

**DATA MINING UNTUK MENGETAHUI LAMA STUDI
MAHASISWA DI UNIVERSITAS MUHAMMADIYAH
PONOROGO**

SKRIPSI

Diajukan dan Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Jenjang Strata Satu (S1)
Pada Program Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Ponorogo



Disusun Oleh:

Aji Pratama Vektor Muhtariz

11531185

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK INFORMATIKA
UNIVERSITAS MUHAMMADIYAH PONOROGO**

2015

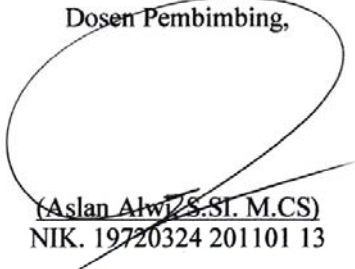
HALAMAN PENGESAHAN

Nama : Aji Pratama Vektor Muhtariz
NIM : 11531185
Program Studi : Teknik Informatika
Fakultas : Teknik
Judul Skripsi : Data Mining Untuk Mengetahui Lama Studi Mahasiswa Di
Universitas Muhammadiyah Ponorogo

Isi dan formatnya telah disetujui dan dinyatakan memenuhi syarat untuk melengkapi persyaratan guna memperoleh Gelar Sarjana pada Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Ponorogo

Ponorogo, September 2015

Menyetujui,
Dosen Pembimbing,


(Aslan Alwi, S.SI, M.CS)
NIK. 19720324 201101 13

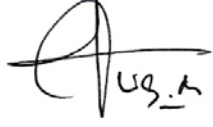
Mengetahui,

Dekan Fakultas Teknik,



(Ir. Aliyadi, MM, M.Kom)
NIK. 19640103 199009 12

Ketua Program Studi
Teknik Informatika,


(Munirah Muslim, S.Kom, MT)
NIK. 19791107 200912 13

HALAMAN BERITA ACARA UJIAN

Nama : Aji Pratama Vektor Muhtariz
NIM : 11531185
Program Studi : Teknik Informatika
Fakultas : Teknik
Judul Skripsi : Data Mining Untuk Mengetahui Lama Studi Mahasiswa Di
Universitas Muhammadiyah Ponorogo

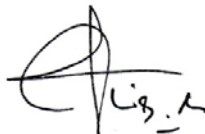
Telah diuji dan dipertahankan dihadapan

Dosen penguji tugas akhir jenjang Strata Satu (S1) pada:

Hari : Selasa
Tanggal : 8 September 2015
Nilai :
Ponorogo, September 2015

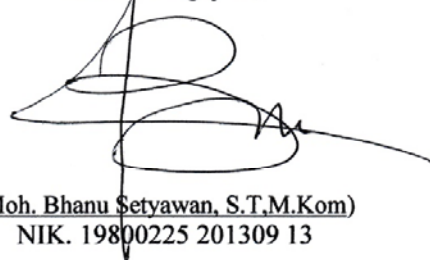
Dosen Penguji

Dosen Penguji I,



(Munirah Muslim, S.Kom, MT)
NIK. 19791107 200912 13

Dosen Penguji II,



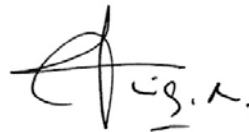
(Moh. Bhanu Setyawan, S.T.M.Kom)
NIK. 19800225 201309 13

Dekan Fakultas Teknik,



(Ir. Aliyadi, MM, M.Kom)
NIK. 19640103 199009 12

Ketua Program Studi
Teknik Informatika,




(Munirah Muslim, S.Kom, MT)
NIK. 19791107 200912 13

BERITA ACARA

BIMBINGAN SKRIPSI

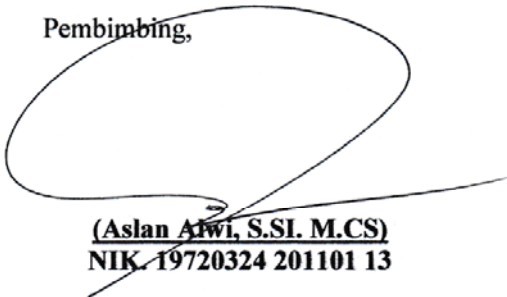
1. Nama : Aji Pratama Vektor Muhtariz
2. NIM : 11531185
3. Program Studi : Informatika
4. Fakultas : Teknik
5. Judul Skripsi : Data Mining Untuk Mengetahui Lama Studi Mahasiswa Di Universitas Muhammadiyah Ponorogo
6. Dosen Pembimbing : Aslan Alwi, S.SI. M.CS
7. Konsultasi :

NO	TANGGAL	URAIAN	TANDA-TANGAN
1.	06 Agustus 2015	Ganti judul "lama study"	
2.	08 Agustus 2015	rencana struktur data	
3.	11. Agustus 2015	tes-tes data menggunakan k-mean / clustering	
4	12 Agustus 2015	buat rincian tabel dengan k-mean	
5	13 Agustus 2015	revisi BAB IV	
6	15 Agustus 2015	rencana desain pembuatan mesin learning	
7	20 Agustus 2015	Beri penjelasan kode program	
8	21 Agustus 2015	Acc. Ujian	

8. Tgl. Pengajuan :
9. Tgl. Pengesahan :

Ponorogo, September 2016

Pembimbing,


(Aslan Alwi, S.SI. M.CS)
NIK. 19720324 201101 13

Motto

“Yang penting itu bukan apa yang kita ketahui, tapi apa yang kita bersedia pelajari.”

“Be a good moslem or die as syuhada”

HALAMAN PERSEMBAHAN

Kupersembahkan karyaku ini untuk:

Kedua orang tuaku Bapak Sumaji dan Ibu Sri Alminatin.

Adikku yang tercinta, Anggri Sigma Nur Rohmi dan Prima Mulqia Irsada.

Kepada sahabat ku yang telah tiada Alm. Andryan Dhery Pratama.

Kepada Sahabat Ex-Niit yang tersisa, Muafa Ahmad S., Yhoni Ardhi Pratama, Imam

Muhtadhin, Agus Kurniawan, Ika Diani, Ayu Istimaroh, Ria Ayu.

Dan segenap Crew Class F angkatan 2011 yang tidak bisa aku sebutkan satu persatu.

DATA MINING UNTUK MENGETAHUI LAMA STUDI MAHASISWA DI UNIVERSITAS MUHAMMADIYAH PONOROGO

Aji Pratama Vektor Muhtariz
11531185

Fakultas Teknik Jurusan Informatika
Universitas Muhammadiyah Ponorogo

ABSTRAKSI

Data mining adalah suatu konsep yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam database besar. *Data mining* adalah bagian dari proses KDD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, *data mining*, dan evaluasi hasil (Maimon dan Last, 2000). KDD secara umum juga dikenal sebagai pangkalan data. *Data mining* dipilih sebagai model yang bagus untuk mengetahui atau memprediksi lama studi mahasiswa di Universitas Muhammadiyah Ponorogo. Dengan ini diharapkan, agar Universitas Muhammadiyah Ponorogo bisa mengurangi lama studi mahasiswa dan mengetahui lama studi mahasiswa.

Kata Kunci: Data Mining, KDD, evaluasi, lama studi

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul Data Mining Untuk Mengetahui Lama Studi Mahasiswa Di Universitas Muhammadiyah Ponorogo. Shalawat serta salam senantiasa penulis panjatkan kepada Nabi Besar Muhammad SAW, yang telah membimbing manusia ke jalan yang benar, yaitu jalan yang di Ridhai Allah SWT.

Penulis menyadari bahwa banyak pihak yang telah berpartisipasi dan membantu dalam menyelesaikan penulisan skripsi ini. Untuk itu, iringan do'a dan ucapan terima kasih yang sebesar-besarnya penulis sampaikan, terutama kepada:

1. Ir. Aliyadi, MM, M. Kom, selaku Dekan fakultas Teknik Universitas Muhammadiyah Ponorogo.
2. Munirah Muslim, S.Kom, MT, selaku Ketua Program Studi fakultas Teknik Universitas Muhammadiyah Ponorogo.
3. Aslan Alwi, S.Si, M.Cs, selaku Dosen Pembimbing atas bimbingan, arahan dan masukan selama penyusunan skripsi.
4. Para dosen Teknik Informatika Universitas Muhammadiyah Ponorogo yang telah memberikan banyak bekal ilmu kepada penulis.
5. Kedua orang tua (Bapak Sumaji dan Ibu Sri Alminatin) yang selalu memberikan semangat dan motivasi baik moril maupun spirituil serta pengorbanan dan perjuangannya yang tak pernah kenal lelah dalam mendidik dan membimbing penulis serta ketulusan do'anya kepada penulis.
6. Sahabat-sahabatku seperjuangan dan teman-teman Teknik Informatika angkatan 2011 yang telah memberikan bantuan, dukungan serta motivasi kepada penulis dalam penyusunan skripsi yang tidak bisa disebutkan satu per satu.

Kekurangan terkait keterbatasan referensi dan ilmu penulis. Oleh sebab itu, penulis mengharapkan saran dan kritik yang bersifat membangun dari pembaca dan dari semua pihak demi kesempurnaan dari skripsi ini.

Penulis berharap semoga skripsi ini bermanfaat bagi para pembaca, dan dapat memberikan kontribusi positif terhadap perkembangan ilmu pengetahuan. Amin.

Ponorogo, Agustus 2015

Penulis

DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan	ii
Halaman Berita Acara Ujian Skripsi	iii
Halaman Berita Acara Bimbingan Skripsi	iv
Halaman Motto	v
Halaman Persembahan	vi
Abstraksi	vii
Kata Pengantar	viii
Daftar Isi	x
Daftar Tabel	xiv
Daftar Gambar	xv

BAB I PENDAHULUAN

A. Latar belakang Masalah	1
B. Rumusan Masalah	2
C. Tujuan dan Manfaat	3
D. Batasan Masalah	3
E. Metode Pengambilan Data.....	4
F. Sistematika penulisan.....	4

BAB II DASAR TEORI

A. Pengertian <i>Data Mining</i>	7
1. Pengenalan Pola, <i>Data Mining</i> , dan <i>Machine learning</i>	8
2. Tahap-tahap <i>Data Mining</i>	10
a. Pembersihan <i>Data (Data Cleaning)</i>	10
b. Integrasi <i>Data (Data Integration)</i>	11
c. Seleksi <i>Data (Data Selection)</i>	11
d. Transformasi <i>Data (Data Transformation)</i>	11
e. Proses <i>Mining</i>	12
f. Evaluasi Pola (<i>Pattern evaluation</i>)	12

g. Presentasi Pengetahuan (<i>Knowledge Presentation</i>).....	12
3. Teknik <i>Data Mining</i>	12
4. <i>Predictive Modeling</i>	13
5. <i>Classification</i>	14
6. <i>Clustering</i>	15
7. <i>Link Analysis</i>	16
8. <i>Deviation Detection</i>	17
9. <i>Time Series Analysis</i>	17
10. <i>Time Series Analysis Pattern</i>	18
a. <i>Trends</i>	18
b. <i>Changes</i>	19
c. <i>Cycles</i>	19
d. <i>Seasonal</i>	19
e. <i>Outliers</i>	19
11. Algoritma <i>Microsoft Time Series</i>	20
a. <i>Autoregression</i>	20
b. <i>Autoregression Tree</i>	23
B. Pengertian <i>Data Warehouse</i>	24
1. Definisi <i>Data Warehouse</i>	24
a. <i>Subject-Oriented</i>	24
b. <i>Integrated</i>	25
c. <i>Time-variant</i>	25
d. <i>Non-volatile</i>	25
2. Keuntungan <i>Data Warehouse</i>	26
a. Potensi ROI (<i>Return On Investment</i>) yang besar.....	26
b. <i>Competitive Advantage</i>	26
c. Meningkatkan produktifitas dari pengambil keputusan perusahaan	27
3. <i>Basis Data</i> dan Sistem Manajemen <i>Basis Data</i>	27
a. DDL (<i>Data Definition Language</i>)	29
b. DML (<i>Data Manipulation Language</i>)	29
c. DCL (<i>Data Control Language</i>)	29
4. Kamus <i>Data (Data Dictionary)</i>	30
5. Desain Model Aplikasi	30
a. Diagram Konteks (<i>Context Diagram</i>).....	31

b.DFD (<i>Data Flow Diagram</i>)	31
--	----

BAB III ANALISIS DATA DAN MINING DATA

A. Analisis Data Mining	33
1. Hubungan Lama Studi Dengan IPK	33
2. Hubungan Tingkat Kelulusan Dengan Program Studi.....	33
B. Sumber Data	33
1. Kelulusan Mahasiswa	34
C. Data Yang Digunakan.....	36
D. Transformasi Data.....	37
E. Pemodelan Fungsi.....	38
1. DFD Level 0	38
2. DFD Level 1	39
a) <i>Import Data</i>	39
b) <i>Cleaning, Integrasi, Selection dan Transformasi</i>	39
1) <i>Cleaning Data</i>	39
2) <i>Integrasi Data</i>	40
3) <i>Selection Data</i>	40
4) <i>Transformasi Data</i>	40
c) Proses <i>Mining</i> lama Studi.....	40
d) Proses <i>Mining</i> IPK	40
e) Proses <i>Mining</i> Program Studi.....	40
f) Proses <i>Mining</i> Jenis Kelamin	41
F. Kebutuhan <i>Software</i> dan <i>Hardware</i>	41
1. <i>Software</i>	41
2. <i>Hardware</i>	41

BAB IV IMPLEMENTASI DATA

1. Implementasi Data	42
2. <i>Mining</i> Data	43
a. <i>Transformasi</i> Data	43
b. Pemodelan Data	45
1) <i>Penetapan</i> Jumlah Cluster(K).....	46
2) Hitung Jarak data kepusat Cluster.....	46

3) Pengelompokan data	50
4) Penghitungan Pusat Cluster baru	51
5) Pengujian Hasil	56
3. <i>Source Code</i>	58
 BAB V PENUTUP	
A. Kesimpulan	65
B. Saran	65
DAFTAR PUSTAKA	67

DAFTAR TABEL

Tabel 2.1 <i>Data mining operations and associated techniques</i>	13
Tabel 2.2 Simbol-Simbol Data Dictionary	30
Tabel 2.3 Simbol-Simbol Context Diagram	31
Tabel 2.4 Simbol-Simbol DFD	32
Tabel 3.1 Tabel Data Kelulusan.....	34
Tabel 3.3 Predikat Kelulusan.....	38
Tabel 4.1 Struktur tabel data.....	42
Tabel 4.2 koordinat centroid.....	48
Tabel 4.3 hasil Iterasi 1.....	49
Tabel 4.4 Penghitungan Cluster Baru C1	52
Tabel 4.5 Penghitungan Cluster Baru C2	53
Tabel 4.6 Cluster baru C3	54
Tabel 4.7 Hasil Iterasi 1	54
Tabel 4.8 Hasil Iterasi 2	55
Tabel 4.9 Hasil Iterasi 3	55
Tabel 4.10 Hasil Iterasi 4.....	55
Tabel 4.11 Hasil Iterasi 5	56
Tabel 4.12 Hasil iterasi 6	56
Tabel 4.13 Uji coba untuk mengetahui lama studi	56
Tabel 4.14 Hasil Clustering dengan Centroid baru.....	57

DAFTAR GAMBAR

Gambar 2.1 Data mining merupakan irisan dari berbagai disiplin.	9
Gambar 2.2 Tahap-Tahap Data Mining	10
Gambar 2.3 - Contoh classification menggunakan tree induction.....	14
Gambar 2.4 – Contoh classification menggunakan neural induction	15
Gambar 2.5 – Contoh database segmentation menggunakan scatterplot.....	16
Gambar 2.6 – Contoh visualisasi dari data pada Gambar 2.9	17
Gambar 2.7 – Time series plots	18
Gambar 2.8 – Microsoft Stock Value	20
Gambar 2.9 – Case Transformation	22
Gambar 2.10 – Matrix koefisien dimana r_d autocorrelation coefficient dengan delay d	23
Gambar 2.11 – Regression tree pada data time series	24
Gambar 3.1 DFD level 0.....	38
Gambar 3.2 DFD Level 1.....	39
Gambar 4.1 Data mentah	43
Gambar 4.2 Transformasi data.....	44
Gambar 4.3 Flowchart Clustering (K-Means)	45
Gambar 4.4 Randbetween.....	46
Gambar 4.4 koordinat centroid yang sudah dirandom.....	47