
IMPLEMENTASI METODE *CLUSTERING* PADA DATA PENERIMAAN SISWA BARU MENGUNAKAN *ALGORITMA K – MEAN* (Studi Kasus SMK Darul Falah Kertagena Tengah)

Irwan Darmawan¹⁾, sholeh rachmatullah²⁾, Sandy Vikki A³⁾, Nirwana Haidar H⁴⁾, Anang Faktchur R⁵⁾

Infomatika Universitas Madura

Email: darmawan@unira.ac.id, sholeh@unira.ac.id, sandy@unira.ac.id, haidar@unira.ac.id,
anang@unira.ac.id

ABSTRAK

Dengan semakin berkembangnya Teknologi Informasi dan Komunikasi dewasa ini menyebabkan tuntutan baru dalam dunia pendidikan akan penguasaan dalam bidang Teknologi. Kebutuhan Informasi dari pelaku pendidikan dan masyarakat juga terus meningkat dalam segi kualitas dan kuantitas, maka penguasaan pengetahuan tentang Teknologi Informasi sangat diperlukan.

Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung didalam limpahan data tersebut, yang kemudian melahirkan data mining. Dunia pendidikan memiliki data yang melimpah dan berkesinambungan mengenai siswa yang dibina dan alumni yang dihasilkan hal ini membuka peluang diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik.

Sebagai contoh, tentang Penerimaan Siswa Baru di Sekolah yang hanya disimpan dalam buku sebagai arsip di sekolah dan formulir pendaftaran sebagai tanda bukti untuk siswa yang mendaftar, daftar calon siswa yang lulus ujian masuk sekolah berdasarkan jumlah kelas, daftar nomor induk bagi yang diterima sebagai siswa baru, sehingga banyak waktu yang terbuang untuk mencatat dan mencari data pendaftar dan seringkali nilai hasil ujian penerimaan siswa baru hanya disimpan dan tidak di analisa setiap tahunnya. Dengan demikian, hal ini menjadi dasar untuk pembuatan Tugas Akhir berjudul Implementasi Metode *Clustering* Pada Data Penerimaan Siswa Baru Menggunakan *Algoritma K - Mean* (Studi Kasus SMK Darul Falah Kertagena Tengah).

Kata kunci : *Algoritma K – Mean, Clustering, Data mining.*

I. PENDAHULUAN

1.1 Latar Belakang

Dengan semakin berkembangnya Teknologi Informasi dan Komunikasi dewasa ini menyebabkan tuntutan baru dalam dunia pendidikan akan penguasaan dalam bidang Teknologi. Kebutuhan Informasi dari pelaku pendidikan dan masyarakat juga terus meningkat dalam segi kualitas dan kuantitas, maka penguasaan pengetahuan tentang Teknologi Informasi sangat diperlukan.

Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan menimbulkan

kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung didalam limpahan data tersebut, yang kemudian melahirkan *data mining*. Dunia pendidikan memiliki data yang melimpah dan berkesinambungan mengenai siswa yang dibina dan alumni yang dihasilkan hal ini membuka peluang diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik.

Sebagai contoh, tentang Penerimaan Siswa Baru di Sekolah yang hanya disimpan dalam buku sebagai arsip di sekolah dan formulir pendaftaran sebagai tanda bukti untuk siswa yang mendaftar, daftar calon siswa yang lulus ujian masuk sekolah berdasarkan jumlah kelas, daftar nomor induk bagi yang diterima sebagai siswa

baru, sehingga banyak waktu yang terbuang untuk mencatat dan mencari data pendaftar dan seringkali nilai hasil ujian penerimaan siswa baru hanya disimpan dan tidak di analisa setiap tahunnya. Dengan demikian, hal ini menjadi dasar untuk pembuatan Tugas Akhir berjudul Implementasi Metode *Clustering* Pada Data Penerimaan Siswa Baru Menggunakan Algoritma *K - Mean* (Studi Kasus SMK Darul Falah Kertagena Tengah).

1.2 Perumusan Masalah

Permasalahan yang diangkat dalam penyelesaian tugas akhir ini adalah sebagai berikut :

1. Bagaimana mengefisienkan waktu analisa data Penerimaan Siswa Baru di SMK Darul Falah Kertagena Tengah yang selalu memakan waktu yang cukup lama ?
2. Bagaimana penerapan metode *clustering* dengan algoritma K-Means terhadap data hasil ujian penerimaan siswa baru?
3. Bagaimana menyajikan informasi daftar calon siswa baru dengan pola-pola data menggunakan data mining ?

1.3 Batasan Masalah

Batasan masalah pada sistem pakar ini adalah:

1. Data yang di *cluster* adalah seluruh hasil ujian
2. Data standart yang dipakai sebagai data uji dalam *data mining* dari 2 tahun terakhir.
3. Sistem melakukan pengelompokan data berdasarkan nilai ujian, namun tidak menghitung pembobotan nilai, tetapi berdasarkan nilai terbesar hingga terkecil, apabila terjadi nilai yang sama pada batasan jumlah akhir kelulusan maka siswa yang di luluskan yang memiliki nomer pendaftaran paling kecil.
4. Pendukung yang digunakan dalam membuat web Penerimaan Siswa baru *Xampp, php, dreamweaver dan photoshop*.

1.4 Tujuan

Tujuan yang ingin dicapai dari pelaksanaan dan penulisan tugas akhir ini sebagai berikut :

1. Merancang dan membuat aplikasi sistem informasi penerimaan siswa baru berbasis web di SMK Darul Falah Kertagena Tengah.

2. Menghasilkan aplikasi sistem informasi penerimaan jumlah calon siswa dan nilai hasil ujian menggunakan metode *clustering* dari tahun ke tahun.

1.5 Manfaat Penelitian

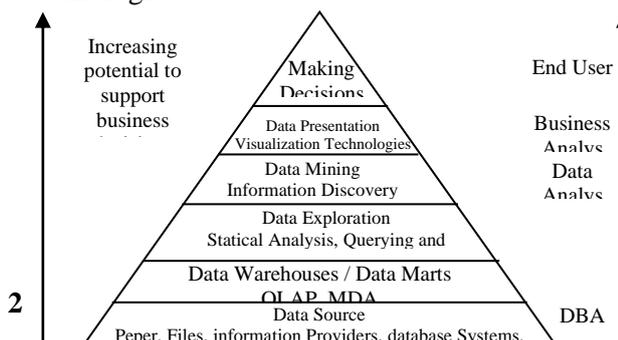
1. Mengetahui cara menerapkan ilmu-ilmu yang telah dipelajari selama ini, dalam merancang dan membuat aplikasi sistem dengan teknik data mining.
2. Membantu proses penerimaan siswa baru
3. Membantu dalam pengambilan keputusanr dari hasil pengelompokan nilai hasil ujian.

II. TEORI DASAR

1. Data Mining

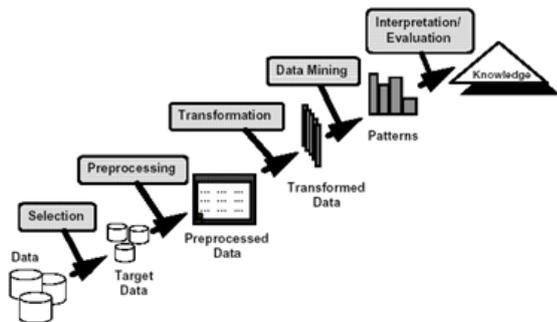
Penggalian data (*data mining*) adalah ekstraksi pola yang menarik dari data dalam jumlah besar. Suatu pola dikatakan menarik apabila pola tersebut tidak sepele, implisit, tidak diketahui sebelumnya, dan berguna. Pola yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru. Penggalian data memiliki beberapa nama alternatif, meskipun definisi eksaknya berbeda, seperti KDD (*knowledge discovery in database*), analisis pola, arkeologi data, pemanenan informasi, dan intelegensia bisnis. Penggalian data diperlukan saat data yang tersedia terlalu banyak (misalnya data yang diperoleh dari sistem basis data perusahaan, e-commerce, data saham, dan data bioinformatika), tapi tidak tahu pola apa yang bisa didapatkan.

Perbedaan *data mining* dengan data *warehouse* dan OLAP (*Online Analytical Processing*) secara singkat bisa dijawab bahwa teknologi yang ada di data *warehouse* dan OLAP (*Online Analytical Processing*) dimanfaatkan penuh untuk melakukan *data mining*. Gambar di bawah menunjukkan posisi masing-masing teknologi.



Gambar 2.1: *Data Mining* dan Teknologi Database Lainnya

Tahapan *data mining* dapat dilihat pada Gambar 2.2.



Gambar 2.2: Tahapan *Data Mining*

Tahapan – tahapan proses data mining :

a. Data Selection

- (1) Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan.
- (2) 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.

b. Pre-processing/ Cleaning

- (1) Pemrosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan noise dilakukan.
- (2) Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD
- (3) Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).
- (4) 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.

c. Transformation

- (1) Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.

(2) Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d. Data mining

(1) Pemilihan tugas data mining : pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dan lain-lain.

(2) Pemilihan algoritma data mining untuk pencarian (searching)

(3) 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.

d. Interpretation/ Evaluation

(1) Penerjemahan pola-pola yang dihasilkan dari data mining.

(2) Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan

(3) Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Penggalian data adalah salah satu bagian dari proses pencarian pola. Berikut ini urutan proses pencarian pola:

- a. Pembersihan Data: yaitu menghapus data pengganggu (*noise*) dan mengisi data yang hilang.
- b. Integrasi Data: yaitu menggabungkan berbagai sumber data
- c. Pemilihan Data: yaitu memilih data yang relevan
- d. Transformasi Data: yaitu mentransformasi data ke dalam format untuk diproses dalam penggalian data.
- e. Penggalian Data: yaitu menerapkan metode cerdas untuk ekstraksi pola.
- f. Evaluasi pola: yaitu mengenali pola-pola yang menarik saja
- g. Penyajian pola: yaitu memvisualisasi pola ke pengguna

Teknik pengumpulan data dan teknologi penyimpanan di berbagai bidang yang berkembang pesat menyebabkan basis data terlalu besar. Analisis data tanpa menggunakan otomasi dari penggalian data adalah tidak memungkinkan lagi, kalau data terlalu banyak, dimensionalitas data terlalu besar, data terlalu kompleks untuk dianalisis manual. Sehingga, lahirlah cabang ilmu penggalian data ini.

Pada dasarnya penggalian data dibedakan menjadi dua fungsionalitas, yaitu deskripsi dan prediksi. Berikut ini beberapa fungsionalitas penggalian data yang sering digunakan:

- a. Karakterisasi dan Diskriminasi: yaitu menggeneralisasi, merangkum, dan mengkontraskan karakteristik data.
- b. Penggalian pola berulang: yaitu pencarian pola asosiasi (association rule) atau pola intra-transaksi, atau pola pembelian yang terjadi dalam satu kali transaksi.
- c. Klasifikasi: yaitu membangun suatu model yang bisa mengklasifikasikan suatu objek berdasar atribut-atributnya. Kelas target sudah tersedia dalam data sebelumnya, sehingga fokusnya adalah bagaimana mempelajari data yang ada agar klasifikator bisa mengklasifikasikan sendiri.
- d. Prediksi: yaitu memprediksi nilai yang tidak diketahui atau nilai yang hilang, menggunakan model dari klasifikasi.
- e. Penggugusan/Cluster analysis: yaitu mengelompokkan sekumpulan objek data berdasarkan kemiripannya. Kelas target tidak tersedia dalam data sebelumnya, sehingga fokusnya adalah memaksimalkan kemiripan intrakelas dan meminimalkan kemiripan antarkelas.
- f. Analisis outlier: yaitu proses pengenalan data yang tidak sesuai dengan perilaku umum dari data lainnya. Contoh: mengenali noise dan pengecualian dalam data.
- g. Analisis trend dan evolusi: meliputi analisis regresi, penggalian pola sekuensial, analisis periodisitas, dan analisis berbasis kemiripan.

Berikut ini adalah 9 algoritma penggalian data yang paling populer berdasarkan konferensi ICDM '06 (International Conference On Data Mining 2006).

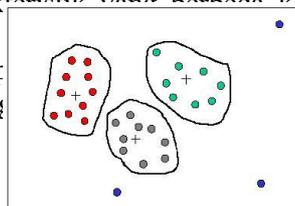
Tabel 2.1 algoritma penggalian data yang paling populer berdasarkan konferensi ICDM '06

No	Algoritma	Suara	No	Algoritma	Suara
1	C4.5	61	6	PageRank	46
2	k-Means	60	7	AdaBoost	45
3	SVM	58	8	KNN	45
4	Apriori	52	9	NaiveBayes	34
5	EM	48			

2. Metode Clustering

Clustering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode Data Mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain.

Setiap data yang semirip mungkin akan digolongkan ke dalam satu wilayah yang sama. Gambar di bawah ini.



Gambar 2.3 Clustering

Ada beberapa pendekatan yang digunakan dalam mengembangkan metode clustering. Dua pendekatan utama adalah clustering dengan pendekatan partisi dan clustering dengan pendekatan hirarki. Clustering dengan pendekatan partisi atau sering disebut dengan partition-based clustering mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam cluster-cluster yang ada. Clustering dengan pendekatan hirarki atau sering disebut dengan hierarchical clustering mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Di samping kedua pendekatan tersebut, ada juga clustering dengan pendekatan automatic mapping (Self-Organising Map/SOM).

3. K-Means

Salah satu metode yang banyak digunakan dalam melakukan clustering dengan partisi ini adalah metode k-means. Secara umum metode k-means ini melakukan proses pengelompokan dengan prosedur sebagai berikut:

- a. Tentukan jumlah cluster
- b. Tentukan pusat clusternya

- c. Hitung rata-rata setiap cluster dari data yang tergabung di dalamnya
- d. Alokasikan kembali semua data ke cluster terdekat
- e. Ulang proses nomor 2, apabila ada data yang berpindah cluster atau apabila ada perubahan nilai centroid.

Prosedur dasar ini bisa berubah mengikuti pendekatan pengalokasian data yang diterapkan, apakah crisp atau fuzzy. Setelah meneliti clustering dari sudut yang lain, saya menemukan bahwa k-means clustering mempunyai beberapa kelemahan.

Pada algoritma 2.1 data clustering menggunakan metode K-Means, secara umum dilakukan dengan algoritma dasar Algoritma 2.1

Input:
 $D = (t_1, t_2, \dots, t_n)$ // Set of elements
 K // Number of desired clusters
 Output
 K // Set of clusters
 K-means algorithm :
 Assign initial values for means m_1, m_2, \dots, m_k ;
 repeat
 assign each item t_i to the cluster which has the closest mean;
 calculate new mean for each cluster;
 until convergence criteria is met;

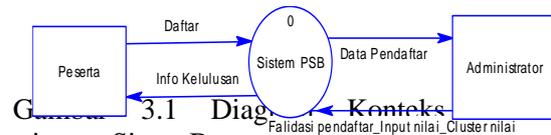
Metode ini mempartisi data ke dalam cluster, sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

III. RANCANGAN SISTEM

1 Data flow diagram (DFD)

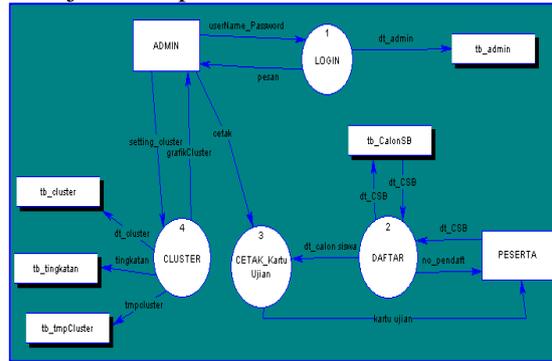
Data flow diagram (DFD) digunakan untuk menggambarkan aliran informasi dan proses data yang bergerak dari pemasukan data hingga keluaran. Data flow diagram memudahkan pemakai yang kurang menguasai bidang komputer untuk mengerti sistem yang akan dikerjakan atau dikembangkan. Data flow diagram dari penerapan sistem informasi Penerimaan Siswa baru dapat dilihat pada gambar

3.1 menjelaskan diagram konteks. Sedangkan diagram level 1 dapat dilihat pada gambar 3.2



Gambar 3.1 Diagram Konteks Penerimaan Siswa Baru

Pengenmbangan proses yang terjadi pada diagram konteks di atas dapat dijabarkan pada DFD level 1 berikut:



Gambar 3.2 DFD Level 1 Sistem Penerimaan Siswa Baru

Pada gambar di atas ada empat proses utama yang dilakukan Sistem PSB , yaitu :

1. Login Admin melakukan proses login
2. Proses Pendaftaran Disini peserta bisa melakukan proses pengisian data pribadi untuk mendaftar.
3. Cetak kartu ujian Peserta yang sudah daftar ulang mendapat kartu ujian.
4. Cluster Nilai Hasil pengolahan data nilai ke dalam bentuk cluster.

VI. IMPLEMENTASI

4.3 Form Informasi Pendaftaran

Melalui menu informasi pendaftaran, calon siswa bisa mendapatkan informasi tentang proses dan syarat-syarat unuk mengikuti seleksi penerimaan siswa baru, seperti syarat pendaftaran, syarat test.



Gambar 4.3 *Form* Informasi Pendaftaran
4.4 *Form* Arsip Visi misi

Berisikan informasi tentang visi dan misi sekolah, dan beberapa penunjang lainnya.

Gambar 4.4 *Form* Arsip Visi dan Misi

4.5 *Form* Daftar Baru

Dengan mengarahkan klik *link* ini, maka calon siswa akan melakukan pendaftaran sebagai peserta seleksi penerimaan siswa baru. Dengan cara mengisi *form* yang telah disediakan.

Gambar 4.5 *Form* Daftar Baru

4.6 *Form* Cetak Nomer Pendaftaran

Setelah mengisi *form* pendaftaran dan mengklik tombol daftar PSB, maka akan tampil nomer pendaftaran di bawah ini, kemudian calon siswa menekan tombol *print* untuk mencetak nomer pendaftaran.

Gambar 4.6 *Form* Cetak Nomer Pendaftaran

4.8 Menu admin

Dari menu login, kita akan masuk ke *form* admin. Kita harus memasukkan *username* dan

password admin, jika *username* dan *password*nya salah maka kita diminta login kembali. Jika benar maka kita akan masuk ke halaman administrator, yang terdiri dari menu :

- a. *Data User* : admin dapat membuat userbaru dan merubah *password* admin.
- b. *Daftar Calon siswa* : berisikan seluruh informasi calon siswa yang telah melakukan pendaftaran.
- c. *Nilai* : merupakan informasi yang menyimpan daftar nilai calon siswa.
- d. *Kelulusan* : berisikan informasi daftar calon siswa yang lulus beserta nilai ujiannya.
- e. *Cetak laporan* : Dapat melakukan cetak lapoaran seperti data pendaftar test masuk, dan daftar calon siswa yang lulus.
- f. *Atur Kelulusan* berisikan pengaturan jumlah kelulusan yang akan diterima dalam PSB.
- g. *Setting Cluster* : berisikan kontrol untuk menghitung nilai cluster yang ada. Dimana *cluster* tersebut dibagi kedalam 5 *cluster*.
- h. *Grafik K-means* : merupakan penampilan hasil *cluster* kedalam bentuk grafik.

Gambar 4.9 *Form* Administrator

4.9 *Form* Daftar Nama Calon Siswa Baru

Siswa yang sudah mengisi *form* pendaftaran akan ditampilkan dalam halaman calon siswa baru, seperti gambar dibawah ini :

Gambar 4.10 *Form* Daftar Calon Siswa Baru

4.10 *Form* Cek Registrasi

Siswa sebelum mendapat kartu ujian, terlebih dahulu dilakukan validasi oleh admin kepada calon siswa seperti gambar di bawah ini :



4.11 Form Cetak Kartu Ujian

Setelah siswa melakukan validasi, maka admin mencetak kartu ujian siswa, seperti gambar dibawah ini :



Gambar 4.12 Form Cetak Kartu Ujian

V. KESIMPULAN

5.1 Kesimpulan

Berdasarkan pembahasan dan analisa program pada bab-bab sebelumnya maka dapat disimpulkan hal-hal sebagai berikut:

- 1) Dengan desain web, peginputan, pengolahan dan manajemen data akan lebih teratur sehingga Program ini mempermudah dan mempercepat kegiatan yang terdapat dalam penerimaan siswa baru
- 2) Implementasi Metode clustering cukup baik untuk mereduksi data dengan pola-pola data kelulusan SMK Darul Falah.
- 3) Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi

tingkat kelulusan khususnya faktor dalam data induk siswa.

5.2 Saran

Untuk menunjang kegiatan pembuatan rekapitulasi laporan maka sangatlah diperlukan :

1. Program yang bisa menentukan kelulusan siswa, jika terjadi nilai yang sama calon siswa yang dinyatakan lulus berdasarkan jumlah kuota yang sudah ditentukan.
2. Diharapkan aplikasi juga dapat dikembangkan untuk clustering yang tidak hanya berdasarkan nilai hasil ujian, namun juga dapat berdasarkan NEM, nilai raport dan sebagainya

REFERENSI

Firdaus. 2007. "7 jam belajar interaktif php dan Msql dengan dreamweaver". Palembang: Maxikom.

Hakim Lukman. 2004. "Cara mudah memadukan web design dan web programing". Jakarta: PT. Elex Media Komputindo.

Jurnal pendidikan dan pengajaran UNDIKSHA, No.3 TH.XXXVI juli 2008

Kusrini.2009." Algoritma data mining".Yogyakarta:ANDI.

Kadir Abdul. 1999. " Konsep dan tuntunan praktis basis data ". Yogyakarta: ANDI.

Kadir Abdul. 2003. "Pemrograman web mencakup : html, css java script dan php". Yogyakarta: ANDI.

Susanto sani,Ph.D , Surady,S.T,M.S.2010. " Pengantar data mining menggali pengetahuan dari bongkahan data ". Yogyakarta : ANDI