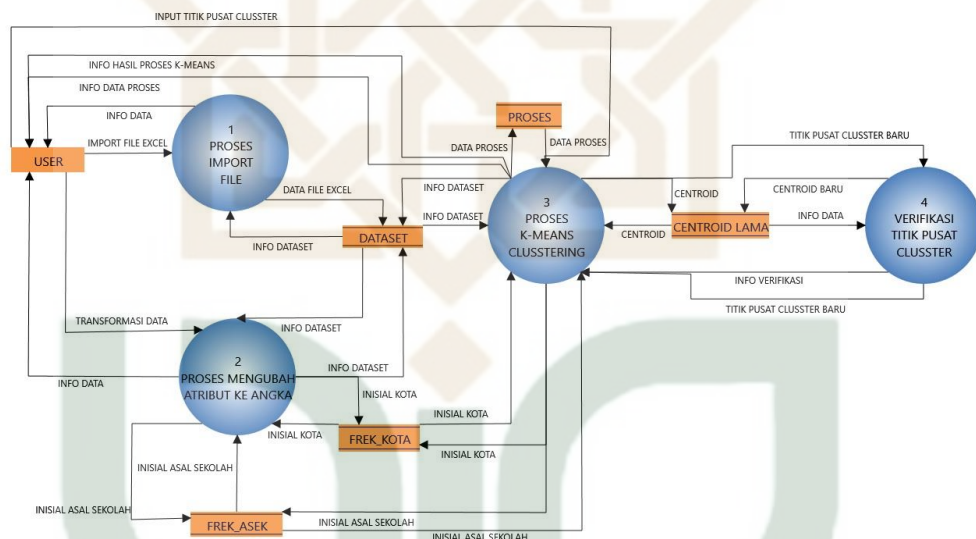
Berdasarkan gambar 4.2 *user* mempunyai 2 alur yaitu alur yang pertama adalah *user* atau pengguna mengimportkan *file excel* kedalam sistem setelah itu data dari *file excel* tersebut disimpan kedalam database dengan nama tabel *dataset* dan *user* mendapatkan informasi bahwa file data tersebut berhasil di import dengan catatan *file excel* yang bisa diimportkan adalah menggunakan format 97-2003 dan file tersebut berisikan sejumlah data yang sudah melalui tahap *pre-processing* data kecuali perubahan data atribut ke nilai angka. Alur kedua adalah pengguna mengisikan titik

pusat cluster yang diinginkan ke dalam sistem lalu sistem akan menampilkan hasil data yang sudah diolah sebelumnya.

b. DFD Level 1

DFD level 1 adalah suatu gambaran tentang informasi jalannya alur dari sebuah sistem yang dijelaskan lebih detail atau spesifik dan merupakan tahap lanjutan dari proses sebelumnya.

DFD level 1 dalam rancangan sistem ini adalah sebagai berikut :



Gambar 4.3 DFD Level 1

Peneliti menggambarkan DFD level 1 pada sistem ini lebih detail daripada diagram konteks. Berdasarkan gambar 4.3 proses yang dijalankan oleh sistem terbagi menjadi 4 proses inti yaitu proses import file, proses mengubah atribut ke angka, proses

algoritma K-Means Clustering, dan proses verifikasi titik pusat cluster atau *centroid*. Penjelasan lebih lengkap untuk setiap proses inti tersebut sebagai berikut :

1. *Proses Import File*

Pada proses pertama dari sistem K-Means Clustering ini adalah mengimportkan *file excel* ke dalam database dengan nama tabel yaitu dataset. Pada proses ini *file excel* yang bisa di importkan adalah *file excel* dengan format 97-2003 dengan isi data sudah dilakukan proses *pre-procesing* kecuali mengubah atribut data ke bentuk angka / transformation data. Didalam proses ini setiap kali akan mengimportkan *file excel* baru sistem akan menghapus *file excel* yang sudah diimportkan sebelumnya karena akan terjadi penumpukan data apabila tidak dihapus terlebih dahulu. Setelah *file excel* berhasil diimportkan akan muncul pemberitahuan bahwa *file excel* berhasil diimportkan.

2. *Proses Mengubah Data Atribut ke Nilai Angka*

Pada proses ini *user* atau pengguna cukup menekan tombol '*proses selanjutnya*' pada sistem untuk masuk ke proses kedua dalam sistem K-Means Clustering. Dalam proses ini sistem akan melakukan salah satu tahapan *pre-processing* data yaitu pengubahan data atribut ke dalam nilai angka. Sistem akan mengambil data atribut dari tabel database dengan nama

dataset lalu data atribut akan dihitung jumlah datanya (frekuensi) sesuai dengan isi data atribut tersebut kemudian memberi inisial dimulai dari frekuensi tertinggi ke frekuensi terendah. Setelah itu data atribut, nilai frekuensi, dan inisial akan diinputkan ke dalam tabel database sesuai jumlah atribut yang di proses. Dalam hal ini atribut yang berisikan data selain angka hanya ada 2 yaitu kota asal dan asal sekolah, maka proses pengubahan data ke nilai angka ini hanya akan disimpan kedalam tabel database *frek_kota* dan *frek_asek*. Setelah data berhasil disimpan dalam tabel database, sistem akan menampilkan hasil proses kepada *user / pengguna*.

3. Proses K-Means Clustering

Setelah langkah menampilkan hasil data dari proses sebelumnya (proses mengubah atribut kedalam nilai angka), *user / pengguna* tinggal menekan tombol '*proses selanjutnya*' untuk dapat melanjutkan ke proses K-Means Clustering. Berdasarkan gambar 4.3 perancangan proses sistem terdapat dua proses yaitu proses K-Means Clustering dan proses verifikasi. Dua proses tersebut saling berkaitan dan berhubungan satu sama lain.

Proses pertama adalah proses K-Means Clustering. Dalam proses tersebut sistem akan mengambil data dari tabel database *dataset*, *frek_kota*, dan *frek_asek* kemudian sistem akan

menampilkan sebuah tabel yang berisikan data gabungan dari tabel database tersebut. Tabel tersebut terdiri dari data atribut kota asal dan asal sekolah beserta inisial atribut masing dan data IPK mahasiswa. Setelah itu *user* atau pengguna akan diminta untuk mengisi nilai titik pusat cluster (*centroid*) yang selanjutnya akan disimpan ke dalam tabel database bernama *centroid_lama*. Sistem K-Means Clustering akan melakukan 3 aktivitas sekaligus yaitu menghitung jarak antara data ke cluster (*distance*), mengelompokkan data kedalam cluster berdasarkan nilai jarak terdekat data ke setiap cluster atau aktivitas clustering data, dan menghitung nilai titik pusat cluster baru. Hasil dari 3 aktivitas tersebut akan ditampilkan kepada *user* dan hanya aktivitas clustering data yang disimpan kedalam database proses sedangkan hasil dari menghitung nilai *centroid* baru akan menjadi data inputan untuk proses ke 4 yaitu verifikasi *centroid*.

4. *Proses Verifikasi Titik Pusat Cluster*

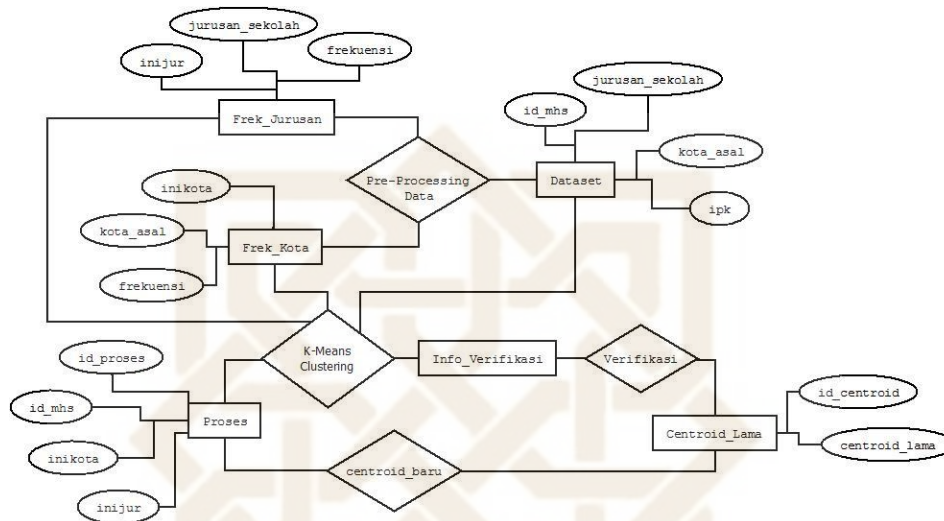
Proses verifikasi *centroid* adalah proses untuk menentukan apakah proses K-Means Clustering akan diulangi atau tidak. Kunci dari proses verifikasi ini adalah berdasarkan titik pusat cluster baru. Inputan data proses ini adalah *centroid* baru yang didapat dari proses K-Means Clustering yang selanjutnya akan disimpan dalam tabel database *centroid_lama*. Logika dari

proses ini adalah jika jumlah data pada tabel database centroid_lama bertambah setelah penginputan *centroid* baru maka proses K-Means Clustering harus diulang lagi dengan menggunakan *centroid* baru sebagai inputannya, sedangkan apabila jumlah data dari tabel database centroid_lamatidak bertambah setelah penginputan *centroid* baru maka artinya adalah *centroid* baru yang diinputkan sama dengan *centroid* sebelumnya dan proses K-Means Clustering selesai. Sistem menampilkan setiap perulangan proses K-Means Clustering kepada *user/pengguna*. Selanjutnya setelah proses K-Means selesai maka sistem akan menampilkan hasil kepada *user/pengguna* dari proses terakhir dalam serangkaian perulangan proses K-Means Clustering.

2. Perancangan ERD

Perancangan ERD atau Entity Relationship Diagram adalah perancangan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. Dalam ERD pemodelan struktur data dan hubungan antar setiap data digambarkan dengan beberapa notasi dan simbol yaitu entitas, atribut, relasi atau hubungan, dan sebuah garis

untuk menghubungkan relasi dan entitas. Perancangan ERD dalam sistem ini adalah sebagai berikut:



Gambar 4.4 ERD (Entity Relationship Diagram)

Berdasarkan rancangan ERD pada gambar 4.4 menjelaskan bahwa proses relasi antar entitas dalam sistem ini adalah entitas dataset yang diimport oleh *user*/pengguna memiliki 4 atribut yaitu *id_mhs*, *asal_sekolah*, *kota_asal*, dan *IPK*. Entitas dataset memiliki relasi dengan tiga entitas lainnya yaitu *frek_asek*, *frek_kota*, dan *proses*. Hubungan antara dataset dengan *frek_asek* adalah *pre-processing* data, yaitu aktivitas untuk mengubah data atribut *asal_sekolah* pada entitas dataset kedalam data yang berupa angka. Hubungan antara dataset dengan *frek_kota* adalah *pre-processingdata*, yaitu aktivitas untuk mengubah data atribut *kota_asal* pada entitas dataset kedalam data yang berupa angka. Kemudian relasi atau hubungan antara dataset, *frek_asek*, dan

frek_kota dengan proses adalah K-Means Clustering, yaitu atribut pada entitas dataset akan diolah kedalam sistem berdasarkan atribut asal_sekolah dan kota_asal yang dihubungkan dengan entitas frek_asek dan kota_asal untuk mendapatkan inisial dari masing-masing atribut. Kemudian hubungan antara entitas proses dengan entitas centroid_lama adalah proses menentukan *centroid* baru, yaitu *centroid* lama akan mendapatkan nilai *centroid* baru dari proses dan selanjutnya akan diverifikasi apakah nilai *centroid* baru sama dengan nilai *centroid* lama.

3. Perancangan Tabel Database

Sistem aplikasi K-Means Clustering ini memiliki beberapa tabel database yang digunakan untuk mendukung berjalannya sistem aplikasi K-Means Clustering. Database yang dipakai menggunakan database MySQL. Tabel – tabel tersebut berada di dalam satu database yaitu database “bayu”. Database tersebut memiliki 6 tabel pendukung sistem. Tabel – tabel yang dimaksud sebagai berikut:

a. Tabel Dataset

Tabel dataset adalah tabel yang berisikan tentang file yang diimportkan pertama kali oleh *user/pengguna* kedalam sistem. Tabel dataset mempunyai 4 kolom yaitu *id_mhs*, *kota_asal*, *asal_sekolah*, dan *ipk*. Data pada tabel ini nantinya yang akan diolah menggunakan algoritma K-Means Clustering pada sistem aplikasi ini. Tabel dataset tersebut adalah sebagai berikut:

Tabel 4.10 Tabel Database Dataset

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
1	id_mhs	Int	11	Primary Key
2	kota_asal	Varchar	30	
3	asal_sekolah	Varchar	30	
4	Ipk	Decimal	10, 2	

b. Tabel Frek_Asal sekolah

Tabel *frek_asal* sekolah adalah sebuah tabel yang dibuat untuk menampung data hasil dari pengolahan salah satu proses *pre-processing* data yaitu *transformation data*/pengubahan data selain angka kedalam nilai angka dengan menghitung frekuensi data asal sekolah dari tabel database dataset dan memberi inisial dimulai dari frekuensi terbesar ke frekuensi terkecil. Dalam tabel ini berisikan beberapa kolom yaitu *injur* sebagai primary key, *asal_sekolah*, *frekuensi*, dan *otomatis*. Maksud nama kolom *injur* adalah singkatan dari inisial asal sekolah dan kolom tersebut berisikan inisial dari asal sekolah mahasiswa. Sedangkan kolom *otomatis* adalah data hasil perkalian antara frekuensi dikalikan dengan inisial. Maksud dari kolom *otomatis* ini adalah untuk mendukung fitur '*otomatis*' yang

terdapat pada form penginputan *centroid*/titik pusat cluster. Fitur otomatis ini adalah untuk mensetting titik pusat cluster secara otomatis tanpa membutuhkan *user* menginputkan *centroid* secara manual. Tabel *frek_asal* sekolah adalah sebagai berikut:

Tabel 4.11 Tabel Database *Frek_Asal* sekolah

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
1	<i>Ini_asek</i>	Int	11	Primary Key
2	<i>asal_sekolah</i>	Varchar	30	
3	Frekuensi	Int	11	
4	Otomatis	Int	11	

c. Tabel *Frek_Kota*

Tabel *frek_kota* adalah sebuah tabel yang dibuat untuk menampung data hasil dari pengolahan salah satu proses *pre-processing* data yaitu *transformation data*/pengubahan data selain angka ke dalam nilai angka dengan menghitung frekuensi data kota asal mahasiswa dari tabel database dataset dan memberi inisial dimulai dari frekuensi terbesar ke frekuensi terkecil. Dalam tabel ini berisikan beberapa kolom yaitu *inikota* sebagai primary key, *kota_asal*, frekuensi, dan otomatis. Maksud nama kolom *inikota* adalah singkatan dari inisial kota dan kolom tersebut berisikan inisial dari kota asal mahasiswa. Sedangkan kolom otomatis adalah data hasil perkalian antara frekuensi dikalikan dengan inisial. Maksud

dari kolom otomatis ini adalah untuk mendukung fitur ‘*otomatis*’ yang terdapat pada form penginputan *centroid* / titik pusat cluster. fitur otomatis ini adalah untuk mensetting titik pusat cluster secara otomatis tanpa membutuhkan *user* menginputkan *centroid* secara manual. Tabel *frek_kota* adalah sebagai berikut:

Tabel 4.12 Tabel Database *Frek_Kota*

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
1	Inikota	Int	11	Primary Key
2	kota_asal	Varchar	30	
3	Frekuensi	Int	11	
4	Otomatis	Int	11	

d. Tabel *Centroid_lama*

Tabel *centroid_lama* adalah sebuah tabel yang dibuat untuk menyimpan *datacentroid* yang diinputkan *user*/pengguna maupun yang diinputkan secara otomatis kedalam sistem aplikasi K-Means Clustering. Tabel *centroid_lama* ini juga berfungsi untuk menentukan dan memverifikasi apakah proses algoritma K-Means akan diulang atau tidak berdasarkan *centroid* baru yang diinputkan kedalam tabel database *centroid_lama*. Tabel tersebut mempunyai 2 kolom yaitu kolom *id_centroid* sebagai *primary key* dan kolom *centroid_lama* dengan memiliki *index* adalah *unique* yang berarti data di dalam kolom tersebut tidak bisa menyimpan data yang

bernilai sama. Perancangan tabel centroid_lama adalah sebagai berikut:

Tabel 4.13 Tabel Database Centroid_lama

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
1	id_centroid	Int	11	Primary Key
2	centroid_lama	Decimal	10, 7	Unique

e. Tabel Nomor

Tabel nomor ini adalah sebuah tabel yang berfungsi sebagai pendukung sistem aplikasi K-Means Clustering yaitu untuk menyimpan dan menampilkan jumlah proses perulangan algoritma K-Means yang telah diproses sistem. Tabel nomor ini memiliki 2 kolom yaitu id_nomor sebagai *primary key* dan kolom nomor. Tabel tersebut adalah sebagai berikut :

Tabel 4.14 Tabel Database Nomor

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
1	id_nomor	Int	11	Primary Key
2	Nomor	Int	11	

f. Tabel Proses

Tabel proses adalah suatu tabel database yang dibuat untuk menyimpan setiap perulangan proses algoritma K-Means Clustering. Sistem akan menghapus data yang diinputkan pada tabel proses setiap terjadi perulangan proses algoritma tersebut maka data yang berada pada tabel proses ini adalah data paling baru dari setiap perulangan algoritma K-Means clustering. Di dalam tabel database proses memiliki beberapa kolom yaitu id_proses sebagai *primary key*, id_mhs sebagai *foreign key*, inikota sebagai *foreign key*, injur sebagai *foreign key*, ipk, clustersatu, clusterdua, dan clustertiga. Kolom id_mhs, inikota, dan iniasek adalah kolom yang mempunyai relasi / hubungan dengan tabel lainnya yaitu tabel dataset, tabel frek_kota, dan tabel frek_asek. Sedangkan kolom clustersatu, clusterdua, dan clustertiga adalah kolom pendukung yang berfungsi sebagai penanda untuk pengelompokan data/clustering data. Tabel proses ini adalah sebagai berikut:

Tabel 4.15 Tabel Database Proses

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
1	id_proses	Int	11	Primary Key
2	id_mhs	Int	11	Foreign Key
3	Inikota	Int	11	Foreign Key

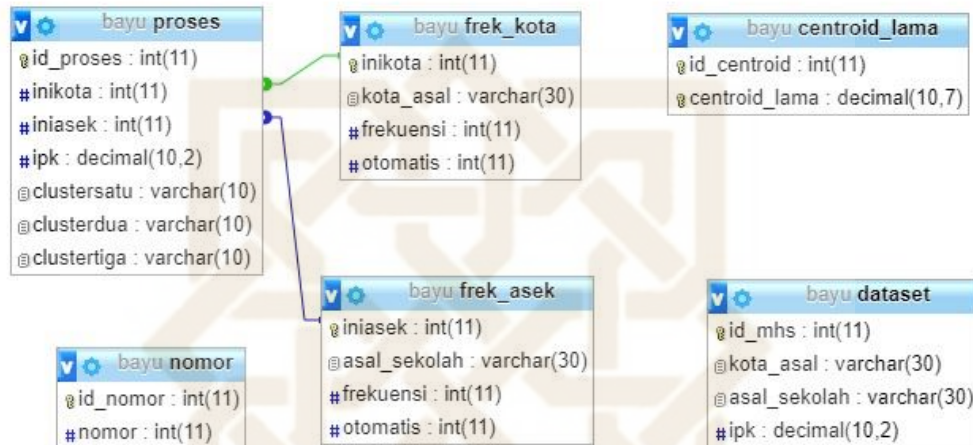
Tabel 4.15 Tabel Database Proses (lanjutan)

NO	Nama Kolom	Tipe data	Ukuran / Value	Keterangan
4	Iniasek	Int	11	Foreign Key
5	Ipk	decimal	10, 2	
6	Clustersatu	Varchar	10	
7	Clusterdua	Varchar	10	
8	Clustertiga	Varchar	10	

g. Relasi Database

Relasi database adalah hubungan antara satu tabel database dengan tabel database lain yang masih berada di dalam satu database dan berfungsi agar data dalam tabel database dapat saling berhubungan dan memudahkan peneliti untuk menggabungkan dan menampilkan data dari tabel yang berbeda beda. Hubungan antar tabel database tersebut ada tiga macam yaitu one-to-one (satu baris data hanya bisa dihubungkan dengan satu baris data pada tabel relasinya), one-to-many (satu baris data bisa dihubungkan dengan satu atau lebih baris data pada tabel relasinya), dan many-to-many

(satu atau lebih baris data bisa dihubungkan dengan satu atau lebih baris data pada tabel relasinya). Perancangan relasi database pada sistem aplikasi K-Means Clustering ini adalah sebagai berikut:

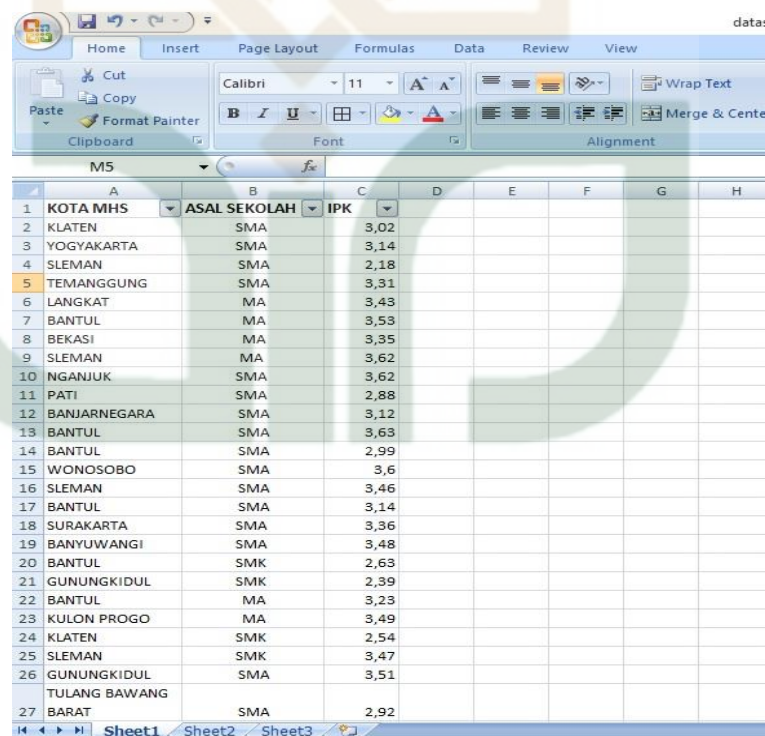


Gambar 4.5 Relasi Database (Sistem K-Means Clustering)

Berdasarkan relasi database pada gambar 4.5 dapat diketahui bahwa tabel proses mempunyai relasi dengan tabel frek_kota, frek_asek, dan dataset. Sedangkan table centroid_lama tidak mempunyai hubungan dengan tabel lain karena tabel centroid_lama adalah hanya sebagai perantara untuk memverifikasi apakah nilai *centroid* baru yang didapatkan mempunyai nilai yang sama dengan nilai *centroid* lama. Sama halnya dengan tabel nomor yaitu hanya sebagai wadah sementara untuk menampilkan nomor proses pada setiap proses yang dialami oleh sistem.

4.5.5 Implementasi Sistem Aplikasi K-Means Clustering

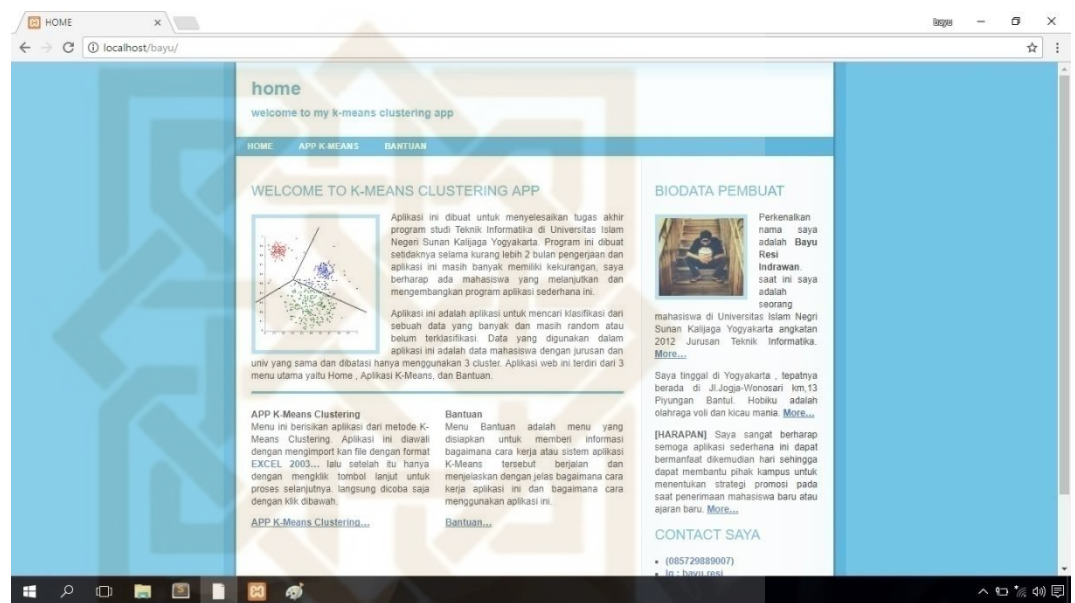
Implementasi sistem adalah menerapkan perancangan sistem DFD, perancangan ERD, dan perancangan database yang sudah dibuat sebelumnya ke dalam sebuah kesatuan sistem yang diberi nama sistem aplikasi K-Means Clustering. Sistem ini memiliki fungsi untuk menerapkan algoritma K-Means Clustering agar dapat membantu pihak universitas untuk menentukan strategi promosi yang akan dipakai berdasarkan data hasil dari sistem ini. Data yang dipakai untuk implementasi sistem ini adalah data yang sama persis dengan data yang dipakai sebelumnya untuk menerapkan algoritma K-Means Clustering menggunakan *Microsoft Excel* 2010. Contoh *file excel* yang akan diimportkan kedalam sistem:



	A	B	C	D	E	F	G	H
	KOTA MHS	ASAL SEKOLAH	IPK					
2	KLATEN	SMA	3,02					
3	YOGYAKARTA	SMA	3,14					
4	SLEMAN	SMA	2,18					
5	TEMANGGUNG	SMA	3,31					
6	LANGKAT	MA	3,43					
7	BANTUL	MA	3,53					
8	BEKASI	MA	3,35					
9	SLEMAN	MA	3,62					
10	NGANJUK	SMA	3,62					
11	PATI	SMA	2,88					
12	BANJARNEGARA	SMA	3,12					
13	BANTUL	SMA	3,63					
14	BANTUL	SMA	2,99					
15	WONOSOBO	SMA	3,6					
16	SLEMAN	SMA	3,46					
17	BANTUL	SMA	3,14					
18	SURAKARTA	SMA	3,36					
19	BANYUWANGI	SMA	3,48					
20	BANTUL	SMK	2,63					
21	GUNUNGKIDUL	SMK	2,39					
22	BANTUL	MA	3,23					
23	KULON PROGO	MA	3,49					
24	KLATEN	SMK	2,54					
25	SLEMAN	SMK	3,47					
26	GUNUNGKIDUL	SMA	3,51					
27	TULANG BAWANG BARAT	SMA	2,92					

Gambar 4.6 Contoh *File excel* Yang Diimport Kedalam Sistem

1. Halaman Depan Sistem Aplikasi K-Means Clustering



Gambar 4.7 Halaman Depan Sistem K-Means Clustering

Halaman depan Sistem Aplikasi K-Means Clustering berisikan tentang tujuan sistem ini dibuat, keterangan tentang menu yang tersedia pada sistem ini, biodata tentang pembuat sistem (peneliti) serta kontak dari peneliti. Pada sistem ini tersedia tiga menu utama yaitu *Home*, *App K-Means*, dan *Bantuan*.

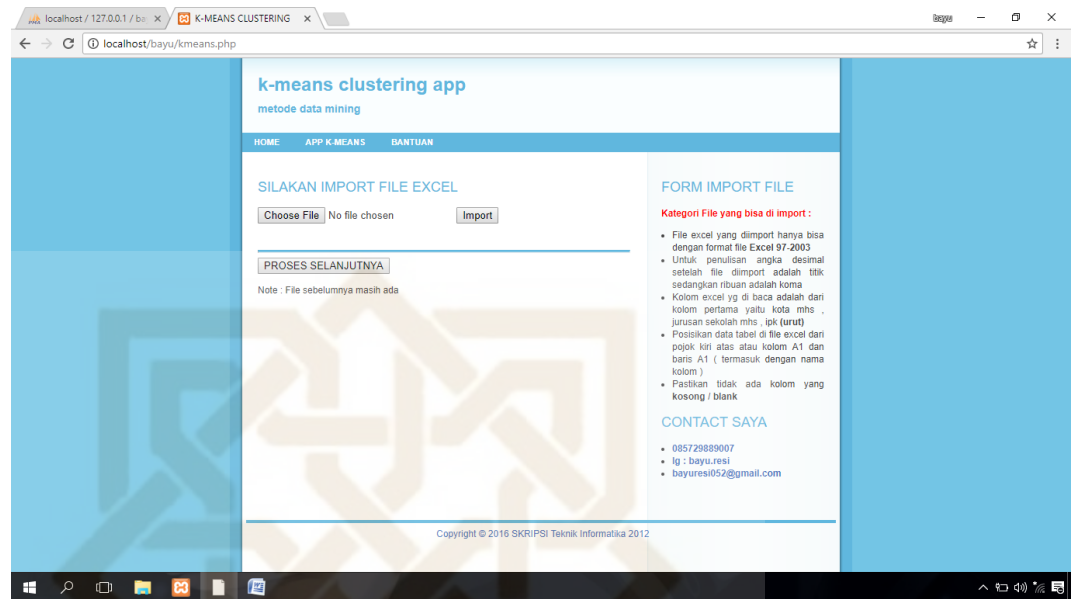
2. Halaman Bantuan



Gambar 4.8 Halaman Bantuan Sistem K-Means Clustering

Halaman bantuan adalah salah satu menu yang terdapat pada sistem aplikasi K-Means Clustering. Halaman ini akan menampilkan sebuah bantuan dan petunjuk mengenai bagaimana cara kerja dan cara pemakaian sistem aplikasi K-Means Clustering ini. Apabila *user* kurang paham dengan cara kerja sistem tersebut *user* bisa menghubungi kontak peneliti yang tersedia pada halaman bantuan dan halaman depan.

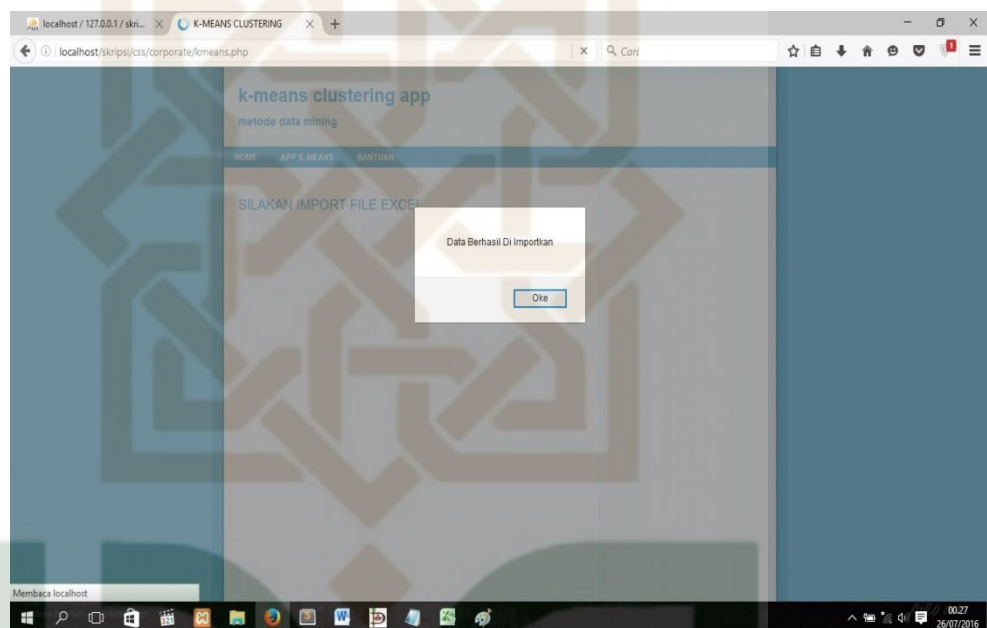
3. Halaman APP K-Means Clustering



Gambar 4.9 Halaman Aplikasi K-Means Clustering

Halaman aplikasi K-Means Clustering adalah halaman awal dari serangkaian proses sistem aplikasi K-Means Clustering ini. Pada halaman ini *user* akan diminta untuk mengimportkan *file excel* dengan berisikan sejumlah data yang selanjutnya akan diolah menggunakan algoritma K-Means Clustering di sistem ini. Pada setiap pengimportkan *file excel* baru sistem akan menghapus *file excel* yang sebelumnya sudah diimport kedalam database jika ada. Pada sisi sebelah kanan halaman akan disertakan keterangan bantuan tentang kriteria *file excel* yang bisa diimportkan kedalam sistem. File yang akan peneliti import adalah *file excel* yang sudah disiapkan sebelumnya, seperti yang sudah dicontohkan pada Gambar 4.6. Proses mengimportkan file ini membutuhkan waktu yang cukup lama tergantung data yang

diimportkan kedalam sistem. Setelah file tersebut berhasil diimportkan kedalam sistem akan ada pemberitahuan bahwa file sudah berhasil diimportkan. Jika sudah ada pemberitahuan bahwa *file excel* berhasil diimport langkah selanjutnya adalah *user* menekan tombol '*proses selanjutnya*' untuk melanjutkan ke proses berikutnya.



Gambar 4.10 Halaman Berhasil Import *File excel*

4. Halaman Transformation Data

The screenshot shows a web browser window with the URL `localhost/bayu/frek.php`. The page title is "k-means clustering app" and the subtitle is "metode data mining". The navigation menu includes "HOME", "APP K-MEANS", and "BANTUAN". The main content area is titled "MEMODIFIKASI ATRIBUT SELAIN ANGKA" and "PROSES SELANJUTNYA".

NO	KOTA ASAL	FREK	INISIAL
1	BANTUL	35	1
2	SLEMAN	23	2
3	YOGYAKARTA	12	3
4	MAGELANG	11	4
5	BANJARNEGARA	7	5
6	BREBES	7	6
7	CLACAP	7	7
8	SUNJUNGKIDUL	7	8
9	KEBUMEN	7	9
10	KULON PROGO	7	10
11	KLATEN	6	11
12	NGAWI	6	12
13	TEMANGGUNG	6	13
14	TEGAL	5	14
15	WONOGIRI	5	15
16	BIMA	4	16
17	GRESIK	4	17
18	MAGETAN	4	18
19	PATI	4	19
20	PONOROGO	4	20
21	BANWIMAS	3	21

NO	ASAL SEKOLAH	FREK	INISIAL
1	SMA	145	1
2	MA	85	2
3	SMK	59	3
4	PON	3	4
5	SMTI	2	5

TRANSFORMATION DATA

Proses ini adalah salah satu tahap Pre-processing atau tahap sebelum diproses yaitu menyiapkan data agar bisa diolah.

Atribut yang digunakan adalah asal kota dan asal jurusan sewaktu SMA yang mana isi data dari atribut tersebut adalah huruf atau selain angka, karena didalam proses K-Means ini data yang bisa diolah adalah hanya data berupa angka maka atribut tersebut dimodifikasi kedalam angka. Untuk mengubah kedua atribut ini sesuai dengan frekuensi terbesar data sampai frekuensi terkecil data dan memberi inisial yang berurutan sesuai frekuensi yang terbentuk.

Seperiti contoh dalam data yang diimport ada beberapa kota yang sama maka data kota yang sama dijadikan satu dan dihitung jumlah mahasiswanya.

Dalam contoh mahasiswa yang berasal dari kota Bantul ada 147 mahasiswa dan jumlah itu adalah jumlah tersebut daripada jumlah mahasiswa dari kota

Gambar 4.11 Halaman Transformation Data (1)

Halaman transformation data adalah halaman yang menampilkan proses pengubahan atribut selain angka ke bentuk nilai angka. Pada sisi sebelah kanan terdapat kolom yang berisikan tentang penjelasan proses transformation data tersebut. Proses transformation data adalah salah satu tahapan *pre-processing* data atau proses untuk menyiapkan data agar data dapat diolah dengan algoritma K-Means Clustering. Sistem tidak memproses semua tahapan *pre-processing* data karena menurut peneliti hanya proses transformation data saja yang membutuhkan waktu lama dalam pengerjaannya karena harus menghitung frekuensi data lalu mengurutkannya dari nilai terbesar dan memberi inisial. Setelah proses ini selesai dengan menggunakan data yang sebelumnya

sudah diimportkan ke dalam sistem peneliti mendapatkan hasil sebagai berikut :

- Pada atribut kota asal terdapat 112 kota asal mahasiswa yang berbeda dengan frekuensi terbesar adalah 35 data sedangkan frekuensi terkecil adalah 1 data.
- Pada atribut jurusan asal sekolah terdapat 5 asal sekolah mahasiswa yang berbeda dengan frekuensi terbesar adalah 90 data sedangkan frekuensi terkecil adalah 1 data.

Setelah proses ini selesai langkah selanjutnya adalah *user/pengguna* menekan tombol '*proses selanjutnya*' untuk melanjutkan ke proses berikutnya.

5. Halaman Dataset dan Form Penginputan Centroid

The screenshot shows a web browser displaying a 'k-means clustering app' interface. The page title is 'k-means clustering app' with the subtitle 'metode data mining'. The navigation menu includes 'HOME', 'APP K-MEANS', and 'BANTUAN'. The main content area is divided into two sections: 'DATASET YANG DIOLAH' and 'INPUTKAN TITIK PUSAT'.

The 'DATASET YANG DIOLAH' section shows a table with 19 records. The table has columns: NO, KOTA ASAL, INISIAL, JURUSAN, INISIAL, and IPK. The data is as follows:

NO	KOTA ASAL	INISIAL	JURUSAN	INISIAL	IPK
1	YOGYAKARTA	4	AKUNTANSI	2	3.88
2	BANTUL	1	AKUNTANSI	2	3.72
3	KULON PROGO	3	TEKNIK KOMPUTER DAN JARINGAN	6	3.88
4	BANTUL	1	ILMU PENGETAHUAN SOSIAL	1	3.01
5	BANTUL	1	AKUNTANSI	2	3.81
6	SLEMAN	2	TEKNIK GAMBAR BANGUNAN	17	3.57
7	SLEMAN	2	AKUNTANSI	2	3.11
8	BANTUL	1	AKUNTANSI	2	3.69
9	LAMPUNG	20	ILMU PENGETAHUAN SOSIAL	1	3.30
10	KULON PROGO	3	AKUNTANSI	2	3.57
11	BANTUL	1	AKUNTANSI	2	3.32
12	SLEMAN	2	AKUNTANSI	2	3.28
13	SLEMAN	2	ADMINISTRASI PERKANTORAN	4	3.46
14	BATANG	18	AKUNTANSI	2	3.33
15	SLEMAN	2	ILMU PENGETAHUAN SOSIAL	1	3.18
16	SLEMAN	2	AKUNTANSI	2	3.33
17	KULON PROGO	3	ILMU PENGETAHUAN SOSIAL	1	3.71
18	BANTUL	1	ILMU PENGETAHUAN ALAM	3	3.36
19	SLEMAN	2	AKUNTANSI	2	3.04

The 'INPUTKAN TITIK PUSAT' section contains three clusters, each with input fields for 'ASAL KOTA', 'JURUSAN SEKOLAH', and 'IPK'. There are also 'RANDOM' and 'PROSES' buttons at the bottom of the input section.

Gambar 4.13 Halaman Dataset (1)

Halaman dataset adalah halaman yang menampilkan sebuah tabel yang berisikan kota asal, inisial kota tersebut, asal sekolah, inisial asal sekolah tersebut, dan ipk mahasiswa. Jumlah data yang akan diolah adalah sejumlah 709 data mahasiswa. Pada sisi sebelah kanan terdapat sebuah form untuk menginputkan titik pusat cluster/*centroid* setiap cluster dan sebuah informasi tentang form penginputan titik pusat cluster tersebut. Untuk penginputan *centroid* ini sistem menawarkan kepada *user*/pengguna dua pilihan yaitu menginputkan *centroid* sesuai keinginan *user* atau dengan menggunakan fitur *random*. Fitur *random* tersebut adalah sebuah fitur untuk menentukan nilai *centroid* secara random atau acak yang akan ditentukan oleh sistem secara otomatis. Nilai tersebut diambil dari nilai rata – rata setiap atribut pada dataset untuk cluster 2 kemudian untuk cluster 1 dan cluster 3 ditentukan secara random dan berubah ubah apabila di *refresh*. Pada implementasi saat ini peneliti memakai dua *centroid* yang berbeda yaitu dengan *centroid* yang pertama adalah nilai *centroid* yang sama dengan implementasi dengan *excel* sedangkan *centroid* kedua adalah menggunakan fitur *random* karena fitur tersebut mempermudah peneliti untuk menentukan *centroid* di setiap cluster.

The screenshot shows a web browser window with the URL `localhost/skrpa/css/corporate/dataset.php?kd=otomatis`. The page title is "k-means clustering app" and the subtitle is "metode data mining". The main content area is divided into two sections: "DATASET YANG DIOLAH" and "INPUTKAN TITIK PUSAT 'RANDOM'".

The "DATASET YANG DIOLAH" section displays a table with 19 records of student data. The table has columns: NO, KOTA ASAL, INISIAL, JURUSAN, INISIAL, and IPK. The "INPUTKAN TITIK PUSAT 'RANDOM'" section contains three tables for Cluster 1, Cluster 2, and Cluster 3, each with input fields for ASAL KOTA, JURUSAN SEKOLAH, and IPK. A "PROSES" button is located below these tables. A red box highlights a "Refresh" button and a message: "Refresh bisa centroid tidak selesai atau error".

NO	KOTA ASAL	INISIAL	JURUSAN	INISIAL	IPK
1.	YOGYAKARTA	4	AKUNTANSI	2	3.88
2.	BANTUL	1	AKUNTANSI	2	3.72
3.	KULON PROGO	3	TEKNIK KOMPUTER DAN JARINGAN	6	3.88
4.	BANTUL	1	ILMU PENGETAHUAN SOSIAL	1	3.01
5.	BANTUL	1	AKUNTANSI	2	3.81
6.	SLEMAN	2	TEKNIK GAMBAR BANGUNAN	17	3.57
7.	SLEMAN	2	AKUNTANSI	2	3.11
8.	BANTUL	1	AKUNTANSI	2	3.69
9.	LAMPUNG	20	ILMU PENGETAHUAN SOSIAL	1	3.30
10.	KULON PROGO	3	AKUNTANSI	2	3.57
11.	BANTUL	1	AKUNTANSI	2	3.32
12.	SLEMAN	2	AKUNTANSI	2	3.28
13.	SLEMAN	2	ADMINISTRASI PERKANTORAN	4	3.40
14.	BATANG	18	AKUNTANSI	2	3.33
15.	SLEMAN	2	ILMU PENGETAHUAN SOSIAL	1	3.18
16.	SLEMAN	2	AKUNTANSI	2	3.33
17.	KULON PROGO	3	ILMU PENGETAHUAN SOSIAL	1	3.71
18.	BANTUL	1	ILMU PENGETAHUAN ALAM	3	3.30
19.	SLEMAN	2	AKUNTANSI	2	3.02

Gambar 4.15 Menentukan *Centroid*

Berdasarkan Gambar 4.15 nilai *centroid* akan terisi secara random. Peneliti mensetting nilai *centroid* tersebut berbeda dengan nilai *centroid* awal yang digunakan pada penerapan algoritma K-Means Clustering menggunakan aplikasi *Microsoft Excel* 2010. Untuk melanjutkan ke proses selanjutnya *user/pengguna* menekan tombol *proses* untuk melanjutkan ke proses berikutnya sedangkan tombol *refresh* apabila *user* menginginkan untuk mendapatkan nilai *centroid* yang berbeda secara acak.

6. Halaman Proses K-Means Clustering

The screenshot shows a web browser window with the URL `localhost/bayu/proses.php`. The page title is "k-means clustering app" and the subtitle is "metode data mining". There are navigation links for "HOME", "APP K-MEANS", and "BANTUAN". The main content area is titled "PROSES 1" and displays "RECORD : 294 DATA MAHASISWA". Below this is a table with the following data:

NO	KOTA ASAL	ASAL SEKOLAH	IPK	CENTROID C1	CENTROID C2	CENTROID C3	C1	C2	C3
1	11	1	3.02	9.1067053	17.0531522	101.0824223	v		
2	3	1	3.14	1.8111046	25.0407747	109.0755523	v		
3	2	1	2.18	0.5500000	26.0192928	110.0850094	v		
4	13	1	3.31	11.1275514	15.0803216	99.0821397	v		
5	72	2	3.43	70.0302792	44.0194968	40.1143386		v	
6	1	2	3.53	2.3685439	27.0367916	111.0409384	v		
7	31	2	3.35	29.0681682	3.2423603	81.0566976	v		
8	2	2	3.62	2.2271262	26.0432333	110.0411019	v		
9	87	1	3.62	85.0232915	59.0275350	25.3188487		v	
10	19	1	2.88	17.0458939	9.0872218	93.0908293	v		
11	5	1	3.12	3.3496418	23.0434372	107.0770942	v		
12	1	1	3.63	2.2360690	27.0600744	111.0722288	v		
13	1	1	2.99	1.8880758	27.0325156	111.0752250	v		
14	44	1	3.60	42.0461758	16.0993913	68.1179338	v		
15	2	1	3.46	1.8300000	26.0537061	110.0733251	v		
16	1	1	3.14	1.8111046	27.0377588	111.0741919	v		
17	104	1	3.38	102.0146700	76.0166929	8.9566121		v	
18	30	1	3.48	28.0510495	2.6171740	82.0982491	v		
19	1	3	2.63	2.4494897	27.0233251	111.0245018	v		
20	8	3	2.39	8.3700549	20.0269845	104.0291959	v		

Gambar 4.16 Halaman Proses K-Means Clustering (1)

Halaman proses adalah halaman setelah *user*/pengguna menekan tombol proses pada form penginputan titik pusat cluster atau *centroid*. Setelah *user* menekan tombol proses maka sistem akan otomatis memproses dataset menggunakan algoritma K-Means Clustering. Proses yang dilakukan adalah menghitung jarak data ke setiap cluster, mengelompokkan data ke setiap cluster berdasarkan jarak terdekat data ke cluster, dan menghitung nilai titik pusat cluster baru. Sistem akan menampilkan setiap hasil dari proses – proses tersebut kedalam sebuah tabel. Dalam tahap ini setiap proses tersebut hampir bekerja secara bersamaan dan menggunakan dataset yang jumlah datanya cukup banyak sehingga tahap ini memerlukan waktu yang cukup lama yaitu sekitar 2 menit 15 detik pada setiap prosesnya. Hal tersebut bisa

dikatakan adalah kelemahan dari sistem K-Means Clustering yang dibuat oleh peneliti.

RECORD : 294 DATA MAHASISWA

NO	KOTA ASAL	ASAL SEKOLAH	IPK	CENTROID C1	CENTROID C2	CENTROID C3	C1	C2	C3
1.	11	1	3.02	4.3883992	25.3466885	77.0034504	V		
2.	3	1	3.14	3.7705626	33.3425490	85.0028785	V		
3.	2	1	2.18	4.8855690	34.3612185	86.0094875	V		
4.	13	1	3.31	6.3570087	23.3459903	75.0031840	V		
5.	72	2	3.43	65.3119274	35.6676269	16.0039586		V	
6.	1	2	3.53	5.6973941	35.3346384	87.0009866	V		
7.	31	2	3.35	24.3125987	5.3387115	57.0009220		V	
8.	2	2	3.02	4.7040749	34.3352994	86.0013328	V		
9.	87	1	3.02	80.3155422	50.6731861	1.2885624		V	
10.	19	1	2.88	12.3428159	17.3559895	68.0044694	V		
11.	5	1	3.12	1.8629259	31.3432717	83.0029780	V		
12.	1	1	3.03	5.7473077	35.3429815	87.0035439	V		
13.	1	1	2.99	5.7489411	36.3431929	87.0031401	V		
14.	44	1	3.60	37.3201617	7.7087414	44.0067615		V	
15.	2	1	3.46	4.7517590	34.3419284	86.0030121	V		
16.	1	1	3.14	5.7420952	35.3420256	87.0028123	V		
17.	104	1	3.36	97.3143745	67.6709715	18.0152148		V	
18.	30	1	3.48	23.3243087	6.3803245	58.0045402	V		
19.	1	3	2.63	5.8633684	35.3621339	87.0121075	V		
20.	8	3	2.39	2.0320679	28.3762508	80.0154917	V		

CLUSTER 1

ASAL KOTA	7.2414009
ASAL SEKOLAH	1.7529000
IPK	3.3281030

CLUSTER 2

ASAL KOTA	57.5429000
ASAL SEKOLAH	1.7714000
IPK	3.3274290

CLUSTER 3

ASAL KOTA	87.5000000
ASAL SEKOLAH	1.7000000
IPK	3.2580000

Next Proses

KETERANGAN :

KOTA ASAL = Inisial Kota
ASAL SEKOLAH = Inisial Asek

CLUSTER 1 = Cluster 1

Maksud dari centang tersebut adalah

Gambar 4.17 Halaman Proses K-Means Clustering (2)

Setelah proses K-Means Clustering selesai maka sistem akan menampilkan perhitungan nilai titik pusat cluster / *centroid* baru dan keterangan maupun petunjuk tentang halaman proses ini pada sisi sebelah kanan halaman. Karena nilai *centroid* baru yang didapatkan mempunyai nilai yang berbeda dengan nilai *centroid* yang lama maka proses K-Means Clustering harus dilanjutkan lagi dengan menekan tombol *next proses*. Dalam halaman proses ini sistem akan terus melakukan perulangan proses K-Means Clustering sampai mendapatkan nilai titik pusat cluster baru yang sama dengan nilai titik pusat cluster lama. Dan sistem akan menampilkan hasil algoritma K-Means Clustering disetiap proses perulangannya.

The screenshot shows a web browser displaying a K-Means clustering application. The page title is "k-means clustering app" and the subtitle is "metode data mining". The application is running on a local host at the URL "localhost/dayu/proses.php".

The main content area is titled "PROSES 5" and indicates "RECORD : 294 DATA MAHASISWA". It displays a table with the following columns: NO, KOTA ASAL, ASAL SEKOLAH, IPK, CENTROID C1, CENTROID C2, CENTROID C3, C1, C2, and C3. The table contains 20 rows of data, representing individual students and their attributes.

To the right of the table, there is a section titled "PUSAT CLUSTER BARU" (New Cluster Center) which lists the results for three clusters:

- CLUSTER 1:**
 - ASAL KOTA: 7.4915000
 - ASAL SEKOLAH: 1.7458000
 - IPK: 3.3280230
- CLUSTER 2:**
 - ASAL KOTA: 38.2388000
 - ASAL SEKOLAH: 1.7910000
 - IPK: 3.3278120
- CLUSTER 3:**
 - ASAL KOTA: 87.5000000
 - ASAL SEKOLAH: 1.7000000
 - IPK: 3.2580000

Below the cluster results, the text "PROSES SELESAI !!" (Process Completed !!) is displayed, followed by a "hasil proses k-means" (K-Means process result) button. A "HASIL" (Result) button is also visible. A "KETERANGAN:" (Explanation) section at the bottom indicates "KOTA ASAL = Inisial Kota" (Home City = Initial City).

Gambar 4.18 Halaman Proses K-Means Clustering (3)

Setelah membutuhkan 18 kali proses perulangan K-Means Clustering akhirnya peneliti mendapatkan nilai titik pusat cluster baru dan nilai titik pusat cluster lama yang bernilai sama. Ketika *centroid* baru yang muncul sama dengan nilai *centroid* lama maka sistem otomatis akan menghentikan proses perulangan dalam algoritma K-Means Clustering. Untuk mendapatkan hasil dari proses tersebut *user/pengguna* menekan tombol *hasil* untuk proses berikutnya.

4.5.6 Hasil Implementasi Sistem Aplikasi K-Means Clustering

Pada halaman hasil proses K-Means Clustering sistem akan menampilkan jumlah perulangan proses K-Means, menampilkan data kota asal mahasiswa dan asal sekolah mahasiswa yang sudah dikelompokkan ke dalam setiap cluster beserta frekuensi dan rata – rata ipk mahasiswa pada setiap cluster. Pada sisi sebelah kanan terdapat keterangan mengenai hasil dari proses sistem aplikasi K-Means clustering ini. Selain itu terdapat fitur EXIT yang berfungsi sebagai tombol untuk keluar dari sistem aplikasi K-Means Clustering.

HASIL PROSES K-MEANS CLUSTERING

RECORD : 709 DATA MAHASISWA | 11 KALI PROSES |

CLUSTER 1					
CLUSTER 1 TERDIRI DARI 508 MAHASISWA YANG TERDIRI DARI					
ASAL KOTA		ASAL SEKOLAH			
KOTA	JUMLAH	ASAL SEKOLAH JUMLAH			
BANTUL	75	SMA	295		
SLEMAN	63				
YOGYAKARTA	49				
MAGELANG	30				
KLATEN	25				
KEBUMEN	22				
GUNUNGGIDUL	19				
KULONPROGO	19				
PURWOREJO	19				
BANJARNEGARA	17			SMK	113
BREBES	17				
CILACAP	17				
NGAWI	14			MA	85
TEMANGGUNG	13				
PATI	12				
TEGAL	12			PON	11
WONOGIRI	11				
MAGETAN	8				
SRAGEN	8				
BANYUMAS	8				
BOYOLALI	7	SMTI	4		
BIMA	6				
JEPARA	6				
PACITAN	6				
BEKASI	5				
CIAMIS	5				
CIREBON	5				
GRESIK	5				
LAMONGAN	5				

RATA-RATA IPK CLUSTER 1 : 3.20

Gambar 4.19 Halaman Hasil Proses K-Means (cluster satu)

Berdasarkan hasil dari cluster 1 diatas dapat diketahui bahwa terdapat 508 data mahasiswa yang termasuk kedalam kelompok cluster 1 dengan nilai rata – rata IPK 3,20. Dalam cluster 1 tersebut terdapat informasi tentang persebaran wilayah kota asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah mahasiswa dari kota dan asal sekolah masing – masing tersebut. Jumlah data mahasiswa pada cluster 1 adalah jumlah data yang paling banyak diantara cluster yang lain sehingga dapat dikatakan bahwa mayoritas mahasiswa prodi teknik informatika paling banyak masuk kedalam kelompok cluster 1. Berdasarkan hal tersebut pihak prodi dapat mempertimbangkan strategi promosi di tahun depan menggunakan hasil persebaran wilayah dan asal sekolah pada cluster 1.



CLUSTER 2			
CLUSTER 2 TERDIRI DARI 138 MAHASISWA YANG TERDIRI DARI			
ASAL KOTA		ASAL SEKOLAH	
KOTA	JUMLAH		
PURBALINGGA	5		
SUKOHARJO	5		
BATANG	4		
BLORA	4		
DEMAK	4		
MADIUN	4		
OGAN KOMERING	4		
PEKALONGAN	4		
PONOROGO	4		
SEMARANG	4		
TASIKMALAYA	4		
WONOSOBO	4		
LOMBOK TENGAH	3		
MANDAILING NATAL	3		
REMBANG	3		
ROKAN HULU	3		
BANYUWANGI	3		
GROBOGAN	3		
JOMBANG	3		
KARAWANG	3		
LABUHANBATU UTARA	3		
LAMPUNG BARAT	3		
KUNINGAN	2		
LAMPUNG TIMUR	2		
MAJALENGKA	2		
METRO	2		
AMBON	2		
MUARA ENIM	2		
BALIKPAPAN	2		
MUSI BANYUASIN	2		
BANGKA	2		
PALANGKARAYA	2		
BLITAR	2		
PASURUAN	2		
DEPOK	2		
PEKANBARU	2		
HULU SUNGAI UTARA	2		
PONTIANAK	2		
INDRAMAYU	2		
SALATIGA	2		
JAKARTA TIMUR	2		
SUBANG	2		
JAKARTA UTARA	2		
SURAKARTA	2		
JEMBER	2		
TANGERANG	2		
KEDIRI	2		
TUBAN	2		
KENDAL	2		
TULUNGAGUNG	2		
TANGERANG	1		
ASAHAN	1		
		ASAL SEKOLAH	JUMLAH
		SMA	60
		MA	47
		SMK	23
		PON	8
RATA-RATA IPK CLUSTER 2 : 3.21			

Gambar 4.20 Halaman Hasil Proses K-Means (cluster dua)

Berdasarkan hasil dari cluster 2 dapat diketahui bahwa terdapat 138 data mahasiswa yang termasuk kedalam kelompok cluster 2 dengan nilai rata – rata IPK 3,21. Dalam cluster 2 tersebut terdapat informasi tentang persebaran wilayah kota asal mahasiswa maupun asal sekolah mahasiswa

beserta dengan jumlah mahasiswa dari kota dan asal sekolah masing – masing. Jumlah data mahasiswa pada cluster 2 adalah jumlah data yang kurang dari cluster 1 tetapi lebih dari cluster 3. Informasi yang menonjol dari cluster 2 adalah informasi mengenai persebaran asal sekolah mahasiswa yaitu banyaknya jumlah asal sekolah yang berbeda beda dengan frekuensi yang sangat sedikit. Berdasarkan hal tersebut dapat diketahui bahwa asal sekolah dengan frekuensi paling sedikit adalah asal sekolah yang kurang meminati untuk masuk di jurusan/prodi Teknik Informatika.



CLUSTER 3			
CLUSTER 3 TERDIRI DARI 62 MAHASISWA YANG TERDIRI DARI			
ASAL KOTA		ASAL SEKOLAH	
KOTA	JUMLAH		
BENGKALIS	1		
KARANGANYAR	1		
MAJENE	1		
SIAGI	1		
BENGGULU UTARA	1		
KARANGASEM	1		
MAKASSAR	1		
SIDDAHU	1		
BINTAN	1		
KENDARI	1		
MALANG	1		
SIMALUNGUN	1		
BOGOR	1		
KERTAK HANYAR	1		
MATAHAM	1		
SITUBONDO	1		
BONDOWOSO	1		
KUDUS	1		
MEDAN	1		
SUKABUMI	1		
BULELENG	1		
KUTAI	1	SMA	34
MUSI RAWAS	1		
SUMENEP	1		
BULUKUMBA	1		
LAPAI	1		
NGANJUK	1		
TAMBAK	1		
CIANJUR	1		
LAMPUNG SELATAN	1		
PANDEGLANG	1	MA	18
TANJ. JABUNG BARAT	1		
DENPASAR	1		
LAMPUNG TENGAH	1		
PEGADAN	1		
TANJUNG SALAI	1		
DHARMASRAYA	1		
LANGKAT	1		
PEMALANG	1		
TANJUNG PINANG	1		
DOMPU	1	SMK	11
LINGSUMBAWE	1		
PINRANG	1		
TRENGGALEK	1		
GAHUT	1		
LOMBOK TIMUR	1		
PONGSEWU	1		
TRIMULYA	1		
INDRAGIRI Hilir	1		
LUNAN NEGARA	1		
PONDI	1	PON	2
TULANG BAWANG BAHAI	1		
JAMBI	1		
LUBUKLINGGAU	1		
PURWAKANTA	1		
BATAM	1		
JOHOR BAHRU	1		
LUMAJANG	1		
HASANGGODU	1		
BAWANG SAKTI JAYA	1		
KAMIYAN	1		
LUIWU UTARA	1		
SARDLANGUN	1		
RATA-RATA IPK CLUSTER 3: 3,21			

Gambar 4.21 Halaman Hasil Proses K-Means (cluster tiga)

Berdasarkan hasil dari cluster 3 dapat diketahui bahwa terdapat 62 data mahasiswa yang termasuk kedalam kelompok cluster 3 dengan nilai rata – rata IPK 3,21. Dalam cluster 3 tersebut terdapat informasi tentang persebaran wilayah kota asal mahasiswa maupun asal sekolah mahasiswa beserta dengan jumlah mahasiswa dari kota dan asal sekolah masing – masing. Jumlah data mahasiswa pada cluster 3 adalah jumlah data yang paling sedikit dibandingkan dengan cluster yang lain. Informasi yang

menonjol dari cluster 3 adalah informasi mengenai persebaran wilayah kota asal mahasiswa yaitu banyaknya jumlah kota asal mahasiswa yang berbeda beda dengan frekuensi yang sangat sedikit. Berdasarkan hal tersebut dapat diketahui bahwa kota asal dengan frekuensi paling sedikit adalah kota yang kurang meminati untuk masuk di jurusan/prodi Teknik Informatika.

4.5.7 Analisa

Setelah dianalisa ditemukan perbedaan dalam jumlah proses. Menurut saya yang menjadikan perbedaan dalam jumlah proses yaitu nilai angka dibelakang koma pada titik pusat cluster. Jika di dalam proses Excel nilai dibelakang koma tidak dibatasi, sedangkan di dalam proses PHP nilai titik pusat cluster dibatasi menjadi dua angka dibelakang koma. Untuk hasil yang didapatkan dari kedua proses (Excel dan PHP) tidak ada perbedaan.

Hasil analisa yang lain ditemukan sebagian kota dari di Jawa Tengah berada di cluster kedua. Penyebab perbedaan ini terletak pada penentuan titik awal pusat cluster. Sehingga menyebabkan sebagian kota di Jawa Tengah berada di cluster kedua.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menerapkan metode algoritma K-Means Clustering menggunakan dua aplikasi yaitu *Microsoft Excel* 2010 dan Sistem Aplikasi K-Means Clustering yang dirancang dan dibuat oleh peneliti.
2. Strategi promosi bagi calon mahasiswa baru di UIN Yogyakarta prodi Teknik Informatika yang tepat sasaran adalah dengan mempertimbangkan kota asal mahasiswa dan asal sekolah mahasiswa berdasarkan hasil cluster yang sudah terbentuk sesuai potensi akademik mahasiswa. Jadi untuk sasaran promosi Teknik Informatika UIN Sunan Kalijaga ke depannya bisa dilakukan berdasarkan cluster 1 pada SMA yang berada di wilayah kota Bantul. Cluster 2 diperoleh sasaran untuk promosi dengan asal sekolah SMA yang berada di kota asal terbanyak yaitu Bekasi dan Ciamis. Dan cluster 3 berisikan 62 kota dengan jumlah paling sedikit dan asal sekolah terbanyak masih berasal dari SMA.
3. Dari penelitian yang telah diolah menggunakan metode k-means clustering dengan menggunakan aplikasi excel dan bahasa pemrograman PHP didapatkan hasil bahwa mahasiswa angkatan 2009 sampai 2016 yang dibagi menjadi 3 cluster dimana pada cluster 1 terdapat mahasiswa dengan asal kota dan asal sekolah terbanyak. Jumlah mahasiswa terbanyak di jurusan Teknik

Informatika berasal dari kota Bantul dengan asal sekolah paling banyak dari SMA. Kemudian pada clusster 2 berisikan mahasiswa dengan jumlah frekuensi kota asal lebih sedikit dibandingkan dengan clusster 1. Sedangkan pada cluster 3 adalah berisikan jumlah frekuensi asal kota paling sedikit diantara ketiga hasil cluster tersebut.

5.2 Saran

Saran yang dapat peneliti berikan untuk penelitian lebih lanjut :

1. Meningkatkan kecepatan proses dalam sistem K-Means Clustering.
2. Menambahkan sumber data dalam penelitian ini dari hanya mencakup data pada Prodi Teknik Informatika menjadi data Fakultas ataupun mencakup semua data di Universitas itu sendiri.

DAFTAR PUSTAKA

- Akhmad. 2017. *Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Iain Salatiga*, Yogyakarta:UIN Sunan Kalijaga.
- Azmi. 2012. *Data Mining Menggunakan Algoritma Agglomerative Hierarchical Clustering dan Algoritma Apriori pada Data Transaksi Swalayan*.
- Dias Rima Ramadhani. 2013. *Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*. Jurnal Sistem Informasi.
- Fadli. 2011. *Konsep Data Mining*. <http://ilmukomputer.org/category/datamining>.
Diakses : 28 Juli 2016.
- Failasufa, Fusna. 2014. *Penerapan Data Mining untuk analisis pola pembelian konsumen dengan algoritma apriori pada data transaksi penjualan pamella supermarket*. Yogyakarta: UIN Sunan Kalijaga.
- Fathansyah. 1999. *Basis Data*. Bandung : Informatika Bandung.
- Febrian, Nopal. 2017. *Penerapan Data Mining Untuk Menentukan Pengelompokan Uang Kuliah Tunggal (UKT) UIN Sunan Kalijaga Menggunakan Algoritma K-Means Clustering*, Yogyakarta:UIN Sunan Kalijaga.
- Gunadi. 2012. *Penerapan Metode Data Mining Market Basket Analysis terhadap Data Penjualan Buku dengan Menggunakan Algoritma Apriori dan frequent pattern growth (FP-growth)*.

- Huda. 2010. *Aplikasi Data Mining untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*.
- Johan Oscar Ong. 2013. *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*. Jurnal Ilmiah Teknik Industri. vol. 12. no. 1. pp. 10-13.
- Kadir, A. 2005. *Dasar Pemrograman Web Dengan ASP*. Yogyakarta: s.n.
- Kusrini, luthfi taufiq Emha. 2009. *Algoritma Data Mining*. Penerbit Andi:Yogyakarta.
- Lingar, S. 2004. *Hypertext Preprocessor*.
<https://saraswatylingar.files.wordpress.com/2014/03/PHP.pdf>. Diakses: 28 Juli 2016.
- Mcleod, Ray Jr. 1995. *Sistem Informasi Manajemen*. PT.Prenhalindo. Jakarta.
- Murdick, Ross. 1993. *Perancangan dan Pembangunan Sistem Informasi*. Penerbit Andi. Yogyakarta.
- Moertini, Veronica. 2002. *Data Mining Sebagai Solusi Bisnis Integral*. Vol 07 no.1, April.
- Roger S. Pressman, 2002. *Rekayasa Perangkat Lunak Pendekatan Praktisi (Buku Satu)*. ANDI : Yogyakarta.
- Scout, 1996. *Definisi Sistem*.
<http://k4c1r.net.au.net/Analisis%20&%20Perancangan%20Sistem%20Infor>

masi/Definisi%20Sistem_%20Informasi_%20Dan%20Sistem%20Informasi.pdf. Diakses: 28 Juli 2016.

Wijanarko, Andi. 2016. *Penerapan Data Mining Untuk Menentukan Strategi Promosi Universitas PGRI Yogyakarta Menggunakan Algoritma K-Means Clustering*, Yogyakarta:UIN Sunan Kalijaga.

Yakub. 2008. *Sistem Basis Data Tutorial Konseptual*. Penerbit Graha Ilmu: Yogyakarta.



LAMPIRAN

LAMPIRAN A Tabel IP Semester 2 angkatan 2013 sampai 2016

NIM	NAMA MAHASISWA	IPK
13650001	REZA GAMA HIDAYAT	3,62
13650002	NUGROHO ANJAR SETYARSO	3,58
13650003	Sidiq Setyadji	2,18
13650004	HEDI ADI SANTOSA	2,74
13650005	YUDHAWAN ARIF PRATAMA	3,63
13650006	HARRY MUSTIKA HADI	3,22
13650007	JUMADIL ABDUL RAHMAN SELIAN	3,43
13650008	MUHAMMAD ZAKI R	3,26
13650009	NOPAL FEBRIYAN	3,5
13650010	MONALISA ARCELIA	3,49
13650011	FINA NURDIANI	3,41
13650012	LAKSAMANA EKA SAPUTRA	3,34
13650013	MUHAMMAD ARIS ASROR	2,88
13650014	HARI SURRISYAD	3,62
13650015	MUHAMMAD AMIRUDIN MUSTHOFA	3,18
13650016	OKI MAHPUDIN	2,83
13650017	M, RANDI SETYAWAN	3,31
13650019	MUHAMMAD HAEDAR Z,H,	3,2
13650020	NINDA'UL QASANNAH	3,26
13650021	ADAM YAHYA A, HASIBUAN	3,35
13650022	AHMAD SUBHAN YAZID	3,78
13650023	MUHAMMAD AZZAM MUJADDID	3,63
13650024	AGIL MUBAROK	3,02
13650025	MAULANA RUSHARDIANTO	3,14
13650026	MUSLIH ZAINURI	2,39
13650027	IRHAM DZUHRI	3,5
13650028	NUR ROHMAN	3,31
13650029	HALIMAH WARDANANTI	3,23
13650030	FEBRI EKA SETYAWAN	3,53
13650031	FAUZI SAMPURNO PRIBADI	3,64
13650032	MUHAMMAD FARIZ	3,12
13650033	CATUR ADI NUGROHO	2,87
13650034	MAHBUB PUBA FAWZAN	3,44
13650035	ADI WIDYA FATNA	3,47
13650036	HIKMAWAN HUDZAIFAH F	3,36

13650037	M NURKHOIRI H	3,36
13650038	AFFIFAH MUTIARA PERTIWI	3,49
13650039	NURUL HIKHMATUL AZMI	2,99
13650040	AISA SAIKHUSITA	3,14
13650041	INTAN PUTRI PRATIWI	3,46
13650042	TONI EFENDI	3,8
13650043	RAZENDRA BINTANG KHARISMA	3,6
13650044	DEVARA EKO KATON M	2,64
13650045	MIJIL SURYANINGTYAS	3,51
13650046	YOGA INDRA HERISTYA	3,56
13650048	ANISA NURUL WILDA	3,37
13650049	TRI SETYA DARMAWAN	3,48
13650050	TONTOWI PRASETYO	3,06
13650053	MUHAMMAD IQBAL	2,63
13650056	MUTIA AFIFAH	3,61
13650070	BUSTANUL ALBET	3,52
13650071	ROBBY TADEO REINWALD	3,16
13650081	MOHAMMAD UBAIDILLAH	3,44
13650082	NUR BANI ISCHAK	3,26
13650085	YASMITHA LUSY EKAYANTI	3,43
13650088	DEVI SUSANTI	3,26
13650090	TSURAYYA ATS-TSAURI	3,74
13650100	NAEMAH WAEBUERAHENG	3,06
13651047	RAMADHANTI EKA WAHYUDI	3,32
13651051	LUCKY ALFARUQ	2,92
13651052	YUHA BANI MAHARDHIKA	3,33
13651054	FAVIAN KUSUMA RAMZY	2,64
13651055	AJI RACHMAD KURNIAWAN	2,54
13651057	ALIFAH SURYA GAMİYANTI	3,51
13651058	PISCA PRASETYO AWAY WECAN	3,37
13651059	MOHAMMAD MULTI AKBAR	3,31
13651060	DANANG AJI BIMANTORO	3,83
13651061	DILON DAMARA	3,23
13651062	AHMAD ZAHID ALIMUDDIN	3,13
13651063	ALVIYAN RAHMAT DWI PUJO PRAKOSO	3,49
13651064	RISKI INDO RAHARJO	2,51
13651065	RESTU TULUS RAMADHANI	3,38
13651066	YUDI ISTIANTO	3,3
13651067	MUHAMMAD KHOIRUL MUSTOFA ARIS	3,67
13651068	MUHAMMAD HUDALLOH	3,5
13651069	MAULANA SYAEIKHI NAEFIQ	3,12

13651072	DINI NUR ISLAMI	3,56
13651073	RIZKI DEWANTARA	3,43
13651074	BIBITA HABIBI MUSTAFA	3,75
13651075	MOH,TAUFIK	2,88
13651076	HANIFAH RAHMAWATI	3,42
13651077	NADYA PRATAMA PUTRI	3,39
13651078	ANGGORO PRISTYO PENDROJOYO PRANOMO	3,11
13651079	EKO JHONY PRANATA	3,21
13651080	ARIES FIRMANSYAH	3,24
13651083	ARIS MUFLIHUL AINI	3,13
13651084	AMRUL HANIF	2,91
13651086	IIN INTAN ULJANAH	3,75
13651087	USMAN IBRAHIM NASUTION	2,87
13651089	WAHYU RETNO KURNIASARI	3,19
13651091	SENDY ADITYA SURYANA	2,98
13651092	MOHAMAD FAHRONI	3,18
13651093	ASEP WISNU	3,18
13651094	MOH FAHRIZAL YUSUF	2,84
13651095	FAJRI YANDI	2,51
13651096	MIFTAH FADHIL HAQ	3,25
13651097	AJI WAHYU SASMITO MUHAMMAD ZAILANI	3,01
13651098	AYU NINGSIH	3,5
13651099	ALFI ROHMATIN CHASANAH	3,59
14650001	M, IRHAN HAFIEZ	3,26
14650002	WELLY KRISNABAYU	2,31
14650003	TRI SETYO DERMAWAN	3,62
14650004	DANANG DWI APRIANSYAH BALANI	3,2
14650005	AGUS SUPARMAN	3,31
14650006	LUQMAN ARIEF DHARMAWAN	3,19
14650007	IRE ICCA AGUSTIORINI HARSEHANTO	3,4
14650008	REZKI NUR SAFITRI	2,98
14650009	LIES TSANIA FIRHATUSSHIAMY	3,42
14650010	RAHMAWATI ULFA	3,55
14650011	YENNY PUSPITA DEWI	3,15
14650012	MUHAMMAD NURUDDIN ABDULLOH	3,34
14650013	DITA SEPTASARI	3,46
14650014	RISALATUL MUAWANAH	3,43
14650015	RAKA ADI NUGROHO	3,71
14650016	AGUNG KOESDIANTORO	2,71

14650017	SAHRI RAMADHAN	3,59
14650018	RASYID RIDHO	1,74
14650019	BAMA ABPAMA SEVSA	3,71
14650020	NURUL HASNAH	3,27
14650021	MUHAMMAD NAUFAL	2,52
14650022	IQBAL ADI NURMANSYAH	3,56
14650023	ALFIAN WICAKSONO	2,97
14650024	RIDWAN WULIDA SIAM	3,6
14650025	TYAS ABIMANYU	3,41
14650026	LUSIANA LESTARI	3,62
14650027	IKA TRISMIATI	3,57
14650028	HUSNUN KARIMAH OPRALIA	3,44
14650029	RIZIA PRAJA ROSALIN	3,33
14650030	M, NOFELIANSYAH PUTRA	3,2
14650031	AMIROH MUJAHIDAH	3,51
14650032	MUHAMMAD REZA VELAYANI	3,31
14650033	KAREN DHARMAKUSUMA	3,53
14650034	RULYSA PUTRI	3,49
14650035	M ADRIAN RAHMANDANU	3,24
14650036	SULAIMAN MUSTHOFA S	3,37
14650037	DENI PRATAMA PUTRA	3,18
14650038	YOGI MUSTAFA BISRI	2,87
14650039	HILMI MUKTITAMA	3,33
14650040	BRYLIAN ABDURRAHMAN	3,39
14650041	KHOIRUL BARIYYAH	3,21
15650002	LANA RAHIM	3,54
15650003	ANNISA KHODISTA SYAKA	3,69
15650004	MUHAMMAD FAUZI ROBBANI	3,2
15650006	ILHAM ROHMAD DANI	3,56
15650007	MUHAMMAD FAUZAN AL-BAIHAQI	3,41
15650008	IRSALINA SANTI KHASANAH	3,57
15650009	BRIAN RAKHMAT AJI	3,2
15650010	DHONI ARI NUGROHO	3,23
15650011	DINIATI RUAIKA	3,51
15650012	RIZQI ZIDANUL FARHAN	3,4
15650013	YUDHA RIWANTO	3,16
15650014	YUNITA ENGGARSARI	3,09
15650015	SHERLY ANDINI	3,13
15650016	MUFTIA CHALIDA	3,63
15650017	ENCIK NURUL HAFIDAH	2,79
15650018	NAFI'A NISA'UL HAMIDAH	3,41

15650019	YUSSUF FAISAL	3,22
15650020	ERNA NOVIA SAFITRI	3,56
15650021	YULINAR ANGREASTI	3,07
15650022	TESYA NURINTAN	3,63
15650023	ZIDNI LUTVI HAMDAN	3,49
15650024	FATIMAH NUR HIDAYATI	3,13
15650025	AJIE DWIHASTADI	3,47
15650026	HAMID MUSAFA	3,27
15650027	FACHRUL AMRI FIRMANSYAH	3,2
15650028	BEDA CAHYUDA ARUM MUKTI	3,19
15650029	MARDIYAH NURHAYATI	3,64
15650030	AHMAD LUDFI NI'AM	3,37
15650031	HASAN MA`RUF	3,37
15650032	MUHAMMAD SABDA ISA A	3,21
15650033	ANDRI PUJI PRASETIYO	3,42
15650034	ENDAH DWI HARRI Riantami	2,87
15650035	HAFID IQBALGIS	3,24
15650036	KITAMI AKROMUNNISA	3,39
15650037	KHOIRUL UMAM	3,37
15650038	RIKO PUTRO NUGROHO	3,46
15650039	JATANG WANNUAR	3,28
15650040	AHMAD DEDDY SHAHERTIAN	2,96
15650041	REZA KURNIAWAN	3,14
15650042	MISBAHUL ULUM	3,28
15650043	ANNISA SOLEHATUL JANNAH	3,21
15650044	FIKRI WIBOWO	3,41
15650045	AFIF MUHAMMAD	3,54
15650046	WAHYU IMANDA	3,58
15650047	MUHAMMAD HABIBIE	3,42
15650048	ARFIANSAH RIDWAN PRATAMA	3,46
15650049	MUHAMMAD FAHMIL ILMI	2,69
15650050	WILDAN NADIYAL AHSAN	2,86
15650051	DIDIK EKO PRAMONO	3,49
15650052	AKHMAD FAUZI	3,02
15650053	MUSYAFA' IRFAN EFENDI	3,31
15650054	ARDI SETIAWAN	3,46
15650055	JAZILATUL ATIYAH	3,53
15650056	MR, FITREE TAHE	3,16
15650057	NAUFAL ASYHAB	3,3
15650058	ZAHID SUGIH PANUWUN	3,39
15650059	RADEN MAS HAIQAL AGNISFATULLAH	2,66

16650001	RIZKA ANANDA	2,12
16650002	FAJRI RAMADHAN	3,05
16650003	LINA PUSPITASARI	3,34
16650004	M, RIZQI DHIYA'UL HAQ	3,39
16650005	SASQYA NILA DEVI	3,4
16650006	NUR ANITA	3,44
16650007	DINDA AYU DEWANTI	3,43
16650008	JUANG SABIT AL NADIEV	3,42
16650009	RIZKA FEBRI SURYANI	3,41
16650010	MUHAMMAD WAHYU LAMBULUSY	3,51
16650011	IKA PUTRI KUSUMA ASTUTI	3,33
16650012	NUR FAIZAH	3,48
16650013	SITI MARYAM	3,74
16650014	NADIA SHOLIHAH	3,59
16650015	NURCHULIS	3,23
16650016	NUR LAILA JAMIL	3,44
16650017	MAGHFIRA IZZANI R,	3,45
16650018	DENIS DARWA SAPUTRA	3,48
16650019	ISNAN HANIFI	1,63
16650020	KHURIN 'IEN MUKHOYYAROH	3,64
16650021	FADQURROSYIDIK	3,56
16650022	INTAN PRADINA RAMADANTY	3,38
16650023	SEKAR MINATI	3,63
16650024	USSI SANDA LUSIANITA	3,41
16650025	MUHAMMAD BAYU IRFAN PRATAMA	3,61
16650026	KHUSAIRI ABDY	3,42
16650027	ARNANDO FIRHAN PRAYUDHA	3,48
16650028	ULFA MULYA	3,59
16650029	AHMAD NUR FAUZI	3,6
16650030	AZIZ FUADY NEGARAWAN	3,61
16650031	MOH ALI RIDWAN	3,57
16650032	IRFAN MAULANA HAKIM	3,64
16650033	HENDRA DEA ARIFIN	3,54
16650034	SYAFA`AT ADI NUGRAHA	3,45
16650035	MUHAMMAD AKID MUSYAFA	3,53
16650036	AZIS ALVRIYANTO	3,51
16650037	ADITYA PRATAMA NUGRAHA	3,53
16650038	M RIZAL AKBARUDIN ASLAM	3,42
16650039	FAPRILIA SHEHANA	3,24
16650040	AGUS AHMAD KHOIR	3,5
16650041	MUHAMMAD RAFFIUDIN	3,58

16650042	MUTI'AH RIHUL JANNAH	3,68
16650043	YUSUF FUADI	3,34
16650044	MUHAMMAD SHOLIHUDDIN NUR	3,36
16650045	GHAIDA NURUL RAHMA HAKIKI	3,41
16650046	YAYAH SITI NURKOMARIAH	3,48
16650047	YA'KIN ARIF PRABOWO	3,46
16650048	NUR RAHMAN L	3,35
16650049	YAYANG TRI WIJAYA	3,56
16650050	MUHAMMAD FAJAR HADIL UMAM	2,99
16650051	MOCHAMAD GHIFARI 'AZMI	3,34
16650052	AMBAR ARUM JULIYANTI	3,23
16650053	AHMAD SEPTIAWAN	3,35
16650054	WINDA VIONITHA	3,49
16650055	SITI SHOFIYAH	3,53
16650056	IKA FITRIANA WULANDARI	3,53
16650057	ERWIN AJI NUGROHO	3,55
16650058	MILADIA NUR AMALINA	3,56
16650059	NELLY AMALIA	3,58
16650060	ARI LUKMAN PURNAWIRAWANTO	3,43
16650061	AHMAD ARDIYANTO	3,5
16650062	RIZANDA EGGI PRATAMA	3,07
16650063	DEA NUR NAFI'	3,61
16650064	TITIK NASIROH	3,68
16650065	RAMADHAN SHALAHUDIN AL AYYUBI	3,81
16650066	HUSNI MAULANA LUTHFI	3,08
16650067	WILDANUN NABIL	3,08
16650068	KHAMDAN NAHARI	3,39
16650069	INARATUL MAHBUBAH	3,31
16650070	MUHAMMAD RIFKI ADAM ALBARISYI	3,39
16650071	MANAARUL HIDAYAT	3,43
16650072	KHOLIQ AMRULLOH	3,54
16650073	RORO INTAN PANUNTUN	3,29
16650074	AMALIA NURLITA HARTATI	3,44
16650075	KHAERULLAH	3,19
16650076	REZI SEPTIANTO	3,39
16650077	GAYUH RIDHO	3,49
16650078	YULIA SITI AMBARWATI	3,68
16650079	MUHAMMAD FAUZAN ASHARI	3,43
16650080	JOHAN MU'ALIMIN	3,36
16650081	MUHAMMAD FAJAR WIJAYA	3,73
16650082	MUHAMMAD DHIAUDDIN	3,39

16650083	AMIN MA'RUF	3,39
16650084	AFIF HILMI RAIHAN	3,14
16650085	PUTRA NAWAWI	3,46
16650086	FAUZAN ARIF SANI	3,4
16650087	IHSAN PRAYOGA	3,34
16650088	RAHMAT NUR LATIF	3,24
16650089	RIDWAN PRAMUDYA WARDANI	3,32
16650090	ALFIA SHOFIYATUN	3,1
16650091	FAIZAL NOVIAN HIDAYAT	3,74
16650092	ABDUL AZIZ	3,2
16650093	ALFIAN NURUL HUMAIDA	3,36
16650094	ERIZTA ALIFA PRABANDANI	3,64
16650095	SOFYAN	3,31
16650096	FARHAN ARMAWAN ASDIN	3
16650097	M, SYAIFUL ANWAR	2,8

LAMPIRAN B Tabel Dataset Mahasiswa

KOTA MHS	ASAL SEKOLAH	IPK
KLATEN	SMA	3,02
YOGYAKARTA	SMA	3,14
SLEMAN	SMA	2,18
TEMANGGUNG	SMA	3,31
LANGKAT	MA	3,43
BANTUL	MA	3,53
BEKASI	MA	3,35
SLEMAN	MA	3,62
NGANJUK	SMA	3,62
PATI	SMA	2,88
BANJARNEGARA	SMA	3,12
BANTUL	SMA	3,63
BANTUL	SMA	2,99
WONOSOBO	SMA	3,6
SLEMAN	SMA	3,46
BANTUL	SMA	3,14
SURAKARTA	SMA	3,36
BANYUWANGI	SMA	3,48
BANTUL	SMK	2,63
GUNUNGKIDUL	SMK	2,39
BANTUL	MA	3,23

KULON PROGO	MA	3,49
KLATEN	SMK	2,54
SLEMAN	SMK	3,47
GUNUNGKIDUL	SMA	3,51
TULANG		
BAWANG BARAT	SMA	2,92
SUKOHARJO	SMA	3,44
PROBOLINGGO	SMA	3,37
BANTUL	SMA	3,38
GUNUNGKIDUL	SMA	3,3
SLEMAN	SMK	2,64
LAMPUNG		
BARAT	SMK	3,12
NGAWI	SMA	3,37
TEMANGGUNG	SMA	3,5
KEDIRI	MA	3,52
MUSI RAWAS	MA	3,13
GRESIK	SMA	3,67
NGAWI	SMA	3,49
BATANG	SMA	2,51
MAGELANG	SMA	3,56
MAGELANG	SMA	3,36
NGAWI	SMA	3,8
BANTUL	MA	3,32
KLATEN	MA	3,33
BANTUL	SMTI	2,87
KULON PROGO	SMTI	3,83
YOGYAKARTA	MA	3,61
GRESIK	MA	3,23
BANTUL	MA	3,26
CILACAP	MA	3,26
BALIKPAPAN	SMK	3,25
GRESIK	SMK	3,44
YOGYAKARTA	SMK	3,5
SLEMAN	SMK	3,59
BOGOR	MA	3,74
BOYOLALI	MA	3,19
BREBES	SMA	3,11
PALANGKARAYA	SMA	3,43
MANDAILING		
NATAL	SMK	2,87
JAKARTA UTARA	SMK	2,88
BANYUWANGI	SMA	3,75

PASURUAN	SMA	2,84
TASIKMALAYA	SMA	3,18
SLEMAN	SMA	3,06
BLITAR	PON	3,75
KEBUMEN	PON	3,51
KLATEN	SMA	3,42
LAMPUNG		
BARAT	SMA	3,21
GRESIK	MA	3,13
SIAK	PON	3,39
MAGETAN	MA	3,06
TASIKMALAYA	MA	3,56
CILACAP	SMK	3,31
TANJUNG BALAI	SMK	3,16
YOGYAKARTA	SMA	3,24
BANTUL	SMA	3,01
SLEMAN	SMA	3,43
SEMARANG	SMA	2,91
MAGELANG	SMK	2,64
TASIKMALAYA	SMK	2,98
SITUBONDO	SMK	2,51
GROBOGAN	SMK	3,18
BANTUL	SMA	3,22
YOGYAKARTA	SMA	3,34
GUNUNGKIDUL	SMA	3,2
MEDAN	MA	3,31
MAGELANG	SMA	3,18
BREBES	SMA	2,74
BANTUL	SMA	3,26
CIAMIS	SMA	3,26
LAMONGAN	MA	3,78
MAGETAN	MA	3,63
LOMBOK TIMUR	MA	3,5
MUSI		
BANYUASIN	MA	3,49
PURBALINGGA	SMA	3,41
BEKASI	SMA	3,5
CILACAP	SMA	3,64
LUWU UTARA	SMA	3,58
DHARMASRAYA	SMA	2,83
PEKANBARU	SMA	3,26
KENDARI	SMA	3,2
KARANGANYAR	SMA	3,62

DEPOK	SMA	3,31
BANTUL	SMA	3,34
BANJARNEGARA	SMK	3,71
SIMALUNGUN	SMK	2,71
LOMBOK		
TENGAH	SMK	3,55
SEMARANG	SMK	3,43
SEMARANG	MA	1,74
BATAM	MA	3,4
KUNINGAN	SMK	3,42
LAMPUNG		
BARAT	SMK	3,15
PRINGSEWU	SMA	3,46
PONTIANAK	SMA	2,31
BANTUL	SMK	3,59
BONDOWOSO	SMK	3,19
PEMALANG	SMA	2,98
ASAHAN	SMA	3,27
MUSI		
BANYUASIN	SMA	3,71
MAKASSAR	SMA	2,52
SLEMAN	SMK	3,41
CILACAP	SMK	3,53
KLATEN	SMA	3,24
SLEMAN	SMA	2,97
LAMPUNG		
TENGAH	SMA	3,44
KEBUMEN	SMA	3,56
MUARA ENIM	SMA	3,2
KUNINGAN	SMA	3,62
KULON PROGO	SMK	3,57
BANTUL	SMK	3,51
MATARAM	SMK	3,31
HULU SUNGAI		
UTARA	SMK	3,49
BALIKPAPAN	SMA	3,33
KULON PROGO	SMA	3,6
YOGYAKARTA	MA	3,37
SUKOHARJO	MA	3,33
SLEMAN	SMA	3,12
BANTUL	SMA	3,18
SLEMAN	SMA	3,39
BANTUL	SMA	2,89

WONOGIRI	SMK	3,46
NGAWI	SMK	3,3
KEBUMEN	MA	3,41
HULU SUNGAI		
UTARA	MA	3,42
MADIUN	MA	3,46
BREBES	MA	3,02
SALATIGA	MA	2,86
WONOGIRI	MA	3,49
LUAR NEGERI	SMA	3,16
BLORA	SMA	3,31
TEMANGGUNG	MA	3,53
SLEMAN	MA	3,54
CIREBON	SMA	3,58
BREBES	SMA	3,69
BANTUL	SMA	3,39
BANTUL	SMA	2,66
BANYUMAS	SMA	3,56
MAGETAN	SMA	3,09
LUBUKLINGGAU	MA	3,63
BLORA	MA	3,63
BANYUMAS	MA	3,49
METRO	MA	3,69
BIMA	SMA	3,2
BANTUL	SMA	3,41
KARANGASEM	MA	3,54
BANTUL	MA	3,16
BANJARNEGARA	SMA	3,13
BIMA	SMA	2,79
KLATEN	SMA	3,57
BIMA	SMA	3,07
TULUNGAGUNG	SMA	3,41
BENGKULU		
UTARA	SMA	3,23
SLEMAN	SMK	3,4
BREBES	SMK	3,22
SUKOHARJO	SMA	3,22
CIREBON	SMA	3,2
BREBES	MA	3,51
BANTUL	MA	3,47
GUNUNGKIDUL	SMA	3,46
BANJARNEGARA	SMA	3,27
MUARA ENIM	SMA	3,19

BLITAR	SMA	3,37
PONOROGO	MA	3,21
PURBALINGGA	MA	3,42
PEKALONGAN	SMA	3,37
JEPARA	SMA	3,39
KEBUMEN	SMA	3,37
KENDAL	SMA	3,37
SLEMAN	SMA	3,28
WONOGIRI	SMA	3,2
SAROLANGUN	SMA	3,64
YOGYAKARTA	SMA	3,12
LABUHANBATU UTARA	MA	2,78
TUBAN	MA	2,87
PEKANBARU	MA	3,21
BANTUL	MA	3,28
BANTUL	SMA	3,15
BANTUL	SMA	1,63
TEGAL	MA	3,39
DENPASAR	MA	3,4
MAGETAN	MA	3,48
TEGAL	MA	2,12
BANTUL	SMK	3,56
SLEMAN	SMK	3,38
RABANGODU	SMA	3,05
WONOGIRI	SMA	3,59
TRIMULYA	SMA	3,44
PEGAGAN	SMA	3,45
PONOROGO	MA	3,34
BIMA	SMA	3,33
SRAGEN	SMK	3,47
CILACAP	MA	3,74
TEGAL	SMK	3,42
BANTUL	SMK	3,23
AMBON	MA	3,51
CILACAP	MA	3,48
KULON PROGO	MA	3,44
GUNUNGKIDUL	SMK	3,41
KUDUS	SMA	3,64
JEMBER	MA	3,63
PONOROGO	SMA	3,41
YOGYAKARTA	SMA	3,61

TEGAL	SMA	3,51
KEBUMEN	SMA	3,42
YOGYAKARTA	SMA	3,64
MAGELANG	SMA	3,35
MAGELANG	SMA	3,16
BANJARNEGARA	SMA	3,53
BANTUL	SMK	3,53
WONOGIRI	SMA	3,36
TRENGGALEK	SMA	2,99
BANJARNEGARA	SMA	3,56
TEMANGGUNG	SMK	3,41
PURWOREJO	SMK	3,49
PONOROGO	SMA	3,58
YOGYAKARTA	SMA	3,48
TAMBAN	SMA	3,42
JOHOR BAHRU	SMA	3,59
MAJALENGKA	MA	3,48
YOGYAKARTA	SMA	3,46
BANTUL	SMK	3,35
PACITAN	SMA	3,56
KEBUMEN	MA	3,34
BANTUL	MA	3,23
SLEMAN	SMK	3,53
PACITAN	MA	3,53
INDRAMAYU	MA	3,58
KEBUMEN	MA	3,55
PATI	MA	3,57
SLEMAN	SMA	3,6
BANTUL	SMA	3,61
PURWOREJO	SMK	3,54
TEMANGGUNG	SMA	3,45
PURWOREJO	SMA	3,24
SLEMAN	SMK	3,5
BANTUL	SMK	3,68
KULON PROGO	SMA	3,36
NGAWI	SMA	3,68
PATI	MA	3,39
TEGAL	SMA	3,08
MALANG	SMK	3,07
NGAWI	MA	3,34
AMBON	SMA	3,68
SLEMAN	SMA	3,24

GROBOGAN	SMK	3,32
BAWANG SAKTI JAYA	SMK	3,54
KULON PROGO	SMA	3,39
BANJARNEGARA	SMA	3,49
INDRAMAYU	MA	3,39
WONOSOBO	MA	3,1
SLEMAN	MA	3,43
BANTUL	MA	3,36
CIREBON	MA	3,19
BREBES	MA	3,44
KERTAK HANYAR	MA	3,39
SLEMAN	MA	3,14
BANTUL	SMA	3,74
PANDEGLANG	SMA	3,2
MAGELANG	SMK	3,5
MAGELANG	SMK	3,43
PATI	MA	3,61
BANYUMAS	MA	3,43
JOMBANG	MA	3,08
MAGELANG	MA	3,39
CILACAP	SMK	3,29
GUNUNGKIDUL	SMK	3,81
MAGELANG	SMA	3,73
TEMANGGUNG	SMA	3,31
KARAWANG	MA	3,4
TANJUNG PINANG	MA	3,46
SLEMAN	SMK	3,31
MAGELANG	SMK	3,64
DEPOK	MA	2,8
YOGYAKARTA	MA	3

LAMPIRAN C Tabel Perhitungan Jarak Data ke Setiap Cluster (M.Excel)

CLUSTER 1			CLUSTER 2			CLUSTER 3		
2	1	1,63	28	2	2,12	112	5	3,83

9,10670632	17,0531522	101,0824223
1,811104635	25,04077475	109,0755523
0,55	26,01929284	110,0850694
11,12755139	15,08032161	99,08213966
70,03027917	44,01949682	40,11433659
2,368543856	27,0367916	111,0409384
29,06816816	3,242360251	81,05695775
2,227128196	26,04323329	110,0411019
85,02329151	59,02753595	25,31884871
17,04589393	9,087221798	93,0908293
3,349641772	23,04343724	107,0770942
2,236067977	27,06067442	111,0722288
1,688075828	27,03251561	111,075225
42,04617581	16,09939129	68,11793376
1,83	26,05370607	110,0733251
1,811104635	27,03775878	111,0741919
102,01467	76,0166929	8,956612083
28,06104952	2,617174048	82,09824907
2,449489743	27,02332511	111,0245018
6,370054945	20,02680454	104,0291959
2,13541565	27,02280703	111,0421542
8,274031665	18,05206082	102,0446745
9,264345633	17,03456486	101,0280362
2,71764604	26,05422231	110,0187693
6,287638666	20,07316866	104,0773866
109,0076332	83,00987893	4,222333478
24,06815531	2,596613179	86,09385634
94,01610288	68,0188393	16,49883632
2,015564437	27,04787607	111,0729603
6,228073538	20,05972084	104,0782441
2,240557966	26,0244193	110,0246159
21,14757906	5,196152423	89,02530034
10,15025123	16,07987873	100,0810252
11,15781789	15,09650291	99,08132468
62,03686082	36,02721194	48,09465771
84,01934301	58,0087933	26,18186395
15,13808442	11,15358687	95,08430785
10,17150923	16,08965195	100,0805456
45,00860362	19,03029427	65,13633702
2,779370432	24,06394814	108,074386

2,644409197	24,05280857	108,0750707
10,23273668	16,11900741	100,0799725
2,203656053	27,02665351	111,0417043
9,213576938	17,04300736	101,0457817
4,305531326	27,1765064	111,0041513
9,210863152	18,32823232	102
2,43318721	25,04436264	109,0414985
15,11820095	11,0558627	95,04925039
2,157985171	27,02405595	111,0419961
5,353213988	21,03092009	105,0443949
27,12239665	1,810220981	83,02611878
15,24060694	11,12395613	95,02185064
2,914944253	25,05802067	109,0188465
2,8002857	26,06071565	110,0184421
48,05675915	22,05956482	62,07260346
50,03432422	24,02384024	60,07836216
4,265020516	22,04495634	106,0778884
86,01883515	60,0226299	24,33433788
80,03460252	54,01446566	30,08191483
54,05148009	28,02815727	56,04375523
28,08014245	2,767110406	82,09754199
88,00831836	62,01224395	22,38258475
25,04800391	1,767370929	85,09654811
1,43	26,03619788	110,0753964
30,22406988	4,75992647	80,00628975
7,844386528	19,1554718	103,0053513
9,176279202	17,07893439	101,0800084
21,05935422	5,214220939	89,09199964
15,10794493	11,04627086	95,04993425
98,06170302	72,03896793	12,0496307
16,09487185	10,04408284	94,05101222
25,09432007	1,75316856	85,05335326
5,641134638	21,05744761	105,0203333
104,0304806	78,01334245	6,359944968
1,89528362	25,04504741	109,0749655
1,704230031	27,03316667	111,0750755
1,8	26,05218033	110,07343
23,03558986	3,259463146	87,09676458
3,003348132	24,02645209	108,0250716
25,11618004	1,655173707	85,02777487
100,0238692	74,00778405	10,28311237

33,09686541	7,150076923	77,02871218
1,878323721	27,04089496	111,0737237
1,980934123	25,04971856	109,0744704
6,202007739	20,05408687	104,0788014
82,02330401	56,01264232	28,16505636
2,530316186	24,04420096	108,0760033
4,151156465	22,03144117	106,081045
1,912302277	27,04255165	111,0735112
51,02604139	25,04594977	59,13818479
68,04132935	42,03279196	42,10703623
16,15549442	10,11336245	94,04807281
72,0312217	46,02069534	38,11966553
38,05863371	12,07795099	72,06327497
41,03862083	15,08854201	69,11712089
29,06022884	3,450275351	81,09937669
5,388886712	21,07867168	105,0763346
75,02534572	49,03194469	35,22871698
53,01358317	27,02783935	57,14892825
40,03319747	14,08188908	70,11650947
64,01925413	38,02849458	46,17788323
59,03355063	33,04920574	51,15705328
32,04406965	6,198072281	78,1042278
1,980934123	27,04604222	111,0731295
4,162499249	23,07657037	107,0187572
99,02608949	73,00923298	11,23629832
71,05410896	45,03381951	39,05225218
23,15685644	3,422878905	87,02390476
23,02199166	3,023970899	87,07679427
44,04693973	18,04545372	66,06954593
36,09991828	10,13360745	74,02815748
21,14971395	5,202009227	89,02506613
93,0180031	67,02085944	17,46816819
92,00251301	66,00784878	18,50163236
2,973482806	27,05847187	111,018276
49,06560506	23,04658109	61,03613356
91,01001319	65,01338016	19,43508426
43,03126305	17,06817213	67,12163288
38,05688374	12,14611461	72,11112536
78,0050774	52,01115265	32,27562703
2,677386786	26,05118231	110,018982
5,710516614	21,07102513	105,0194744

9,14287154	17,06617708	101,0808988
1,34	26,03310393	110,0760628
69,02373577	43,03187656	41,19650592
7,261191362	19,08071277	103,0779943
37,03329448	11,09803586	73,11222128
36,05495944	10,16120072	74,10832679
8,471339918	18,08597523	102,0199373
2,921369542	27,05424366	111,0184777
81,04210264	55,02196016	29,07353436
34,1095236	8,177829785	76,02707149
27,05346558	1,861209284	83,09783391
8,238986588	18,08840513	102,0786604
2,242231032	25,03123049	109,0422469
24,0809053	2,337541443	86,05376227
1,49	26,03843313	110,0749931
1,844586675	27,03929733	111,0739506
1,76	26,05019962	110,0735827
1,608601877	27,02948205	111,0760262
13,27964231	13,10708206	97,02132188
10,33387149	16,07458864	100,0214022
7,291666476	19,04374175	103,044536
34,06176889	8,104936767	76,06029253
76,0286058	50,01795278	34,13410172
4,351103308	22,0184014	106,0455379
96,01308713	70,00391132	14,3506411
13,17040622	13,07198914	97,04697625
73,0160318	47,02213947	37,22161872
31,04548921	5,236038579	79,10291019
11,20758672	15,06612425	99,04589845
2,155945268	26,03874805	110,0412836
20,09483765	6,255525557	90,08919192
4,499288833	22,07860729	106,0755372
2,024252949	27,04834376	111,0729202
1,435583505	27,02390793	111,0782107
19,09777212	7,216203988	91,08826983
16,06647441	10,09657863	94,08797798
74,03377608	48,02374517	36,12533737
31,08054054	5,223035516	79,05719449
19,11699767	7,132804498	91,05007194

83,03158194	57,02161783	27,16651616
14,0877571	12,08992969	96,08536257
2,041665986	27,04929019	111,0728428
60,03872167	34,02964002	50,09075863
2,083482661	27,02002221	111,0425545
3,354101966	23,04387337	107,0770283
14,04797494	12,06021973	96,08892548
9,206714941	17,09100641	101,0795113
14,0738623	12,07901072	96,08630287
110,0144009	84,01585624	4,021989557
47,02722616	21,05307816	63,12970775
2,670748959	26,05068905	110,0190206
4,746377566	22,05017007	106,0206211
24,05261108	2,491987159	86,09513401
20,06152786	6,177896082	90,09104783
4,531489821	22,04386763	106,0429272
2,320689553	27,03372893	111,0411167
6,272870156	20,06976831	104,0775523
3,419005703	23,05043384	107,0762046
37,03287188	11,09706718	73,11230813
30,05041763	4,308421985	80,10125842
18,09686161	8,073914788	92,05098804
41,0512375	15,05622795	69,06640355
90,01681843	64,0200164	20,40126467
56,02765032	30,04351677	54,14973315
7,213016013	19,06731497	103,078667
63,02402399	37,03461219	47,17214856
1,65	26,04506863	110,0740773
13,09446066	13,08305775	97,08448331
97,02082302	71,02330885	13,60279751
1,794463708	25,03996805	109,0756806
67,01732985	41,00531185	43,11731091
108,0117475	82,00342981	3,731166038
40,04368115	14,04236803	70,06699936
2,173131381	27,02490703	111,0418952
1,819450466	27,03813788	111,0741302
1	27,02295506	111,0938342
12,16953573	14,05748555	98,0468949
2,851823978	28,02924187	112,0409965
16,13761135	10,09205628	94,04851142

12,05156006	14	98,06081837
2,95379417	27,05685865	111,0183449
2,657536453	26,04971401	110,0191006
95,01061204	69,01351245	15,54375759
13,14692359	13,12100987	97,08273585
107,0153078	81,01692663	5,015186936
89,01860704	63,02197157	21,38093543
18,10867472	8,092490346	92,05020424
14,10283659	12,10223533	96,08459814
101,0365558	75,01881431	9,226570327
5,518342142	21,06239303	105,042887
12,29650763	14,09574404	98,02126351
2,749545417	27,04130359	111,0196379
26,08705426	1,39	84,05416349
5,4242511	21,04399202	105,0434315
8,262935314	18,0483351	102,0448534
6,570266357	20,06649197	104,0200769
66,03059972	40,04135862	44,18185261
55,04543578	29,03928546	55,08212051
18,08779699	8,164808632	92,08787325
2,218197466	25,06431926	109,0735917
12,14637394	14,10432912	98,08212069
7,22524048	19,07065809	103,078456
2,245016704	25,06612056	109,0735353
2,637877935	24,05229511	108,0751146
2,518114374	24,04332756	108,0761255
3,551056181	23,06486722	107,0751605
2,93428015	27,0552786	111,0184219
13,11460636	13,09723635	97,08357688
106,0087242	80,01097987	5,718881009
3,56719778	23,06672062	107,0750807
11,32114835	15,08854201	99,02109068
22,16888811	4,344755459	88,02338098
18,1053169	8,193387578	92,08725482
2,102974084	25,05692719	109,0739313
103,0155527	77,01746555	8,072676136
57,03368829	31,0509404	53,15127092
77,02871218	51,01813011	33,13793144
2,08540164	25,05584962	109,0739974
2,821063629	27,04649515	111,0190542
39,04772593	13,11768272	71,11309936

7,274895188	19,03912813	103,0448451
2,13541565	27,02280703	111,0421542
2,758622845	26,0574001	110,0185893
39,05905785	13,07624181	71,06398525
35,06854003	9,117653207	75,06039235
7,327100381	19,05373717	103,0440605
17,13953325	9,116057262	93,04873777
1,97	26,06128163	110,0729435
2,218197466	27,05956578	111,0722666
22,17313916	4,360779747	88,02320205
11,14954708	15,09201444	99,08150382
22,0588327	4,272516823	88,09283796
2,738046749	26,05579398	110,0186752
3,033562262	27,06351049	111,0181179
8,184919059	18,07035141	102,0794832
10,20796258	16,1069426	100,0800804
17,12009346	9,089163878	93,04941483
12,08728671	14,06846118	98,08446615
79,03843116	53,01794508	31,07374454
10,19431704	16,04644509	100,0461898
26,08069209	1,852997572	84,09531794
1,61	26,04331776	110,0742845
33,10371731	7,172168431	77,02765802
46,08305654	20,07526837	64,03189908
8,191312471	18,07243481	102,0793495
3,529815859	23,0624565	107,0752801
35,05848827	9,089163878	75,06126564
42,03761292	16,0299844	68,07005876
2,059126028	26,032981	110,0416285
2,234479805	27,02845908	111,0415278
20,08565657	6,094661598	90,05226038
4,502899066	22,03956442	106,0431615
65,03151236	39,02067273	45,10203543
1,811104635	26,02	110,0430648
2,334973233	27,0670353	111,0720852
87,01416494	61,01775479	23,35373418
3,390707891	24,06043225	108,019021
3,352610923	24,05651887	108,0192575
17,14410686	9,122505138	93,0486346
19,11125323	7,121523713	91,05031576

58,02673953	32,01439676	52,09186597
2,845628226	24,03357859	108,0425546
5,635210733	21,05632684	105,0204342
6,689723462	20,09617128	104,0192309
2,9	24,07471911	108,074095
11,12755139	15,08032161	99,08213966
61,03386683	35,02339789	49,09363401
105,020707	79,01136374	5,84267918
2,611972435	26,04642202	110,0194092
3,469884724	24,06886786	108,018684
32,03699268	6,038410387	78,06446631
1,96898451	25,01548321	109,0444354

LAMPIRAN D Tabel Clustering Data (M.Excel)

CLUSTER 1	CLUSTER 2	CLUSTER 3
ok		
ok		
ok		
ok		
		ok
ok		
	ok	
ok		
		ok
	ok	
ok		
ok		
ok		
	ok	
ok		
ok		
		ok
	ok	
ok		
ok		
ok		
ok		

ok		
ok		
ok		
		ok
	ok	
		ok
ok		
ok		
ok		
	ok	
ok		
ok		
	ok	
		ok
	ok	
ok		
	ok	
ok		
ok		
ok		
ok		
ok		
ok		
ok		
	ok	
ok		
ok		
	ok	
	ok	
ok		
ok		
	ok	
	ok	
ok		
		ok
		ok
	ok	
	ok	
		ok



	ok	
ok		
	ok	
ok		
ok		
	ok	
	ok	
		ok
	ok	
	ok	
ok		
		ok
ok		
ok		
ok		
	ok	
ok		
	ok	
		ok
	ok	
ok		
ok		
ok		
		ok
	ok	
	ok	
	ok	
		ok
	ok	
	ok	
	ok	
ok		
		ok
	ok	
	ok	
	ok	
	ok	
	ok	



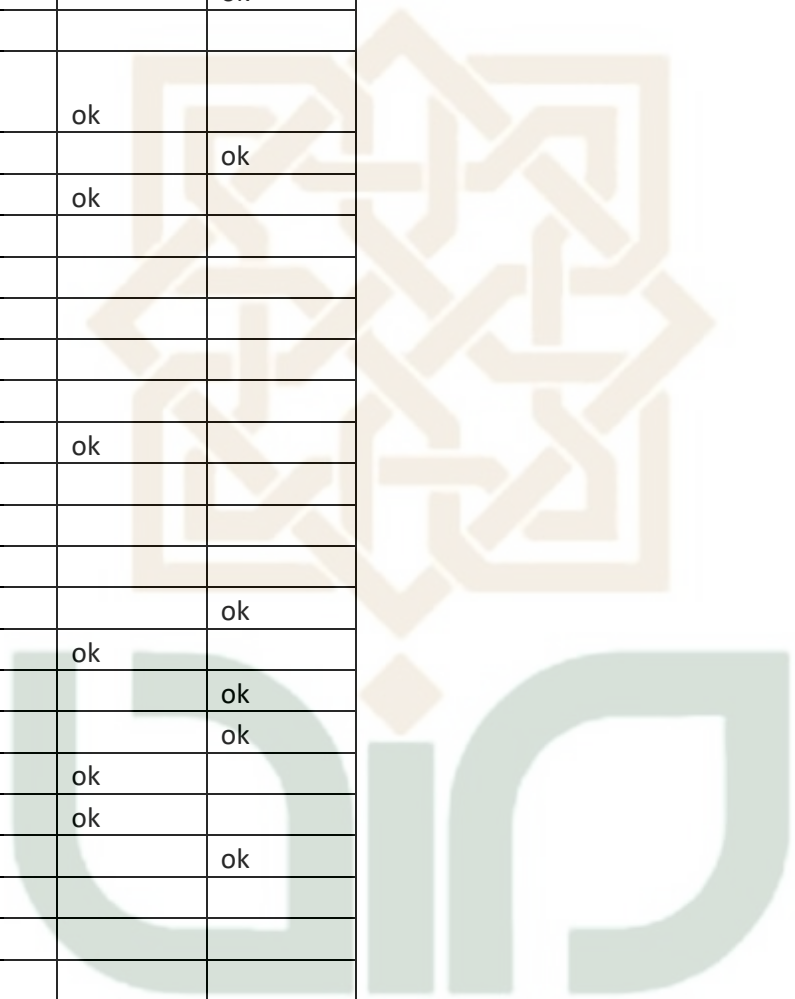
ok		
ok		
		ok
		ok
	ok	
	ok	
	ok	
	ok	
	ok	
		ok
		ok
ok		
	ok	
		ok
	ok	
	ok	
		ok
ok		
ok		
ok		
ok		
		ok
ok		
	ok	
	ok	
ok		
ok		
	ok	
ok		
ok		
ok		
ok		
	ok	
ok		
ok		



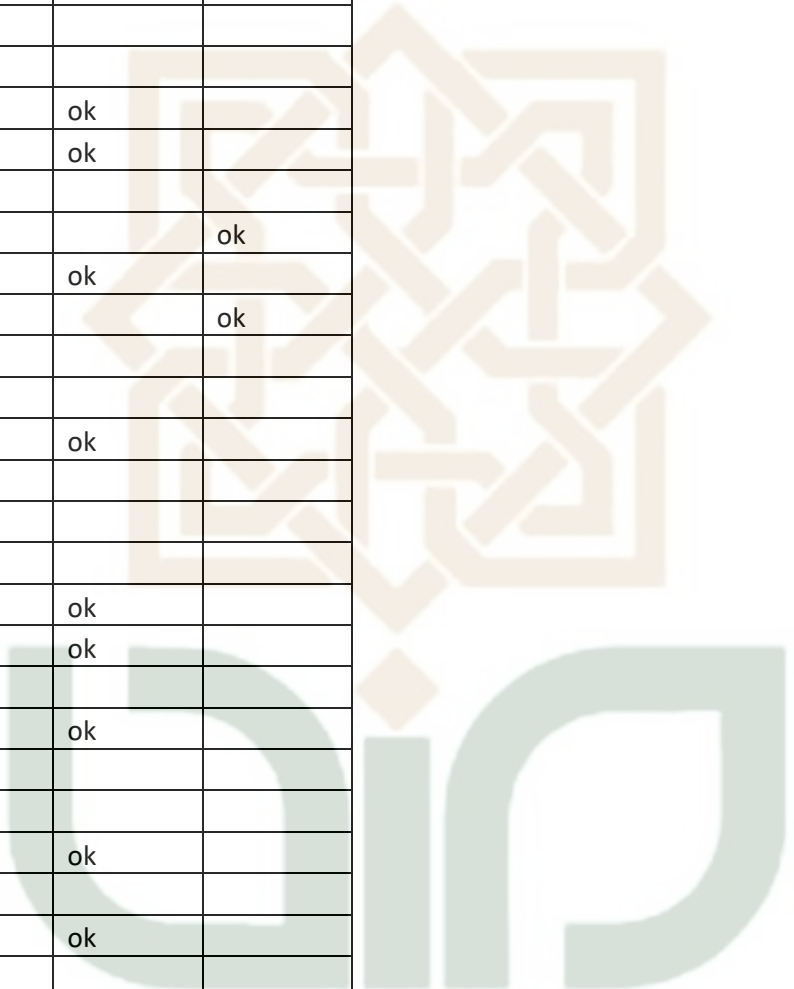
	ok	
		ok
ok		
		ok
	ok	
		ok
	ok	
ok		
ok		
	ok	
ok		
ok		
ok		
	ok	
	ok	
		ok
	ok	
	ok	
		ok
	ok	
ok		
	ok	
ok		
ok		
	ok	
	ok	
ok		
ok		
ok		
ok		
	ok	
	ok	
	ok	



	ok	
		ok
	ok	
ok		
	ok	
ok		
	ok	
		ok
ok		
	ok	
		ok
	ok	
ok		
ok		
ok		
ok		
	ok	
ok		
ok		
ok		
		ok
	ok	
		ok
		ok
	ok	
	ok	
		ok
ok		
ok		
ok		
	ok	
ok		
ok		
ok		
	ok	
	ok	
	ok	
ok		
ok		



ok		
ok		
ok		
ok		
ok		
ok		
	ok	
		ok
ok		
ok		
	ok	
	ok	
ok		
		ok
	ok	
		ok
ok		
ok		
	ok	
ok		
ok		
ok		
	ok	
	ok	
ok		
	ok	
ok		
ok		
ok		
	ok	
ok		
		ok
ok		
	ok	
ok		



	ok	
	ok	
ok		
ok		
	ok	
	ok	
ok		
ok		
	ok	
ok		
	ok	
ok		
ok		
		ok
ok		
ok		
	ok	
	ok	
	ok	
ok		
ok		
ok		
ok		
	ok	
		ok
ok		
ok		
	ok	
ok		

LAMPIRAN E Source Code PHP (index.PHP)

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>HOME</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />

```

```

<link href="default.css" rel="stylesheet" type="text/css" />
</head>
<body>
<div id="header">
<h1>HOME<br />
<span class="text1">welcome to my K-Means Clustering APP</span></h1>
</div>
<div id="menu">
<ul>
<li><a href="index.PHP">Home</a></li>
<li><a href="kmeans.PHP">APP K-Means</a></li>
<li><a href="help.PHP">bantuan</a></li>
</ul>
</div>
<div id="content">
<div id="left">
<h2>Welcome to K-Means Clustering App</h2>
<p> Aplikasi ini dibuat untuk menyelesaikan tugas akhir program
studi
Teknik Informatika di Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
Program ini dibuat setidaknya selama kurang lebih 1 bulan pengerjaan dan
aplikasi ini masih banyak memiliki kekurangan, saya berharap ada
mahasiswa yang melanjutkan dan mengembangkan program aplikasi sederhana
ini.</p>
<p>Aplikasi ini adalah aplikasi untuk mencari klasifikasi dari sebuah data yang
banyak dan masih random atau belum terklasifikasi.
Data yang digunakan dalam aplikasi ini adalah data mahasiswa dengan jurusan
dan univ yang sama dan dibatasi hanya menggunakan 3 cluster.
Aplikasi web ini terdiri dari 3 menu utama yaitu Home, Aplikasi K-Means,
dan Bantuan.</p>
<hr id="border-top" />
<div id="colA">
<p><strong>APP K-Means Clustering</strong><br />
Menu ini berisikan aplikasi dari metode K-Means Clustering. Aplikasi ini
diawali dengan mengimport kan file dengan format <a>EXCEL
2003</a> lalu setelah itu hanya dengan mengklik tombol lanjut untuk
proses selanjutnya. langsung dicoba saja dengan klik dibawah.</p>
<a href="kmeans.PHP">APP K-Means Clustering</a></p>
</div>
<div id="colB">
<p><strong>Bantuan</strong><br />
Menu Bantuan adalah menu yang disiapkan untuk memberi informasi
bagaimana cara kerja atau sistem aplikasi K-Means tersebut berjalan dan
menjelaskan dengan jelas bagaimana cara kerja aplikasi ini dan bagaimana
cara menggunakan aplikasi ini.</p>

```

```

<a href="help.PHP">Bantuan&#8230;</a></p>
</div>
</div>
<div id="right">
<h2>biodata pembuat</h2>

<p> Perkenalkan nama saya adalah <strong>Bayu Resi Indrawan</strong>. saat
ini saya ialah seorang mahasiswa di Universitas Islam Negeri Sunan Kalijaga
Yogyakarta angkatan 2012 Jurusan Teknik Informatika. <a
href="#">More&#8230;</a></p>
<p> Saya tinggal di Yogyakarta, tepatnya berada di Jl.Cendrawasih Nogotirto
IV Gamping Sleman. Hobiku adalah olahraga saya sangat suka basket dan
sepak bola, club favorit saya adalah Juventus yang berasal dari Italia. <a
href="#">More&#8230;</a></p>

<p><strong>[HARAPAN]</strong> Saya sangat berharap semoga aplikasi
sederhana ini dapat bermanfaat dikemudian hari sehingga dapat membantu
pihak kampus untuk menentukan strategi promosi pada saat penerimaan
mahasiswa baru atau ajaran baru. <a href="#">More&#8230;</a></p>
<h2>Contact Me</h2>
<ul>
<li><a>085729889007 </a></li>
<li><a>bayuresi052@gmail.com</a></li>
</ul>
</div>
</div>
<div id="footer">
<p>Copyright &copy; 2016 SKRIPSI Teknik Informatika 2012 </p>
</div>
</body>
</html>

```

LAMPIRAN F Source Code PHP (import.PHP)

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>K-MEANS CLUSTERING</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link href="default.css" rel="stylesheet" type="text/css" />
</head>
<body>

```



```

<div id="header">
<h1>K-Means Clustering APP<br />
<span class="text1">metode Data Mining</span></h1>
</div>
<div id="menu">
<ul>
<li><a href="index.PHP">Home</a></li>
<li><a href="kmeans.PHP">APP K-Means</a></li>
<li><a href="help.PHP">bantuan</a></li>
</ul>
</div>
<div id="content">
<div id="left">
<h2>SILAKAN IMPORT FILE EXCEL</h2>

<?PHP
//koneksi ke database, username,password dan namadatabase menyesuaikan
MySQL_connect('localhost', 'root', "");
MySQL_select_db('skripsi');

//memanggil file excel_reader
require "excel_reader.PHP";

//jika tombol import ditekan
if(isset($_POST['submit'])){

    $truncate = "TRUNCATE TABELdataset";
MySQL_query($truncate);

    $target = basename($_FILES['filepegawaiall']['name']) ;
    move_uploaded_file($_FILES['filepegawaiall']['tmp_name'], $target);

    $data = new
Spreadsheet_Excel_Reader($_FILES['filepegawaiall']['name'],false);

// menghitung jumlah baris file xls
$baris = $data->rowcount($sheet_index=0);

// import data excel mulai baris ke-2 (karena tabel xls ada header pada baris 1)
for ($i=2; $i<=$baris; $i++)
{
// membaca data (kolom ke-1 sd terakhir)
$kota_asal = $data->val($i, 1);
$jurusan_sekolah = $data->val($i, 2);
$ipk = $data->val($i, 3);

```

```

// setelah data dibaca, masukkan ke tabel pegawai sql
$query = "INSERT into dataset
(kota_asal,jurusan_sekolah,ipk)values('$kota_asal','$jurusan_sekolah','$ipk')";
$hasil = MySQL_query($query);
}

if($hasil){
// jika import sukses

echo "<script language='JAVASCRIPT'>";
echo "alert('Data Berhasil Di Importkan')";
echo "</SCRIPT>";

}else{
// jika impor gagal
die(MySQL_error());
}

// hapus file xls yang udah dibaca
unlink($_FILES['filepegawaiaall']['name']);
}

?>

<form name="myForm" id="myForm" onSubmit="return validateForm()"
action="kmeans.PHP" method="post" enctype="multipart/form-data">
<input type="file" id="filepegawaiaall" name="filepegawaiaall" />
<input type="submit" name="submit" value="Import" /><br /><br />
</form><br />
<hr id="border-top" />
<form name="Frek" id="Frek" action="Frek.PHP" method="post" >
<input type="submit" name="submit" value="PROSES SELANJUTNYA"
/><br /><br />
<?PHP $sql = "SELECT * FROM dataset";
$result = MySQL_query($sql);
while($row = MySQL_fetch_array($result)){
$cek = $row['id_mhs'];
if ($cek == 1) {
echo "Note : File sebelumnya masih ada";
}else{ }
}
}
?>
</form>

```

```

</div>
<div id="right">
<h2>FORM IMPORT FILE</h2>
<div id="red"><p> Kategori File yang bisa di import : </p></div>
<ul>
<li>File excel yang diimport hanya bisa dengan format file <strong>Excel 97-2003</strong></li>
<li> Untuk penulisan angka desimal setelah file diimport adalah titik sedangkan ribuan adalah koma </li>
<li> Kolom excel yg di baca adalah dari kolom pertama yaitu kota mhs, jurusan sekolah mhs, ipk <strong>(urut)</strong></li>
<li> Posisikan data tabel di file excel dari pojok kiri atas atau kolom A1 dan baris A1 ( termasuk dengan nama kolom )</li>
<li> Pastikan tidak ada kolom yang <strong>kosong / blank</strong></li>
</ul>

<h2>Contact Me</h2>
<ul>
<li><a>085729889007 </a></li>
<li><a>bayuresi052@gmail.com</a></li>
</ul>
</div>
</div>
<div id="footer">
<p>Copyright &copy; 2016 SKRIPSI Teknik Informatika 2012 </p>
</div>
</body>
</html>
<script type="text/javascript">
// validasi form (hanya file .xls yang diijinkan)
function validateForm()
{
function hasExtension(inputID, exts) {
var fileName = document.getElementById(inputID).value;
return (new RegExp('(' + exts.join('|').replace(/\.\/g, '\\.') + '$')).test(fileName);
}

if(!hasExtension('filepegawai', ['.xls'])){
alert("Hanya file XLS (Excel 2003) yang diijinkan.");
return false;
}
}
</script>

```

LAMPIRAN G Source Code PHP (frek.PHP)

```
<?PHP
MySQL_connect('localhost', 'root', '');
MySQL_select_db('skripsi');

include ('function.PHP');

if(isset($_POST['submit'])){ ?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>K-MEANS CLUSTERING</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link href="default.css" rel="stylesheet" type="text/css" />
</head>
<body>
<div id="header">
<h1>K-Means Clustering APP<br />
<span class="text1">metode Data Mining</span></h1>
</div>
<div id="menu">
<ul>
<li><a href="#">Home</a></li>
<li><a href="#">APP K-Means</a></li>
<li><a href="#">bantuan</a></li>
</ul>
</div>
<div id="content">
<div id="left">

<h2>Memodifikasi Atribut selain Angka</h2>
<form name="dataset" id="dataset" action="dataset.PHP" method="post" >
<input type="submit" name="submit" value="PROSES SELANJUTNYA" />
</form>
<br/>

<?PHP $frek = frekuensi('frek_jurusan','jurusan_sekolah','inijur'); $frek2 =
frekuensi('frek_kota','kota_asal','inikota'); ?>

<div ><tabel border="1" width="200px" id="kiri"><?PHP
inisial('frek_kota','kota_asal','KOTA ASAL','inikota'); ?></tabel></div>
```

```
<div ><tabel border="1" width="200px" id="kanan"><?PHP
inisial('frek_jurusan','jurusan_sekolah','JURUSAN SEKOLAH','inijur');
?></tabel></div>
```

```
</div>
```

```
<div id="right">
```

```
<h2 align="center">TRANSFORMATION DATA</h2>
```

```
<p> Proses ini adalah salah satu tahap Pre-procesing atau tahap sebelum
diproses yaitu menyiapkan data agar bisa diolah. </p>
```

```
<p> Atribut yang digunakan
```

```
    adalah asal kota dan asal jurusan sewaktu SMA yang mana isi data dari
atribut tersebut adalah huruf atau selain angka,
```

```
    karna didalam proses K-Means ini data yang bisa diolah adalah hanya data
berupa angka maka atribut tersebut dimodifikasi kedalam angka. Untuk
mengubah kedua
```

```
    atribut ini sesuai dengan frekuensi terbesar data sampai frekuensi terkecil
data dan memberi inisial yang berurutan sesuai frekuensi yang terbentuk</p>
```

```
<p> Seperti contoh dalam data yang diimport ada beberapa kota yang sama.
maka data kota yang sama dijadikan satu dan dihitung jumlah
mahasiswanya</p>
```

```
<p> Dalam contoh mahasiswa yang berasal dari kota Bantul ada 147 mahasiswa
dan jumlah itu adalah jumlah tersebar daripada jumlah mahasiswa dari kota
yang lain
```

```
    maka kota Bantul berinisial kan 1</p>
```

```
<h2>Contact Me</h2>
```

```
<ul>
```

```
<li><a>085729889007 </a></li>
```

```
<li><a>bayuresi052@gmail.com</a></li>
```

```
</ul>
```

```
</div>
```

```
</div>
```

```
<div id="footer">
```

```
<p>Copyright &copy; 2016 SKRIPSI Teknik Informatika 2012 </p>
```

```
</div>
```

```
</body>
```

```
</html>
```

```
<?PHP
```

```
} else {
```

```
    echo "<META http-equiv='Refresh' content='0; URL=kmeans.PHP'>";
```

```
}
```

```
?>
```

LAMPIRAN H Source Code PHP (dataset.PHP)

```
<?PHP
MySQL_connect('localhost', 'root', '');
MySQL_select_db('skripsi');

include ('function.PHP');

if(isset($_POST['submit'])){ ?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>K-MEANS CLUSTERING</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link href="default.css" rel="stylesheet" type="text/css" />
</head>
<body>
<div id="header">
<h1>K-Means Clustering APP<br />
<span class="text1">metode Data Mining</span></h1>
</div>
<div id="menu">
<ul>
<li><a href="#">Home</a></li>
<li><a href="#">APP K-Means</a></li>
<li><a href="#">bantuan</a></li>
</ul>
</div>
<div id="content">
<div id="left">

<h2>Dataset yang diolah</h2>
<h4>Record : <strong><?PHP echo record(); ?></strong>data mahasiswa</h4>

<div>
<table border="1">
<?PHPdataset(); ?>
</table>
</div>

</div>

<div id="right">
```

```
<h2>INPUTKAN TITIK PUSAT</h2>
<table border="1" width="100px">
<form name="dataset" id="dataset" action="proses.PHP" method="post">
<tr align="center"><td colspan="2">CLUSTER 1 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td align="center"><input type="number" name="kota1" required ></td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td><input type="number" name="jurusan1" required ></td>
</tr>
<tr>
<td> IPK </td>
<td><input type="decimal" name="ipk1" required ></td>
</tr>

<tr align="center"><td colspan="2">CLUSTER 2 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td><input type="number" name="kota2" required ></td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td><input type="number" name="jurusan2" required ></td>
</tr>
<tr>
<td> IPK </td>
<td><input type="decimal" name="ipk2" required ></td>
</tr>
<br/>

<tr align="center"><td colspan="2">CLUSTER 3 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td><input type="number" name="kota3" required ></td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td><input type="number" name="jurusan3" required ></td>
</tr>
<tr>
<td> IPK </td>
<td><input type="decimal" name="ipk3" required ></td>
</tr>
<br/>
```

```
</table>
<br />
<a href="dataset.PHP?kd=otomatis"><button> * RANDOM * </button></a>
<input type="submit" name="submit" value="PROSES" />
</form>
```

<p>Dalam tahap ini data sudah melalui tahap preprocessing dan data sudah siap untuk diolah kedalam proses K-Means Clustering. langkah selanjutnya yaitu menentukan

titik pusat / centroid didalam setiap Cluster yang tersedia</p>

<p> *Random
Kita bisa menentukan titik pusat cluster sesuai keinginan kita ataupun bisa dengan bantuan random, yaitu menset titik pusat cluster secara

 acak / random dan sudah diatur oleh sistem untuk nilai akan bernilai acak. untuk cluster 1 kurang dari cluster 2 dan cluster 3 lebih dari cluster 2.</p>

```
<h2>Contact Me</h2>
```

```
<ul>
```

```
<li><a>085729889007 </a></li>
```

```
<li><a>bayuresi052@gmail.com</a></li>
```

```
</ul>
```

```
</div>
```

```
</div>
```

```
<div id="footer">
```

```
<p>Copyright &copy; 2016 SKRIPSI Teknik Informatika 2012 </p>
```

```
</div>
```

```
</body>
```

```
</html>
```

```
<?PHP
```

```
//----- OTOMATIS KLIK -
-----//
```

```
} elseif ($_GET['kd'] == 'otomatis') { ?>
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

```
<html xmlns="http://www.w3.org/1999/xhtml">
```

```
<head>
```

```
<title>K-MEANS CLUSTERING</title>
```

```
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
```

```
<link href="default.css" rel="stylesheet" type="text/css" />
```



```

</head>
<body>
<div id="header">
<h1>K-Means Clustering APP<br />
<span class="text1">metode Data Mining</span></h1>
</div>
<div id="menu">
<ul>
<li><a href="#">Home</a></li>
<li><a href="#">APP K-Means</a></li>
<li><a href="#">bantuan</a></li>
</ul>
</div>
<div id="content">
<div id="left">

<h2>Dataset yang diolah</h2>
<h4>Record : <strong><?PHP echo record(); ?></strong>data mahasiswa</h4>

<div>
<table border="1">
<?PHPdataset(); ?>
</table>
</div>

</div>

<div id="right">
<h2 align="center">INPUTKAN TITIK PUSAT 'RANDOM'</h2>
<table border="1" width="210px">
<form name="dataset" id="dataset" action="proses.PHP" method="post">
<tr><td colspan="2" align="center"><a href=""></a><strong>Refresh bila centroid tidak selesai atau
eror</strong></td></tr>
<tr align="center"><td colspan="2">CLUSTER 1 </td></tr>
<tr>
<td width="130px"> ASAL KOTA</td>
<td align="center"><input type="number" name="kota1" value="<?PHP
otomatis('frek_kota',3,1,"); ?>" hidden ><?PHP otomatis('frek_kota',3,1,");
?></td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td align="center"><input type="number" name="jurusan1" value="<?PHP
otomatis('frek_jurusan',3,1,"); ?>" hidden ><?PHP

```

```

otomatis('frek_jurusan',3,1,"); ?></td>
</tr>
<tr>
<td> IPK </td>
<td align="center"><input type="decimal" name="ipk1" value="<?PHP
otomatis('dataset',4,2,"); ?>" hidden ><?PHP otomatis('dataset',4,2,"); ?></td>
</tr>

<tr align="center"><td colspan="2">CLUSTER 2 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td align="center"><input type="number" name="kota2" value="<?PHP echo
otomatis('frek_kota',1,",""); ?>" hidden ><?PHP echo otomatis('frek_kota',1,","");
?></td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td align="center"><input type="number" name="jurusan2" value="<?PHP
echo otomatis('frek_jurusan',1,",""); ?>" hidden ><?PHP echo
otomatis('frek_jurusan',1,",""); ?></td>
</tr>
<tr>
<td> IPK </td>
<td align="center"><input type="decimal" name="ipk2" value="<?PHP echo
otomatis('dataset',2,",""); ?>" hidden ><?PHP echo otomatis('dataset',2,","");
?></td>
</tr>
<br/>

<tr align="center"><td colspan="2">CLUSTER 3 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td align="center"><input type="number" name="kota3" value="<?PHP
otomatis('frek_kota',5,1,'inikota'); ?>" hidden ><?PHP
otomatis('frek_kota',5,1,'inikota'); ?></td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td align="center"><input type="number" name="jurusan3" value="<?PHP
otomatis('frek_jurusan',5,1,'inijur'); ?>" hidden ><?PHP
otomatis('frek_jurusan',5,1,'inijur'); ?></td>
</tr>
<tr>
<td> IPK </td>
<td align="center"><input type="decimal" name="ipk3" value="<?PHP
otomatis('dataset',6,2,"); ?>" hidden ><?PHP otomatis('dataset',6,2,"); ?></td>

```

```
</tr>
<br/>

</table>
<br />
<td><input type="submit" name="submit" value="PROSES" /></td>
</form>
```

<p>
 Dalam tahap ini data sudah melalui tahap preprosecing dan data sudah siap untuk diolah kedalam proses K-Means Clustering. langkah selanjutnya yaitu menentukan

titik pusat / centroid didalam setiap Cluster yang tersedia</p>

<p> *Random
Kita bisa menentukan titik pusat cluster sesuai keinginan kita ataupun bisa dengan bantuan random, yaitu menset titik pusat cluster secara

 acak / random dan sudah diatur oleh sistem untuk nilai akan bernilai acak. untuk cluster 1 kurang dari cluster 2 dan cluster 3 lebih dari cluster 2.</p>

<h2>Contact Me</h2>

<a>085729889007

<a>bayuresi052@gmail.com

</div>

</div>

<div id="footer">

<p>Copyright © 2016 SKRIPSI Teknik Informatika 2012 </p>

</div>

</body>

</html>

<?PHP

} else {

echo "<META http-equiv='Refresh' content='0; URL=kmeans.PHP'>";

}

?>

LAMPIRAN I Source Code PHP (proses.PHP)

```
<?PHP
MySQL_connect('localhost', 'root', '');
MySQL_select_db('skripsi');

include ('function.PHP');

if(isset($_POST['submit'])){ ?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>K-MEANS CLUSTERING</title>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link href="default.css" rel="stylesheet" type="text/css" />
</head>
<body>
<div id="header">
<h1>K-Means Clustering APP<br />
<span class="text1">metode Data Mining</span></h1>
</div>
<div id="menu">
<ul>
<li><a href="#">Home</a></li>
<li><a href="#">APP K-Means</a></li>
<li><a href="#">bantuan</a></li>
</ul>
</div>
<div id="content">
<div id="left">

<h2> PROSES <?PHP nomorproses(1); ?></h2>
<h4>Record : <strong><?PHP echo record(); ?></strong>data mahasiswa</h4>

</div>
<?PHP

    $kota1   = $_POST['kota1'];
    $jurusan1 = $_POST['jurusan1'];
    $ipk1    = $_POST['ipk1'];
    $kota2   = $_POST['kota2'];
    $jurusan2 = $_POST['jurusan2'];
    $ipk2    = $_POST['ipk2'];
    $kota3   = $_POST['kota3'];
```

```

$jurusan3 = $_POST['jurusan3'];
$ipk3     = $_POST['ipk3'];

echo "<tabel border='1' width='480px'>";

kmeans($kota1,$jurusan1,$ipk1,$kota2,$jurusan2,$ipk2,$kota3,$jurusan3,$ipk3)
;
echo "</tabel>"; ?>
</div>

</div>

<div id="right">
<h2>Pusat Cluster Baru</h2>
<?PHP newcluster(); ?>

<p> KETERANGAN : </p>
<p>I.K = Inisial Kota</p>
<p>I.J = Inisial Jurusan</p>
<p>C1 = Cluster 1</p>
<p>Maksud dari centang tersebut adalah data pertama / data baris pertama
masuk kedalam cluster 1 karena cluster 1 adalah jarak paling dekat
diantara cluster yang lain, dan begitu seterusnya</p>
<p> Proses K-Means bisa mengalami <strong>berkali kali</strong> proses
tergantung dengan data yang diolah, sampai akhirnya menemukan titik pusat
cluster yang tidak berubah ubah lagi</p>

<h2>Contact Me</h2>
<ul>
<li><a>085729889007 </a></li>
<li><a>bayuresi052@gmail.com</a></li>
</ul>
</div>
</div>
<div id="footer">
<p>Copyright &copy; 2016 SKRIPSI Teknik Informatika 2012 </p>
</div>
</body>
</html>

<?PHP
} else {

echo "<META http-equiv='Refresh' content='0; URL=hasil.PHP?id=hasil'>";

```

```
}  
?>
```

LAMPIRAN J Source Code PHP (hasil.PHP)

```
<?PHP  
MySQL_connect('localhost', 'root', '');  
MySQL_select_db('skripsi');  
  
include ('function.PHP');  
  
if ($_GET['id'] == 'hasil'){ ?>  
  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml">  
<head>  
<title>K-MEANS CLUSTERING</title>  
<meta http-equiv="content-type" content="text/html; charset=utf-8" />  
<link href="default.css" rel="stylesheet" type="text/css" />  
</head>  
<body>  
<div id="header">  
<h1>K-Means Clustering APP<br />  
<span class="text1">metode Data Mining</span></h1>  
</div>  
<div id="menu">  
<ul>  
<li><a href="#">Home</a></li>  
<li><a href="#">APP K-Means</a></li>  
<li><a href="#">bantuan</a></li>  
</ul>  
</div>  
<div id="content">  
<div id="left">  
  
<h2><span> HASIL PROSES K-MEANS CLUSTERING</span></h2>  
<h3>  
<strong>  
<?PHP echo "record : "; echo record(); echo " data mahasiswa | ";?>  
<?PHP $sql = "SELECT MAX(id_nomor) as proses FROM nomor"; $result =  
MySQL_query($sql); $row = MySQL_fetch_array($result); $p = $row['proses'];
```

```

echo "$p";?> kali proses |
</strong>
<a href="hasil.PHP?id=exit"><button> EXIT </button></a>
</h3>

<tabel border="1" width="430px">
<tr height="25px"><td colspan="2" align="center"><strong> CLUSTER
1</strong></td></tr>
<tr height="23px"><td colspan="2" align="center"> CLUSTER 1 TERDIRI
DARI <strong><?PHP recordcluster('clustersatu'); ?></strong> MAHASISWA
YANG TERDIRI DARI </td></tr>
<tr><td align="center"> ASAL KOTA </td><td align="center"> JURUSAN
SEKOLAH </td></tr>
<tr>
<td>
<tabel>
<tr>
<th align="left"> KOTA </th>
<th align="center"> JUMLAH </th>
</tr>
<?PHP hasil('inikota','clustersatu','frek_kota','inikota'); ?>
</tabel>
</td>

<td>
<tabel height="340px">
<tr>
<th align="left"> JURUSAN SEKOLAH </th>
<th align="center"> JUMLAH </th>
</tr>
<?PHP hasil('inijur','clustersatu','frek_jurusan','inijur'); ?>
</tabel>
</td>
</tr>
<tr height="25px"><td colspan="2" align="center"> RATA-RATA IPK
CLUSTER 1 <strong> : <?PHP rata_rata_ipk('clustersatu');
?></strong></td></tr>
</tabel>
<br />

<tabel border="1" width="430px">
<tr height="25px"><td colspan="2" align="center"><strong> CLUSTER
2</strong></td></tr>
<tr height="23px"><td colspan="2" align="center"> CLUSTER 2 TERDIRI
DARI <strong><?PHP recordcluster('clusterdua'); ?></strong> MAHASISWA
YANG TERDIRI DARI </td></tr>

```

```

<tr><td align="center"> ASAL KOTA </td><td align="center"> JURUSAN
SEKOLAH </td></tr>
<tr>
<td>
<tabel height="503px">
<tr>
<th align="left"> KOTA </th>
<th align="center"> JUMLAH </th>
</tr>
<?PHP hasil('inikota','clusterdua','frek_kota','inikota'); ?>
</tabel>
</td>

<td>
<tabel>
<tr>
<th align="left"> JURUSAN SEKOLAH </th>
<th align="center"> JUMLAH </th>
</tr>
<?PHP hasil('inijur','clusterdua','frek_jurusan','inijur'); ?>
</tabel>
</td>
<tr height="25px"><td colspan="2" align="center"> RATA-RATA IPK
CLUSTER 2 <strong> : <?PHP rata_rata_ipk('clusterdua');
?></strong></td></tr>
</tabel>
<br />

<tabel border="1" width="430px">
<tr height="25px"><td colspan="2" align="center"><strong> CLUSTER
3</strong></td></tr>
<tr height="23px"><td colspan="2" align="center"> CLUSTER 3 TERDIRI
DARI <strong><?PHP recordcluster('clustertiga'); ?></strong> MAHASISWA
YANG TERDIRI DARI </td></tr>
<tr><td align="center"> ASAL KOTA </td><td align="center"> JURUSAN
SEKOLAH </td></tr>
<tr>
<td>
<tabel>
<tr>
<th align="left"> KOTA </th>
<th align="center"> JUMLAH </th>
</tr>
<?PHP hasil('inikota','clustertiga','frek_kota','inikota'); ?>
</tabel>

```



```

</td>

<td>
<tabel height="684px">
<tr>
<th align="left"> JURUSAN SEKOLAH </th>
<th align="center"> JUMLAH </th>
</tr>
<?PHP hasil('inijur','clustertiga','frek_jurusan','inijur'); ?>
</tabel>
</td>
</tr>
<tr height="25px"><td colspan="2" align="center"> RATA-RATA IPK
CLUSTER 3 <strong> : <?PHP rata_rata_ipk('clustertiga');
?></strong></td></tr>
</tabel>

</div>

<div id="right">
<h2>HASIL K-MEANS</h2>

<p> Form ini adalah proses akhir untuk menyajikan output atau hasil dari proses
K-Means Clustering, yaitu data mentah yang diimportkan sudah
dikelompokkan kedalam cluster-cluster yg terbentuk sesuai jarak terdekat
masing2.</p>
<p> Dari hasil proses K-Means tersebut kita memperoleh informasi bahwa
cluster 1 adalah cluster dengan IPK rata-rata paling tinggi dibandingkan cluster 2
ataupun cluster 3,
jadi cluster 1 adalah cluster yang pantas untuk diprioritaskan untuk promosi
yang lebih intensif lagi.</p>

<h2>Contact Me</h2>
<ul>
<li><a>085729889007 </a></li>
<li><a>bayuresi052@gmail.com</a></li>
</ul>
</div>
</div>
<div id="footer">
<p>Copyright &copy; 2016 SKRIPSI Teknik Informatika 2012 </p>
</div>
</body>
</html>

```

```

<?PHP
} elseif ($_GET['id'] == 'exit') {

    $truncate = "TRUNCATE TABELcentroid_lama";
    $truncate2 = "TRUNCATE TABEL nomor";
    $a = MySQL_query($truncate);
    $b = MySQL_query($truncate2);

    if ($a AND $b) {
        echo "<META http-equiv='Refresh' content='0; URL=kmeans.PHP'>";
    } else {
        echo "gagal";
    }

} else {

    echo "<META http-equiv='Refresh' content='0; URL=kmeans.PHP'>";

}
?>

```

LAMPIRAN K Source Code PHP (function.PHP)

```

<?PHP

//MENCARI INISIAL DISETIAP KOTA_ASAL DAN JURUSAN ASAL
function vlookup($kota,$table,$sub_tabel,$col){

    //mencari nama id untuk tabel dataset
    $res = MySQL_query("SELECT * FROM $table");
    $id = MySQL_field_name($res,0);

    //query berdasarkan id input
    $result = MySQL_query("SELECT * FROM $table WHERE $sub_tabel = '$kota' ");
    if($result){
        $row = MySQL_fetch_array($result);
        $fieldnames = $row[$col];
    }
    return $fieldnames;
}

//MENCARI FREKUENSI SETIAP KOTA_ASAL DAN
JURUSAN_SEKOLAH + INISIAL

```

```

function frekuensi($tabel,$berdasarkan,$iniapa){

//menghapus semua data pada tabel frek_kota dan frek_jurusan
//KOTA
$truncate ="TRUNCATE TABLE $tabel";
MySQL_query($truncate);

//MENGHITUNG FREKUENSI KOTA + INISIAL
$inisial=1;
$sql = "SELECT $berdasarkan, COUNT($berdasarkan) as frekuensi FROM
dataset GROUP BY $berdasarkan ORDER BY frekuensi DESC , $berdasarkan
ASC";
$result = MySQL_query($sql);
while($row = MySQL_fetch_array($result)){
    $frekuensi = $row['frekuensi'];
    $val      = $row[$berdasarkan];
    $otomatis = $frekuensi*$inisial;

    $q = "INSERT INTO $tabel ($iniapa,$berdasarkan,frekuensi,otomatis)
VALUES ('$inisial','$val','$frekuensi','$otomatis)";
    $r = MySQL_query($q);

    $inisial++;
}
}

//MENAMPILKAN BERDASARKAN FREKUENSI TERBESAR KE
TERKECIL
function inisial($tabel,$atribut,$atr,$iniapa){

    echo "<tr>
<th align='center'>NO</th>
<th align='center'>$atr</th>
<th align='center'>FREK</th>
<th align='center'>INISIAL</th>
</tr>";

    $no = 1;

    $sql = "SELECT * FROM $tabel ORDER BY frekuensi DESC ,
$iniapa ASC";

    $result = MySQL_query($sql);
    while($row = MySQL_fetch_array($result)){
        $frekuensi = $row['frekuensi'];

```

```

        $val    = $row[$atribut];
        $inisial = $row[$iniapa];

        echo "<tr>
<td align='center'>$no</td>
<td>$val</td>
<td align='center'>$frekuensi</td>
<td align='center'>$inisial</td>
</tr>";

        $no++;
    }
}

//MENAMPILKAN DATASET YANG AKAN DIOLAH
function dataset(){

    echo "<tr>
<th align='center'>NO</th>
<th align='center'>KOTA ASAL</th>
<th align='center'>INISIAL</th>
<th align='center'>JURUSAN</th>
<th align='center'>INISIAL</th>
<th align='center'>IPK</th>
</tr>";

    $no = 1;
    $inisial = 1;

    $sql = "SELECT * FROM dataset";

    $result = MySQL_query($sql);
    while($row = MySQL_fetch_array($result)){
        $kota    = $row['kota_asal'];
        $jurusan  = $row['jurusan_sekolah'];
        $ipk     = $row['ipk'];
        $inikota = vlookup($kota,'frek_kota','kota_asal',0);
        $inijur  =
vlookup($jurusan,'frek_jurusan','jurusan_sekolah',0);

        echo "<tr>
<form method='POST' align='center'>

```

```

<td> $no. </td>
<td> $kota </td>
<td align='center'> $inikota </td>
<td> $jurusan </td>
<td align='center'> $inijur </td>
<td align='center'> $ipk </td>
</form>
</tr>";

        $inisial++;
        $no++;
    }
}

function record(){

$result = MySQL_query("SELECT * FROM dataset");
$num_rows = MySQL_num_rows($result);
// Display the results
return $num_rows;

}

function otomatis($tabel,$kode,$kode3_kode,$sub_tabel){

//Menset Centroid Cluster 2 dengan Rata-Rata Atribut pada Cluster 2
if ($kode == 1) {
    $sql = "SELECT SUM(otomatis) as jumlah FROM $tabel";
    $result = MySQL_query($sql);
    $row = MySQL_fetch_array($result);
    $jumlah = $row['jumlah'];
    $pembagi = record();
    $auto = $jumlah/$pembagi;

    $sval = number_format($auto,2);
    return $sval;
//Centroid dengan Rata-Rata pada Cluster 2
}elseif ($kode == 2) {
    $sql = "SELECT AVG(ipk) as rata_rata FROM $tabel";
    $result = MySQL_query($sql);
    $row = MySQL_fetch_array($result);
    $sval = $row['rata_rata'];

    $sval = number_format($sval,2);
    return "$sval";
}
}

```

```

//Menentukan Centroid secara Random / Acak pada Atribut Kota Asal dan
Jurusan Asal Mahasiswa untuk cluster 1
}elseif ($kode == 3) {
    $cluster2 = otomatis($tabel,$kode3_kode,"");
    $clus1 = number_format((rand()*$cluster2),2);
    if ($clus1 == '0') {
        otomatis($tabel,3,1,"");
    } else {
        echo $clus1;
    }
}
//Menentukan Centroid secara Random / Acak pada Atribut IPK Mahasiswa
untuk cluster 1
}elseif ($kode == 4) {
    $cluster2 = otomatis($tabel,$kode3_kode,"");
    $acak = (rand()*$cluster2);
    $clus1 = number_format($acak,2);
    if ($clus1 == '0') {
        otomatis($tabel,4,2,"");
    } else {
        echo $clus1;
    }
}

//Menentukan Centroid secara Random / Acak pada Atribut Kota Asal dan
Jurusan Asal Mahasiswa untuk cluster 3
}elseif ($kode == 5) {
    $sql = "SELECT MAX($sub_tabel) as max FROM $tabel";
    $result = MySQL_query($sql);
    $row = MySQL_fetch_array($result);
    $max = $row['max'];

    $clus1 = number_format((rand()*$max),2);

    $cluster2 = otomatis($tabel,$kode3_kode,"");

    if ($clus1 <= $cluster2) {
        otomatis($tabel,5,1,$sub_tabel);
    } else {
        echo $clus1;
    }
}
//Menentukan Centroid secara Random / Acak pada Atribut IPK Mahasiswa
untuk cluster 1
}elseif ($kode == 6) {

    $max_ipk = 10;

```

```

$clus1 = number_format((rand()*$max_ipk),2);

$cluster2 = otomatis($tabel,$kode3_kode,"");

if ($clus1 < $cluster2) {
    otomatis($tabel,6,2,"");
} else {
    echo $clus1;
}
}

}

//-----PROSES K-MEANS
CLUSTERING-----//

//MEMBERI NOMOR PROSES
function nomorproses($proses){

    $q = "INSERT INTO nomor (nomor) VALUES ($proses)";
    $r = MySQL_query($q);

    $sql = "SELECT MAX(id_nomor) as proses FROM nomor";
    $result = MySQL_query($sql);
    $row = MySQL_fetch_array($result);

    $p = $row['proses'];

    echo "$p";

}

//PROSES K-MEANS
function
kmeans($kota1,$jurusan1,$ipk1,$kota2,$jurusan2,$ipk2,$kota3,$jurusan3,$ipk3)
{

    $truncate ="TRUNCATE TABLE proses";
    MySQL_query($truncate);

    $sko1 = number_format($kota1,2); $jur1 = number_format($jurusan1,2); $ip1 =
number_format($ipk1,2);
    $sko2 = number_format($kota2,2); $jur2 = number_format($jurusan2,2); $ip2 =

```

```

number_format($ipk2,2);
$ko3 = number_format($kota3,2); $jur3 = number_format($jurusan3,2); $ip3 =
number_format($ipk3,2);

    echo "    <tr>
<th rowspan='2' align='center'> NO </th>
<th rowspan='2' align='center'> I.K </th>
<th rowspan='2' align='center'> I.J </th>
<th rowspan='2' align='center'> IPK </th>
<th colspan='3' align='center'>CENTROID C1</th>
<th colspan='3' align='center'>CENTROID C2</th>
<th colspan='3' align='center'>CENTROID C3</th>
<th rowspan='2' align='center'>C1</th>
<th rowspan='2' align='center'>C2</th>
<th rowspan='2' align='center'>C3</th>
</tr>
<tr>
<th>$ko1</th>
<th>$jur1</th>
<th>$ip1</th>
<th>$ko2</th>
<th>$jur2</th>
<th>$ip2</th>
<th>$ko3</th>
<th>$jur3</th>
<th>$ip3</th>
</tr>";

    $no = 1;

    $sql = "SELECT * FROM dataset";

    $result = MySQL_query($sql);
    while($row = MySQL_fetch_array($result)){
        $kota    = $row['kota_asal'];
        $jurusan  = $row['jurusan_sekolah'];
        $ipk     = $row['ipk'];
        $inikota  = vlookup($kota,'frek_kota','kota_asal',0);
        $inijur   =
vlookup($jurusan,'frek_jurusan','jurusan_sekolah',0);
        $jarak1   =
distance($kota1,$jurusan1,$ipk1,$inikota,$inijur,$ipk);
        $jarak2   =
distance($kota2,$jurusan2,$ipk2,$inikota,$inijur,$ipk);
        $jarak3   =
distance($kota3,$jurusan3,$ipk3,$inikota,$inijur,$ipk);

```



```

        $clustering = clustering($jarak1,$jarak2,$jarak3);
        $clustering2 = clustering2($jarak1,$jarak2,$jarak3);
        $clustering3 = clustering3($jarak1,$jarak2,$jarak3);

        echo "<tr>
<form method='POST' align='center'>
<td align='center'> $no. </td>
<td align='center'>$inikota</td>
<td align='center'>$inijur</td>
<td align='center'>$ipk</td>
<td align='center' colspan='3'>$jarak1</td>
<td align='center' colspan='3'>$jarak2</td>
<td align='center' colspan='3'>$jarak3</td>
<td align='center'>$clustering</td>
<td align='center'>$clustering2</td>
<td align='center'>$clustering3</td>
</form>
</tr>";

        $q = "INSERT INTO proses
(inikota,inijur,ipk,clustersatu,clusterdua,clustertiga) VALUES
('$inikota','$inijur','$ipk','$clustering','$clustering2','$clustering3')";
        $r = "INSERT INTO centroid_lama (centroid_lama)
VALUES
('$kota1'),('$jurusan1'),('$ipk1'),('kota2'),('$jurusan2'),('$ipk2'),('kota3'),('$jurusan3'),('$ipk3)";
        $s = MySQL_query($q);
        $t = MySQL_query($r);

        $no++;
    }
}

//MENGHITUNG JARAK DATA KE CLUSTER
function distance($k,$j,$i,$inikota,$inijur,$ipk){

    $jarak = SQRT( (pow(($inikota - $k),2)) + (pow(($inijur - $j),2)) + (pow(($ipk - $i),2)) );
    return number_format($jarak,7);

}

//MENGELOMPOKKAN DATA KEDALAM CLUSTER BERDASARKAN
JARAK TERDEKAT DATA DENGAN PUSAT CLUSTER (CENTROID)
function clustering($a,$b,$c){

```

```

IF ($a < $b AND $a < $c){
    return "v";
}if ($b < $c AND $b < $a) {
    return "";
} if ($c < $a AND $c < $b) {
    return "";
} else {
    return "fail";
}
}
function clustering2($a,$b,$c){

IF ($a < $b AND $a < $c){
    return "";
}if ($b < $c AND $b < $a) {
    return "v";
} if ($c < $a AND $c < $b) {
    return "";
} else {
    return "fail";
}
}
function clustering3($a,$b,$c){

IF ($a < $b AND $a < $c){
    return "";
}if ($b < $c AND $b < $a) {
    return "";
} if ($c < $a AND $c < $b) {
    return "v";
} else {
    return "fail";
}
}

//MENCARI CLUSTER BARU (RATA-RATA)
function rata_rata($colom,$cluster){

    $sql = "SELECT AVG($colom) as rata_rata FROM proses WHERE $cluster='v'";
    $result = MySQL_query($sql);
    $row = MySQL_fetch_array($result);

    $val = number_format($row['rata_rata'],7);

```

```

return $val;

}

//MENCARI PUSAT (CENTROID) BARU DISETIAP CLUSTER
function newcluster(){

//CLUSTER1
$kota1 = rata_rata('inikota','clustersatu'); $jur1 = rata_rata('inijur','clustersatu');
$ipk1 = rata_rata('ipk','clustersatu');
//CLUSTER2
$kota2 = rata_rata('inikota','clusterdua'); $jur2 = rata_rata('inijur','clusterdua');
$ipk2 = rata_rata('ipk','clusterdua');
//CLUSTER3
$kota3 = rata_rata('inikota','clustertiga'); $jur3 = rata_rata('inijur','clustertiga');
$ipk3 = rata_rata('ipk','clustertiga');

echo" <table border='1' width='210px'>
<form name='newcluster' id='newcluster' action='proses.PHP' method='post' >
<tr align='center'><td colspan='2'>CLUSTER 1 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td><input type='decimal' name='kota1' value='$kota1' hidden > $kota1 </td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td><input type='decimal' name='jurusan1' value='$jur1' hidden > $jur1 </td>
</tr>
<tr>
<td> IPK </td>
<td><input type='decimal' name='ipk1' value='$ipk1' hidden > $ipk1 </td>
</tr>
<br/>

<tr align='center'><td colspan='2'>CLUSTER 2 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td><input type='decimal' name='kota2' value='$kota2' hidden > $kota2 </td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td><input type='decimal' name='jurusan2' value='$jur2' hidden > $jur2 </td>
</tr>
<tr>
<td> IPK </td>
<td><input type='decimal' name='ipk2' value='$ipk2' hidden > $ipk2 </td>

```

```

</tr>
<br/>

<tr align='center'><td colspan='2'>CLUSTER 3 </td></tr>
<tr>
<td> ASAL KOTA</td>
<td><input type='decimal' name='kota3' value='$kota3' hidden > $kota3 </td>
</tr>
<tr>
<td> JURUSAN SEKOLAH</td>
<td><input type='decimal' name='jurusan3' value='$jur3' hidden > $jur3 </td>
</tr>
<tr>
<td> IPK </td>
<td><input type='decimal' name='ipk3' value='$ipk3' hidden > $ipk3 </td>
</tr>
<br/>

</table> ";

    $cek =
verifikasi($kota1,$jur1,$ipk1,$kota2,$jur2,$ipk2,$kota3,$jur3,$ipk3);

    if ($cek == 'Lanjut') {
        echo "<br />
<td colspan='4'><input type='submit' name='submit' value='Next Proses' /></td>
</form>";
    } elseif ($cek == 'stop') {
        echo "<br />
<td colspan='4'><h2><strong> PROSES SELESAI !! </strong></h2></td>
<h1><span class='text1'>hasil proses k-means</span><br/><input type='submit'
name='hasil' value='HASIL' /></h1>";
    }
}

//MEMVERIFIKASI APAKAH CENTROID BARU YG DIDAPATKAN
SAMA ATAU BEDA DENGAN CENTROID YANG LAMA
function verifikasi($a,$b,$c,$a2,$b2,$c2,$a3,$b3,$c3) {

    $resultsatu = MySQL_query("SELECT * FROM centroid_lama");
    $num_rowsatu = MySQL_num_rows($resultsatu);
    $pertama = $num_rowsatu;

    $q = "INSERT IGNORE INTO centroid_lama (centroid_lama) VALUES

```

```

($a),($b),($c),($a2),($b2),($c2),($a3),($b3),($c3)";
$cek = MySQL_query($q);

$resultkedua = MySQL_query("SELECT * FROM centroid_lama");
$num_rowskedua = MySQL_num_rows($resultkedua);
$kedua = $num_rowskedua;

if ($pertama < $kedua) {
    return "Lanjut";
}else{
    return "stop";
}
}

//----- HASIL PROSES K-MEANS CLUSTERING -----//

//HASIL DARI PROSES K_MEANS
function hasil($sub_tabel,$cluster,$tabel_vlookup,$iniapa){

$sql2 = "SELECT $sub_tabel, COUNT($sub_tabel) as jumlah FROM proses
WHERE $cluster = 'v' GROUP BY $sub_tabel ORDER BY jumlah DESC ";
$result2 = MySQL_query($sql2);
while($row2 = MySQL_fetch_array($result2)){

    $jumlah = $row2['jumlah'];
    $inisial = $row2[$sub_tabel];
    $val = vlookup($inisial,$tabel_vlookup,$iniapa,1);

    echo "<tr>
<form>
<td> $val </td>
<td align='center'> $jumlah </td>
</form>
</tr>";
}

}

//JUMLAH DATA MAHASISWA SETIAP HASIL CLUSTER
function recordcluster($cluster){

$result = MySQL_query("SELECT * FROM proses WHERE $cluster = 'v' ");
$num_rows = MySQL_num_rows($result);

```

```
// Display the results
echo $num_rows;

}

//MENGHITUNG RATA-RATA SETIAP HASIL CLUSTER
function rata_rata_ipk($cluster){

    $sql = "SELECT AVG(ipk) as rata_rata FROM proses WHERE $cluster='v' ";
    $result = MySQL_query($sql);
    $row = MySQL_fetch_array($result);

    $val = number_format($row['rata_rata'],2);
    echo $val;

}

?>
```



CURICULUM VITAE

Identitas diri

Nama Lengkap : Bayu Resi Indrawan
Tempat, Tanggal Lahir : Bantul, 25 Januari 1994
Kewarga Negaraan : Indonesia
Agama : Islam
Jenis Kelamin : Laki-laki
Email : bayuresi052@gmail.com
Kontak : 085729889007

Riwayat Pendidikan

2000 - 2006 : SD Negeri Kabregan
2006 - 2009 : SMP Negeri 1 Piyungan
2009 - 2012 : SMA Negeri 1 Sewon
2012 – 2018 : S1 Teknik Informatika UIN Sunan Kalijaga
Yogyakarta

Pengalaman Organisasi

- ❖ Muda Bhakti Sandeyan Kulon
- ❖ HIMA (Himpunan Mahasiswa) Teknik Informatika