

PENERAPAN DATA MINING DENGAN MENGGUNAKAN ALGORITMA C4.5 UNTUK MENGLASIFIKASIKAN PEREKRUTAN KARYAWAN DI SMK NEGERI 1 NGULING

Endang Widia Astuti ¹⁾
SMK NEGERI 1 NGULING

Email : sttastuti@gmail.com

ABSTRACT : In this study, the authors sought to mine the data an Educational Institution to determine the classification of the recruitment of employees at the school. Existing data were analyzed using a decision tree approach (decision tree) ie the algorithm C4.5. With this algorithm can know the data of candidates which are grouped into classes that qualify prospective employees and data which are grouped into classes do not qualify. The pattern then can be used to determine which job candidates who meet the criteria to be included in the agency, so that the principal can take a decision to accept or reject the prospective employee.

Keyword: *algorithm C4.5, the classification of employee recruitment, Data Mining*

1. PENDAHULUAN

Aktifitas Kerja dalam suatu Instansi SMKN 1 NGULING digerakan oleh tenaga kerja yang memiliki Kompetensi dan ketrampilan di bidangnya. Sumberdaya manusia dalam hal ini tenaga kerja menjadi syarat utama dalam mengoprasikan management sekolah. Pengolahan sumberdaya manusia yang tepat, menjadi bagian yang sangat penting dan bahkan proses prerkrutan tenga kerja yang tidak tepat akan menjadi masalah tersendiri bagi Instansi tersebut. Setiap sekolah berupaya untuk menyusun format yang tepat tentang manajemen sumberdaya manusianya (mulai dari proses perekrutan, pendidikan, dan pelatihan, job description yang jelas, sistem upah atau gaji yang tepat, adanya jenjang karir atau pengembangan staf, dan lainnya).

Saat ini persaingan untuk mendapatkan pekerjaan semakin ketat, dimana lapangan kerja yang tersedia semakin sedikit dan proses perekrutan dan seleksi karyawan semakin ketat, sehingga kesempatan orang untuk memperoleh pekerjaan semakin sulit. Kemudian dapat dikatakan lebih lanjut bahwa tujuan utama dari proses rekrutmen dan seleksi pegawai adalah untuk mendapatkan orang yang tepat bagi suatu jabatan tertentu, sehingga orang tersebut mampu bekerja secara optimal dan dapat bertahan di Instansi dalam jangka waktu yang lama. Meskipun tujuannya terdengar sangat sederhana, proses rekrutmen tersebut sangatlah kompleks, memakan waktu

yang lama dan sangat terbuka peluang untuk melakukan kesalahan dalam menentukan orang yang tepat.

2. Metode

2.1. Definisi Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine elarning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [Turban 2005]. Menurut gartner group, data mining adalah suatuV proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dan sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistic dan matematika [Larose 2005].

2.2. Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variable target. Estimasi lebih kearah numeric dari pada kearah kategori.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa datang.

4. Klasifikasi

Dalam kalsifikasi terdapat target variable kategori, sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengkulusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan record-record dalam cluster.

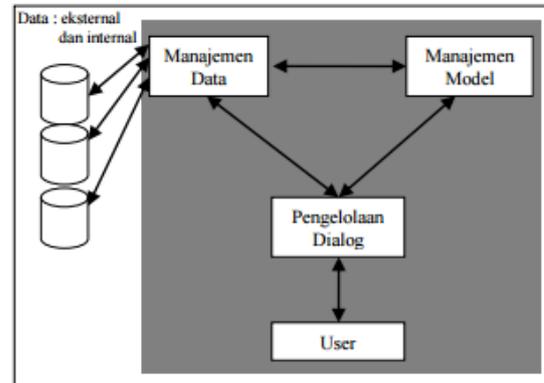
6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang biasa.

2.3. Tahapan Data Mining

Tahapan dalam data mining sesungguhnya sebelum melakukan data mining sudah harus ada gudang data yang akan kita tambang untuk memperoleh pola-pola, atau trend yang timbul dari suatu aktifitas dalam organisasi/perusahaan. Dalam hal ini tidak akan dibahas mengenai Data Warehousing jadi kita asumsikan Data Warehouse telah tersedia. Setelah Data Warehouse tersedia, akan kita manfa'atkan untuk menggali informasi yang kita perlukan untuk analisa pengadaan barang menggunakan teknik data mining.

Tahapan Data Mining diilustrasikan digambar



Gambar 1 Hubungan antara 3 Sistem penunjang Keputusan

1. Cleaning (Data Eksternal Dan Internal)
Tahapan ini adalah melakukan pembersihan data yang tidak konsisten.
2. Integration. (Managemen Data)
Penggabungan data yang telah melalui tahap cleaning.
3. Transformation. (Managemen Model)
Tahapan ini adalah aktifitas penentuan bentuk data (fitur-fitur dan dimensi) dan mereduksi data sehingga sesuai kebutuhan data mining
4. Data Mining.(Pengolahan Dialog)
Mengeskrak data untuk menemukan pola-pola atau trend tertentu pada data.
5. Evaluation.
Evaluasi pola atau trend yang ditemukan.
Presentation
6. Menyampaikan hasil data mining kepada end user.

2.4 Desain Penelitian

1. Metode Penelitian

Agar penelitian yang dilakukan lebih terarah maka peneliti menggunakan suatu metode deskriptif yang bertujuan untuk mendapatkan gambaran lengkap dari proses dan subjek penelitian.

2. Metode Pengumpulan Data

Data untuk penelitian ini diambil dari data primer dan data skunder. Data primer di dapat dari wawancara dan observasi lapangan. Data sekunder di dapat dari data objek penelitian (database), studi literature dan tulisan ilmiah tentang data mining.

a. Wawancara

Penulis melakukan wawancara dengan para pengambil kebijakan

b. Observasi

Penulis melakukan observasi langsung dengan melihat system kerja perikrutan Karyawan yang sedang berjalan.

c. Literatur

Penulis mengumpulkan literature dan sumber-sumber yang mendukung penelitian.

2. ANALISIS, INTERPRETASI, DAN IMPLIKASI PENELITIAN

2.1. Persiapan Data

Faktor – faktor pertimbangan perekrutan Karyawan

Faktor-faktor pertimbangan perekrutan karyawan pada SMKN 1 NGULING pada dasarnya menggunakan prinsip manual, namun pihak Sekolah menggunakan faktor-faktor pertimbangan pemberian Keputusan yang lebih terinci, yaitu :

1. Jenjang Kelulusan

Jenjang kelulusan merupakan penentu keputusan pertama salah satu syarat masuknya seorang calon karyawan yang nantinya akan ditempatkan di bidang tempat sesuai kelulusannya.

2. Nilai Kelulusan

Nilai kelulusan adalah syarat utama masuknya seorang karyawan yang nantinya menjadi salah satu keputusan penting dimana seorang karyawan dipertimbangkan pula masalah akademis kelulusannya.

3. Tes psikologi

Merupakan tes yang digunakan untuk mengetahui IQ seorang calon karyawan yang berkualitas dan juga menerapkan sebuah implementasi dari kecerdasan calon karyawan yang diuji langsung oleh perusahaan.

4. Tes wawancara

Adalah merupakan syarat wajib masunya calon karyawan yang akan diketahui bagaimana status sosial dan juga intelektualnya tentang keadaan sekitar dunia kerja dan apakah ia siap dan mampu untuk bekerja sesuai kondisi yang di butuhkan perusahaan.

5. Tes Kesehatan

Tes yang digunakan untuk melihat kesahatan calon karyawan, apakah calaon karyawan tersebut pernah atau sedang mengalami suatu penyakit yang dapat mengakibatkan terganggunya aktifitas harian apa tidak, dan tes ini adalah hasil akhir dari serangkaian tes dan penentuan dari tes yang telah dilakukan sebelumnya.

sehingga didapatkan suatu data yang lengkap untuk mendapatkan keputusan yang tepat dan akurat.Selanjutnya untuk menghasilkan keputusan yang lebih singkat dan mudah dipahami oleh pemakai, data lengkap tersebut ditranformasikan dengan menggunakan metoda data tranformasi, dan data-data yang didapat dari pengamatan dikonversikan kedalam bentuk angka-angka seperti dalam tabel berikut :

5.1 Metode Statistik Pohon Keputusan

Tabel 1 : Jenis dan Nilai Data

a. Metoda Data Cleaning

Metoda datacleaning adalah suatu metoda yang digunakan untuk menangani data yang tidak lengkap, menghilangkan noise data dan data yang inkonsisten. Teknik yang digunakan pada metoda data cleaning adalah teknik mean substitution. Dalam teknik

Faktor Pertimbangan	Penilaian		
	Baik	Cukup Baik	Kurang
Jenjang Kelulusan	80 – 100	60 - 79	0 - 59
Nilai Kelulusan	80 – 100	60 - 79	0 - 59
Tes Psikologi	80 – 100	60 - 79	0 - 59
Tes Wawancara	80 – 100	60 - 79	0 - 59
Per tiching	80 – 100	60 - 79	0 - 59

ini, data yang tidak lengkap diisi dengan nilai rata-rata dari sample data. Algoritma dari teknik mean substitution ini adalah sebagai berikut :

Algoritma 1 : Teknik mean substitution
 Input : Data yang tidak lengkap
 Output : Data lengkap

Method :

1. Hitung rata-rata dari sample data

$$x = \frac{\sum x_i}{n}$$
2. Isi data tidak lengkap dengan nilai rata-rata

Maka dari algoritma 1 dapat dilihat pada table 2 berikut ini :

Tabel 2 : Data tidak lengkap

Nama	Jenjang Kelulusan	Nilai Kelulusan	Tes Psikologi	Tes Wawancara	Tes Kesehatan	Keputusan
Dina	90	?	60	75	80	Dipertimbangkan
Mami	80	70	?	80	85	Dipertimbangkan
Andi	90	80	75	85	75	Dipertimbangkan
Jono	90	60	65	?	80	Dipertimbangkan
Rosida	90	80	75	80	80	Diterima
M Sohi	90	80	75	80	80	Diterima
Marta	80	70	70	70	80	Dipertimbangkan
Dicky	80	70	70	75	75	Dipertimbangkan
Khafids	90	80	80	75	75	Diterima
Imania alief	90	80	85	80	90	Diterima

Dari tabel 2 diatas dapat dilihat bahwa terdapat data yang tidak lengkap yaitu pada record 2, 3 dan 4. Nilai rata-rata untuk kelayakan Kelulusan, Psikologi dan Wawancara adalah 61, 65 dan 61. Berdasarkan nilai rata-rata tersebut, tabel 2 yang berisi data tidak lengkap dapat menjadi data lengkap yang dapat dilihat pada tabel 3 berikut ini. Untuk mendapatkan data lengkap maka digunakan metode hitung rata-rata dengan sebagai berikut :

1 contoh kita ambil sampel record 2 :

Rata-rata = jumlah keseluruhan data dari sampel dibagi dengan banyak factor Pertimbangan.

$$x = 305 / 5 (faktor) = 61$$

Nama	Jenjang Kelulusan	Nilai Kelulusan	Tes Psikologi	Tes Wawancara	Tes Kesehatan	Keputusan
Dina	90	61	60	75	80	Dipertimbangkan
Mami	80	70	65	80	85	Dipertimbangkan
Andi	90	80	75	85	75	Dipertimbangkan
Jono	90	60	65	61	80	Dipertimbangkan
Rosida	90	80	75	80	80	Diterima
M Sohi	90	80	75	80	80	Diterima
Marta	80	70	70	70	80	Dipertimbangkan
Dicky	80	70	70	75	75	Dipertimbangkan
Khafids	90	80	80	75	75	Diterima
Imania alief	90	80	85	80	90	Diterima

b. Data Transformasi

Data tranformasi adalah suatu metoda yang digunakan untuk mentranformasikan baris data kedalam data yang ditranformasikan yaitu dengan melakukan normalisasi dan pengumpulan data sehingga menjadi sama. Algoritma dari data tranformasi adalah sebagai berikut :

Algoritma 2 : Data tranformasi

Input : Data lengkap
 Output :Data yang ditranformasikan
 Method :

1. Tentukan nilai terbesar (Xmax),dan nilai terkecil (Xmin)
2. Tentukan range nilai (Xrange = Xmax -Xmin).
3. Tentukan jumlah kelas, $K = 1 + 3,3 \log (n)$.
 $n = \text{Sampel}$

Tentukan nilai interval, $\text{Int} = \text{Xrange} / K$.
 Tranformasikan data dengan $\{Xmin, [Xmin + \text{Int}], \dots, [Xmin + n(\text{Int})]\}$.

Dari algoritma 2 dengan menggunakan data lengkap pada tabel 3 maka dapat dilihat

nilai terkecil untuk masing –masing atribut.

1. Atribut Jenjang Kelulusan

- a. nilai terbesarnya adalah 90 dan nilai terkecilnya 80.
- b. range nilai adalah $90 - 80 = 10$.
- c. jumlah kelas adalah $1 + 3,3 \log (10) = 3$
- d. nilai interval adalah $10/3 = 3$

Tranformasi yang dihasilkan

Nilai terkecil dari sampel + Nilai interval =
Data tranformasi
 $80 - 83 = 1$
 $84 - 87 = 2$
 $88 - 91 = 3$

2. Atribut Nilai Kelulusan

- a. nilai terbesarnya adalah 80 dan nilai terkecilnya 60.
- b. range nilai adalah $80 - 60 = 20$.
- c. jumlah kelas adalah $1 + 3,3 \log (10) = 3$
- d. nilai interval adalah $20/3 = 7$

Tranformasi yang dihasilkan

Nilai terkecil dari sampel + Nilai interval =
Data tranformasi
 $60 - 67 = 1$
 $68 - 75 = 2$
 $76 - 83 = 3$

3. Atribut Tes Psikologi

- a. nilai terbesarnya adalah 85 dan nilai terkecilnya 60.
- b. range nilai adalah $85 - 60 = 25$.
- c. jumlah kelas adalah $1 + 3,3 \log (10) = 3$
- d. nilai interval adalah $25/3 = 8$

Tranformasi yang dihasilkan

Nilai terkecil dari sampel + Nilai interval =
Data tranformasi
 $60 - 68 = 1$
 $69 - 77 = 2$
 $78 - 85 = 3$

4. Atribut Tes Wawancara

- a. nilai terbesarnya adalah 80 dan nilai terkecilnya 61.
- b. range nilai adalah $80 - 61 = 19$.
- c. jumlah kelas adalah $1 + 3,3 \log (10) = 3$
- d. nilai interval adalah $19/3 = 6$

Tranformasi yang dihasilkan

Nilai terkecil dari sampel + Nilai interval =
Data tranformasi
 $61 - 67 = 1$
 $68 - 74 = 2$
 $75 - 81 = 3$

5. Atribut Tes Kesehatan

- a. nilai terbesarnya adalah 90 dan nilai terkecilnya 75.
- b. range nilai adalah $90 - 75 = 15$.
- c. jumlah kelas adalah $1 + 3,3 \log (10) = 3$
- d. nilai interval adalah $15/3 = 5$

Tranformasi yang dihasilkan

Nilai terkecil dari sampel + Nilai interval =
Data tranformasi
 $75 - 80 = 1$
 $81 - 86 = 2$
 $87 - 92 = 3$

Berdasarkan nilai diatas, data lengkap dari tabel 3 dapat ditransformasikan kedalam tabel 4.

Tabel 4 Data Transformasi

Nama	Jenjang Kelulusan	Nilai Kelulusan	Tes Psikologi	Tes Wawancara	Tes Kesehatan	Keputusan
Dina	3	1	1	2	1	Dipertimbangkan
Mami	1	2	1	3	2	Dipertimbangkan
Andi	3	3	3	3	1	Dipertimbangkan
Jono	3	1	1	1	1	Dipertimbangkan
Rosida	3	3	3	3	1	Diterima
M. Sohi	3	3	3	3	1	Diterima
Marta	1	2	2	2	1	Dipertimbangkan
Dicky	1	2	2	3	1	Dipertimbangkan
Khafids	3	3	3	3	1	Diterima
Imania alief	3	3	3	3	3	Diterima

d. Penentuan Keputusan

Penentuan keputusan dengan merupakan hasil dari pemrosesan program agar keputusan yang dihasilkan tepat, cepat, dan akurat. Pada pengentrian data tidak lengkap, memproses data dan menghasilkan keputusan apakah keputusan calon karyawan

tersebut ditolak, dipertimbangkan, atau diterima. Keputusan tersebut dihasilkan dengan ketentuan sebagai berikut :

1. Nilai data yang dihasilkan dari data tranformasi minimal = 1
2. Jumlah kelas = $1 + 3,3 \log (n)$, (n) = banyak sampel
3. Jumlah Max = Jumlah atribut (5) x Jumlah kelas (3) = $5 \times 3 = 15$
4. Range = Jumlah Max –Jumlah Min
= $15-1= 14$
5. Jumlah Keputusan = 3
- Diterima
- Dipertimbangkan
- Ditolak
6. Jumlah interval keputusan
= Range / Jumlah keputusan
= $14 / 3$
= 4.6
= 5

Jumlah interval keputusan digunakan untuk nilai pengklasifikasikan tiap–tiap keputusan maka keputusan yang dihasilkan yaitu :

- 1 sampai 5 = 1 (Keputusan Ditolak)
6 sampai 11 =2 (Keputusan Dipertimbangkan)
> dari 11 = 3 (Keputusan Diterima)

5. PENUTUP

A. Kesimpulan.

Dari penerapan data cleaning dan data transformasi akan mempermudah pimpinan dalam hal ini Kepala SMKN 1 Nguling dalam pengambilan keputusan untuk menerima calon karyawan dengan cepat dan tepat. Sehingga sulitnya mendapatkan data yang lengkap dalam melakukan kunjungan kelapangan dalam ketertutupan calon karyawan dalam memberikan data dapat diatasi. Sehingga dapat meningkatkan cara menangani permasalahan klasifikasi tersebut

Lembaga SMKN 1 Nguling sudah memiliki banyak rombel dan merupakan salah satu sekolah Rujukan di wilayah Nguling Kab Pasuruan, maka perlu adanya suatu penelitian lebih lanjut dalam menganalisis. Utamanya yang berhubungan dengan karyawan yang ada dan yang akan mau bekerja di SMKN 1 Nguling Supaya karyawan yang diterima bekerja benar-benar sesuai dengan apa yang diharapkan.

B. Saran

1. Software yang digunakan sebaiknya di tingkatkan dengan menggunakan database yang terbaru untuk menunjang penyimpanan data dan keamanan jaringan.
2. Perlu adanya penelitian lebih lanjut untuk mencari faktor-faktor yang mempengaruhi kualitas calon pelamar, sehingga dalam penentuan keputusan dan kebijakan dapat ditentukan dengan cepat dan tepat serta akurat.

6. DAFTAR PUSTAKA

1. Turban, E., dkk. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset, 2005
2. Mandasari, V., & Tama, B. A. Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining, 2011
3. Fairuz El Said, 2009. Data Mining – Konsep Pohon Keputusan, diakses dari “<http://fairuzelsaid.wordpress.com/2009/11/24/data-mining-konsep-pohonkeputusan/>”, pada tanggal 12 Maret 2010.
4. Ramdhani, M.A dan Suryadi, K., “Sistem Pendukung Keputusan”, PT. Remaja Rosdakarya, 2003
5. Kusriani dan Emha Taufiq Luthfi, ”Algoritma Data Mining”, Andi Offset, Yogyakarta, 2009.
6. Santosa, Budi. 2007. *Data Teknik Pemanfaatan Data untuk Keperluan Bisnis Teori dan Aplikasi*. Yogyakarta: Graha Ilmu.

7. Santosa, Budi. 2007. *Data Mining Terapan dengan Matlab*. Yogyakarta: Graha Ilmu.
8. Susanto, Sani. 2010. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi
9. Jogiyanto, HM. 2001. *Analisis dan Desain Sistem Informasi Pendekatan Terstruktur Teori dan Praktek Aplikasi Bisnis*. Yogyakarta; Andi.