

**PERBANDINGAN 3 METODE DALAM DATA MINING
UNTUK PENJURUSAN SISWA DI SMA N 3 BOYOLALI**



SKRIPSI

Disusun sebagai salah satu syarat menyelesaikan Jenjang Strata I
pada Program Studi Informatika Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta

Oleh :

SYARIFAH NUR HARYATI

NIM : L200110067

**PROGRAM STUDI INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

2015

HAL NPERSETUJUAN

Skripsidenganjudw

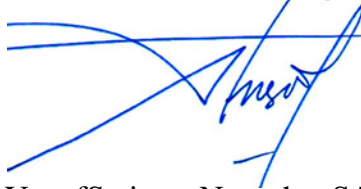
PERBANDINGAN 3 METODE DALAM DATA MINING
UNTUK PENJURUSAN SISWA DI SMA N 3 BOYOLALI

Ini telah diperiksa dan disetujui dalam sidang pendadaran :

Hari : Sabtu

Tanggal: 21 Februari 2015

Pembimbing



Yusuf Swis o Nu oho S.T.M. En

NIK: 1197

HALAMAN PENGESAHAN

PERBANDINGAN 3 :METODE DALAM DATA MINING
UNTUK PENJURUSAN SISWA DI SMAN 3 BOYOLALI

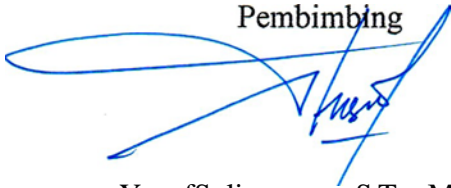
Dipersiapkan dan disusun oleh
SYARIFAH NUR HARYATI

NIM: L200110067

Telah dipertahankan di depan Dewan Penguji
pada tanggal 6 Maret 2015

Susunan Dewan Penguji

Pembimbing



Yusuf Sulis, S.T., M.En
NIK : 1197

Dewan Penguji I



Fatah Yasin, S.T., M.T.
NIK : 738

D T I

Hasyim Asy'ari, S.T., MT.
NIK: 981

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar sarjana

Tanggal .\ . -- — --



Dekan
Fakultas Komunikasi dan Informatika



Husni Thamrin, S.T, M.T, Ph.D
NIK : 706



Ketua
Program Studi Informatika



Dr. Heru Supriyono, M.Sc.
NIK : 970

DAFTAR KONTRIBUSI

Dengan ini saya menyatakan bahwa skripsi ini 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.

Berikut saya sampaikan daftar kontribusi dalam penyusunan skripsi:

1. Yusuf Sulisty Nugroho, ST., M. Eng, selaku dosen pembimbing yang memberikan masukan dan arahan penyusunan skripsi ini.
2. Saya sendiri yang mengolah data dan penghitungan metode *decision tree*, *naive bayes* dan *clustering k-means* berdasarkan sumber-sumber pada buku, literatur dan *internet*.
3. Aplikasi yang saya gunakan untuk membantu skripsi ini yaitu.
 - a. *Rapid Miner 5*.
 - b. *Microsoft Excel 2010*.

Demikian pernyataan dan daftar kontribusi ini saya buat dengan sejujurnya.

Saya bertanggungjawab atas isi dan kebenaran daftar diatas.

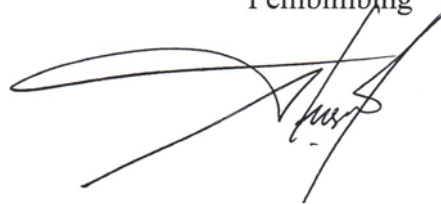
Surakarta, Februari 2015



SYARIFAH NUR HARYATI

Mengetahui :

Pembimbing



Yusuf Sulistyogroho, S.T., M. Eng
NIK: 1197

MOTTO DAN PERSEMBAHAN

MOTTO :

“Sukses adalah berani bertindak dan punya prinsip”

-Penulis-

“Keberhasilan adalah sebuah proses. Niat adalah awal keberhasilan. Peluh keringat adalah penyedapnya. Tetesan air mata adalah pewarnanya. Doa dan orang-orang disekitar adalah bara api yang mematangkannya. Kegagalan disetiap langkah adalah pengawetnya. Maka dari itu bersabarlah ! Allah selalu menyertai orang-orang yang penuh kesabaran dalam proses menuju keberhasilan.”

-Penulis-

“Kesulitan tidak akan datang apabila kita mau berusaha.”

-Penulis-

“Musuh yang paling berbahaya di atas dunia ini adalah penakut dan bimbang. Teman yang paling setia, hanyalah keberanian dan keyakinan yang teguh.”

-Andrew Jackson-

“Pelangi akan hadir Selepas Hujan Lebat”

-Penulis-

PERSEMBAHAN :

Dengan mengucapkan syukur alhamdulillah atas rahmad dan hidayah Allah SWT, akan kupersembahkan karya ini kepada orang-orang yang saya sayangi :

1. Kedua orang tuaku tercinta, untuk kasih sayang yang tidak terbatas, yang selalu senantiasa mendoakanku dan memberi dukungan moril maupun materil yang tak akan pernah terbalaskan demi kesuksesanku.
2. Kakakku Taufik Ismail dan saudara-saudaraku yang telah memberikan perhatian, dukungan dalam menyelesaikan skripsi.
3. Sahabatku Veronica Andriana, yang selalu ada disaat suka maupun duka, berjuang bersama dari awal kuliah sampai berakhirnya kuliah, saling memberikan motivasi untuk lulus bersama.
4. Orang-orang terdekatku Dhody, Irma, Ajik, Dhony, Loga, Nanda, Mira, Nyla yang selalu memberikan semangat, motivasi dan menghiburku disaat aku malas dan jenuh dalam menyelesaikan skripsi .
5. Pasukan Damin (Veronica dan Deny) sebagai partner yang telah berjuang bersama, saling membantu dan memberikan semangat disaat suka maupun duka dari awal sampai terselesaikannya skripsi ini.
6. Teman-teman Teknik Informatika angkatan 2011 yang telah menemaniku selama ini berjuang bersama berbagi ilmu dan pengalamannya.
7. Kakak tingkat angkatan 2010 Mbak Tika, Mbak Nina, Mas Ibnu, Mas Adit, Mas Aziz, Mas Adjie yang selalu memberikan bantuan, semangat dan bimbingan disaat mengalami kesulitan dalam menyelesaikan skripsi.

KATA PENGANTAR

Alhamdulillah, kami panjatkan syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Perbandingan 3 metode dalam data mining untuk penjurusan siswa di SMA N 3 Boyolali”.

Skripsi ini disusun untuk memenuhi kurikulum pada Program Studi Teknik Informatika Universitas Muhammadiyah Surakarta, sebagai kewajiban mahasiswa dalam rangka menyelesaikan program sarjana.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, oleh karena itu kritik saran yang membangun dari berbagai pihak sangat penulis harapkan demi perbaikan-perbaikan ke depan.

Terwujudnya skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Husni Thamrin, S.T, MT., Ph.D. selaku Dekan Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta serta pembimbing akademik selama kuliah.
2. Bapak Dr. Heru Supriyono, S.T., M.Sc. selaku Ketua Jurusan Teknik Informatika Universitas Muhammadiyah Surakarta.
3. Bapak Yusuf Sulistyono, S.T., M.Eng selaku pembimbing skripsi yang telah memberikan, bimbingan, dan pengarahan kepada penulis sehingga dapat menyelesaikan tugas akhir ini.

4. Segenap dosen dan karyawan prodi Teknik Informatika atas bantuan dan ilmu yang diberikan kepada penulis selama masa perkuliahan hingga dinyatakan mendapat gelar Strata 1.
5. Kepada orang tua yang selalu memberikan do'a, semangat dan motivasi dengan tiada hentinya kepada penulis.
6. Semua pihak yang tidak bisa disebutkan satu-persatu yang telah membantu hingga terselesainya skripsi ini.

Akhirnya penulis berharap semoga skripsi ini berguna bagi semua pihak dan bermanfaat bagi penulis khususnya dan pembaca pada umumnya dalam menambah pengetahuan dan wawasan ilmu. Amiin.

Surakarta, Februari 2015

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
DAFTAR KONTRIBUSI.....	iv
MOTTO DAN PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR PERSAMAAN	xiv
DAFTAR LAMPIRAN.....	xv
DAFTAR GAMBAR	xvi
ABSTRAKSI.....	xviii
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah	3
C. Batasan Masalah	3
D. Tujuan Penelitian.....	3
E. Manfaat Penelitian.....	3
F. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	6
A. Telaah Penelitian	6

B. Landasan Teori	11
1. <i>Data mining</i>	11
2. <i>Decision Tree</i>	11
3. Algoritma C4.5.....	12
4. <i>Naive Bayes</i>	13
5. <i>Clustering</i>	14
6. Algoritma <i>K-Means</i>	14
7. Atribut dominan.....	15
8. <i>RapidMiner 5</i>	15
9. <i>Accuracy</i>	16
10. <i>Precision</i>	16
11. <i>Recall</i>	16
BAB III METODOLOGI PENELITIAN	17
A. Waktu dan Tempat	17
B. Peralatan Utama dan Pendukung	17
C. Tahap Penelitian	18
1. Identifikasi Masalah	18
2. Metode Pengumpulan Data	19
a. Data Primer	19
b. Data Sekunder.....	19
c. Studi pustaka.....	19
3. Analisis Kebutuhan	20

a. Penentuan Atribut 20

b. Penentuan Sampel.....	22
4. Pengelompokkan Data.....	22
5. Implementasi <i>data mining</i>	24
a. Penggunaan metode <i>Decision Tree</i>	24
b. Penggunaan metode <i>Naive Bayes</i>	25
c. Penggunaan metode <i>K-Means Clustering</i>	26
BAB IV HASIL DAN PEMBAHASAN	28
A. Hasil Penelitian	28
1. Penentuan Atribut.....	28
2. Penentuan Sampel	29
3. Pengolahan Data.....	29
4. Hasil Implementasi <i>Decision Tree</i> menggunakan <i>RapidMiner 5</i>	30
5. Hasil Implementasi <i>Naive Bayes</i> menggunakan <i>RapidMiner 5</i>	35
6. Hasil Implementasi <i>Clustering(K-Means)</i> Menggunakan <i>RapidMiner 5</i>	39
B. Analisa dan Pembahasan.....	42
1. Implementasi dengan Perhitungan <i>Decision Tree</i>	42
a. Perhitungan untuk mencari <i>root node</i> /simpul akar.....	43
b. Perhitungan untuk <i>internal node</i> /simpul cabang.....	50
c. Perhitungan untuk <i>internal node</i> selanjutnya.....	56
2. Implementasi dengan Perhitungan <i>Naive Bayes</i>	68

3. Implementasi dengan Perhitungan <i>Clustering (K-Means)</i> ..	72
C. Hasil Perbandingan 3 Metode.....	80
D. Interpretasi Hasil Penelitian.....	80
BAB V KESIMPULAN DAN SARAN.....	83
A. Kesimpulan.....	83
B. Saran	84
DAFTAR PUSTAKA	85
LAMPIRAN.....	87

DAFTAR TABEL

Tabel 3.1	Daftar Atribut keseluruhan.....	20
Tabel 3.2	Penentuan Variabel <i>Dependen</i> dan <i>Independen</i>	21
Tabel 3.3	Pengelompokan data	23
Tabel 3.4	Daftar wilayah.....	23
Tabel 4.1	Data <i>Training</i>	30
Tabel 4.2	Data penjurusan siswa.....	73
Tabel 4.3	Hasil <i>centroid</i> per <i>cluster</i>	75
Tabel 4.4	Hasil Perbandingan 3 Metode	80

DAFTAR PERSAMAAN

Persamaan (3.1)	Rumus Slovin	22
Persamaan (3.2)	Rumus <i>Entropy</i>	25
Persamaan (3.3)	Rumus <i>Information Gain</i>	25
Persamaan (3.4)	Rumus <i>Naive Bayes</i>	26
Persamaan (3.5)	Rumus jarak <i>euclidean</i>	27
Persamaan (3.6)	Rumus <i>Centroid</i>	27

DAFTAR LAMPIRAN

Lampiran 1. Surat Keterangan Penelitian

Lampiran 2. Data Training penjurusan siswa

Lampiran 3. Data Testing penjurusan siswa

Lampiran 4. Tampilan *Decision Tree* pada aplikasi *RapidMiner 5*

Lampiran 5. Tampilan *Naive Bayes* pada aplikasi *RapidMiner 5*

Lampiran 6. Tampilan *Clustering K-Means* pada aplikasi *RapidMiner 5*

DAFTAR GAMBAR

Gambar 4.1	Rancangan proses penerapan <i>Decision Tree</i> untuk klasifikasi data penjurusan siswa	31
Gambar 4.2	Hasil skema <i>Decision Tree</i> untuk klasifikasi data penjurusan siswa	32
Gambar 4.3	Tampilan hasil <i>decision tree</i> pada <i>Scatter Plot</i>	34
Gambar 4.4	Rancangan proses <i>Naive Bayes</i> menggunakan <i>data training</i> ...	36
Gambar 4.5	Rancangan proses <i>Naive Bayes</i> menggunakan <i>data testing</i>	36
Gambar 4.6	Hasil <i>Naive Bayes</i> untuk prediksi penjurusan siswa.....	37
Gambar 4.7	Tampilan hasil <i>Naive Bayes</i> pada <i>Text View</i>	37
Gambar 4.8	<i>Scatter Plot Naive Bayes</i> menggunakan <i>data training</i>	38
Gambar 4.9	<i>Scatter Plot Naive Bayes</i> menggunakan <i>data testing</i>	39
Gambar 4.10	Rancangan proses penerapan metode <i>K-Means</i> untuk pengelompokan penjurusan siswa.....	40
Gambar 4.11	Hasil <i>Cluster Model K-Means</i> untuk penjurusan siswa	40
Gambar 4.12	Hasil pengelompokan <i>Cluster</i> data penjurusan siswa.....	41
Gambar 4.13	Tampilan hasil <i>Clustering K-Means</i> pada <i>Scatter Plot</i>	41
Gambar 4.14	<i>Root Node</i>	50
Gambar 4.15	<i>Internal node</i> rata-rata IPS	56
Gambar 4.16	<i>Leaf node</i> rata-rata IPS nilai ≤ 70	57
Gambar 4.17	<i>Internal Node</i> ke-2	62
Gambar 4.18	<i>Leaf node</i> rata-rata IPS nilai > 80	62

Gambar 4.19	<i>Leaf node gender</i> laki-laki	63
Gambar 4.20	<i>Leaf node gender</i> perempuan	68

ABSTRAKSI

SMA N 3 Boyolali merupakan salah satu Sekolah Menengah Atas di kota Boyolali yang terdapat 2 jurusan yaitu IPA dan IPS. Penjurusan siswa ini mengarahkan peserta didik agar lebih fokus mengembangkan kemampuan dan minat yang dimiliki. Jurusan yang tidak tepat bisa sangat merugikan siswa dan masa depannya. Dengan penjurusan tersebut diharapkan dapat memaksimalkan potensi, bakat atau talenta individu, sehingga juga akan memaksimalkan nilai akademisnya. Untuk mengatasi permasalahan tersebut adalah dengan cara menerapkan proses *data mining*.

Teknik *data mining* yang digunakan dalam penentuan jurusan ini menggunakan metode *Decision Tree* Algoritma C4.5, *Naive Bayes* dan *Clustering* Algoritma *K-Means*. Sedangkan atribut yang digunakan terdiri dari *Gender*, *Minat*, *Rata-rata IPA*, *Rata-rata IPS*, *Psikotest IPA*, *Psikotest IPS*, *Asal Sekolah* dan *Jurusan*. Dalam melakukan analisa ini menggunakan bantuan *software RapidMiner 5* untuk mengetahui metode apa yang paling baik.

Pengimplementasian *data mining* menggunakan perbandingan 3 metode dapat diketahui bahwa berdasarkan dari nilai *precision*, metode *naive bayes* lebih baik digunakan untuk penelitian ini dibandingkan dengan metode yang lain dengan nilai 77,51%. Sedangkan berdasarkan nilai *recall* dan *accuracy*, *decision tree* lebih baik digunakan dibanding metode yang lain dengan nilai *recall* sebesar 90,80% dan nilai *accuracy* sebesar 79,14%. Variabel yang paling berpengaruh dalam menentukan penjurusan yaitu rata-rata IPA sehingga perlu dijadikan pertimbangan bagi pihak sekolah.

Kata kunci : *Data Mining, Decision Tree Algoritma C.45, Naive Bayes, Clustering Algoritma K-Means*