

Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree

Disusun oleh :
Budanis Dwi Meilani Achmad dan Fauzi Slamet
Jurusan Sistem Informasi
Fakultas Teknologi Informasi. ITATS
Email : dwimeilanibudanis@yahoo.com

ABSTRAK

Data mining merupakan suatu proses untuk menemukan informasi-informasi yang bermanfaat dari sekumpulan *database* yang berukuran besar. Salah satu teknik yang ada pada data mining adalah klasifikasi. Dengan menerapkan teknik klasifikasi pada data-data karyawan dan jadwal kerja, diharapkan nantinya dapat menghasilkan suatu pola-pola tertentu. Dan dari pola-pola tersebut dapat dibuat jadwal kerja yang baru. menggunakan adalah metode *Decision Tree* dan algoritma yang digunakan untuk membentuk pohon keputusan adalah algoritma C4.5. Metode *Decision Tree* merupakan metode yang merubah fakta yang sangat besar menjadi sebuah pohon keputusan yang merepresentasikan aturan-aturan. Pohon keputusan ini juga berguna untuk mengeksplorasi data, serta menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, C4.5, CART. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun.

Kata kunci: *Data mining*, jadwal kerja, *Decision Tree*, algoritma C4.5.

ABSTRACT

Data minning is a process to find useful informations from an unit large size database. The one of technics in data minning is classificat. With use the classificates to the labour date and the work schedule can results a new specific pattern. From the pattern can make a new work schedule by decision tree method and algorithm that used to form decision tree is algoritm C4.5. The Decision Tree methods can change the useful fact become decision tree that presents the rules. The decision tree also used to explores data, and find inner correration between input variable with target variable. More algoritm can used to form decision tree, ie IDR3, C4.5, CART. Algoritm C4.5 is depelomented from algoritm ID3 that make a decision tree from bottom to uppers it is a roof and the uppers is leaf.

Keyword : Data minning, a work schedule, decision tree, algoritm C4.5.

PENDAHULUAN

Setiap perusahaan menginginkan karyawannya dapat bekerja secara maksimal sesuai dengan keahliannya masing-masing. Banyak sekali faktor-faktor yang mempengaruhi kinerja seorang karyawan, seperti umur, jenis kelamin, tingkat kedisiplinan, dll. Berdasarkan faktor-faktor tersebut, maka harus disusun sebuah jadwal kerja yang sesuai dengan kriteria masing-masing karyawan. Namun, muncul beberapa kendala apabila pembuatan jadwal kerja itu dilakukan secara manual. Salah satunya adalah jumlah karyawan yang besar. Dengan jumlah karyawan yang mencapai ratusan orang, akan sangat sulit untuk membuat jadwal kerja dengan mempertimbangkan kriteria dari masing-masing karyawan.

Untuk mengatasi kendala-kendala tersebut, maka dibuat aplikasi yang mampu menghasilkan jadwal kerja yang baik. Dengan melibatkan proses komputasi seperti itu, maka membuat jadwal kerja yang sesuai dengan kriteria para karyawan akan mudah dilakukan. Dan dengan demikian setiap karyawan akan mampu melakukan pekerjaan mereka secara maksimal.

DASAR TEORI

Data mining adalah suatu proses untuk menemukan informasi yang bermanfaat dari sekumpulan database besar yang tersimpan dalam penyimpanan dengan menggunakan teknik

pengenalan pola seperti teknik statistik, matematika, kecerdasan buatan, dan machine learning. 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. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada penilaian berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah:

- Prediksi harga beras dalam tiga bulan mendatang.
- Prediksi presentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

4. Klasifikasi

Di dalam klasifikasi terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Kemudian untuk menentukan pendapatan seorang pegawai, dipakai cara klasifikasi dalam data mining.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang mempunyai kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Contoh pengklusteran dalam bisnis dan penelitian adalah:

- Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial.
- Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Asosiasi mencari kombinasi jenis barang yang akan terjual untuk bulan depan.

METODE DECISION TREE

Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang disebut sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan suhu. Salah satu atribut merupakan atribut yang menyatakan data solusi per *item* data yang disebut atribut hasil. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, C4.5, CART.

ALGORITMA C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Ross Quinlan pada akhir tahun 1970-an. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun.

Secara umum, algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut:

- Hitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
- Pilih atribut sebagai *Node*.
- Buat cabang untuk tiap-tiap anggota dari *Node*.
- Periksa apakah nilai *entropy* dari anggota *Node* ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai *entropy* anggota *Node* adalah nol, maka proses pun berhenti.
- Jika ada anggota *Node* yang memiliki nilai *entropy* lebih besar dari nol, ulangi lagi proses dari awal dengan *Node* sebagai syarat sampai semua anggota dari *Node* bernilai nol.

Node adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *gain* suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut:

$$Gain(S, A) = Entropy(S) - \left(\sum_{i=1}^n \frac{A_i}{S} * Entropy(A_i) \right)$$

Keterangan: S : Kasus.
 A : Atribut
 n : Jumlah partisi atribut A
 A_i : Jumlah kasus pada partisi ke- i .
 S : Jumlah kasus.

Sementara itu, untuk menghitung nilai *Entropy* dapat dilihat pada persamaan berikut ini:

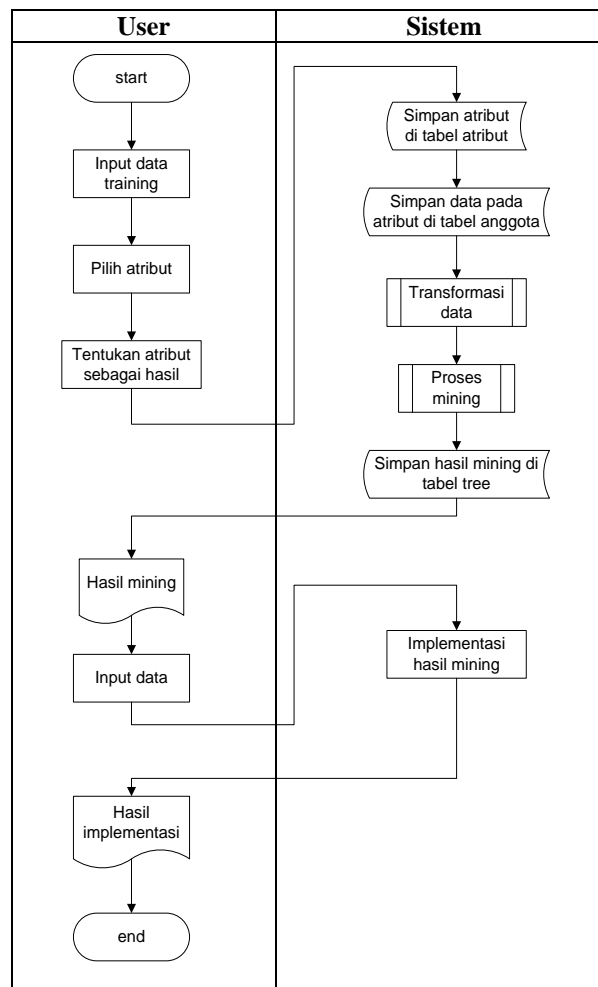
$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan: S : Himpunan kasus.
 n : Jumlah partisi S
 p_i : Proporsi dari S_i ke S .

PERANCANGAN SISTEM

Pada tahap perancangan sistem ini, akan dirancang suatu sistem dalam suatu bagan yang menunjukkan alur proses yang terjadi di dalam sistem tersebut. Berikut ini akan ditunjukkan proses-proses yang berperan penting dalam perancangan sistem yang ditampilkan dalam bagan alir dokumen sistem.

Bagan alir dokumen :



Gambar 1 Bagan Alir Dokumen Sistem

di atas menunjukkan proses-proses yang terjadi di dalam sistem. Diawali dengan menginputkan data yang akan dipakai untuk membuat pohon keputusan, proses selanjutnya adalah memilih atribut, kemudian menentukan atribut hasil dari atribut-atribut yang telah dipilih. Setelah itu atribut yang dipilih beserta anggotanya akan disimpan ke dalam database. Setelah disimpan, seluruh data akan ditransformasi ke dalam format yang dibutuhkan. Kemudian proses mining pun mulai dijalankan, dan hasilnya akan disimpan ke dalam database lebih dulu kemudian ditampilkan ke *user*. Setelah pohon keputusan terbentuk, proses selanjutnya adalah menginputkan data yang belum mempunyai jadwal kerja. Kemudian proses pembentukan jadwal kerja yang baru akan dimulai. Jadwal kerja dibentuk berdasarkan *rule-rule* dari pohon keputusan yang dihasilkan.

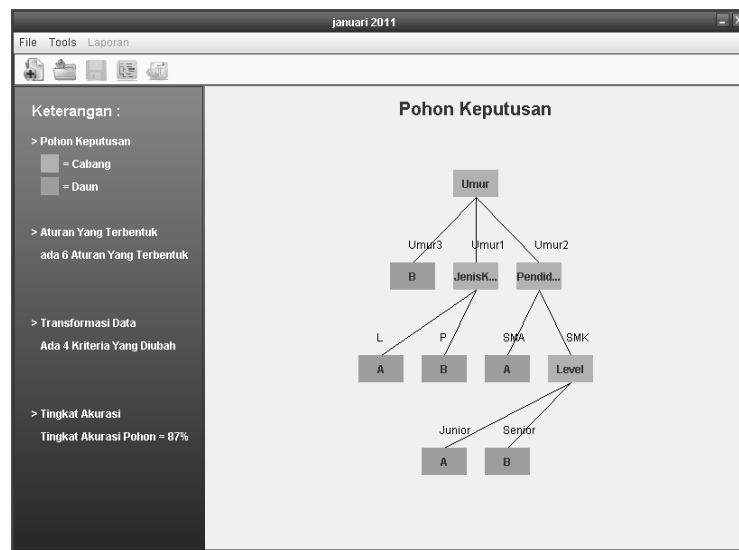
IMPLEMENTASI SISTEM

Setelah melalui tahap perancangan sistem, tahap berikutnya adalah implementasi sistem. Untuk menjalankan sistem ini, hal pertama yang dilakukan adalah menginputkan data untuk membentuk pohon keputusan. Data yang diinputkan sudah tersimpan di dalam database server.

NP	Nama	Umur	JenisKelamin	Pendidikan	Agama	Level	LamaKerja	Kesehatan	Jadwal
2104	Gandung A.	22	L	SMA	Islam	Junior	1	Baik	A
2107	Hadi Purwa...	25	L	SMA	Islam	Junior	3	Baik	A
2103	Intan Nurha...	26	P	SMA	Islam	Senior	3	Tidak Baik	A
2108	Fuad Hasyim	32	L	SMA	Kristen	Senior	6	Baik	B
2106	Anjar Widhi...	27	P	SMK	Islam	Junior	1	Baik	A
2110	Meida Nurk...	28	P	SMK	Islam	Senior	7	Baik	B
2114	Ratih Rah...	29	P	SMA	Islam	Senior	8	Tidak Baik	A
2111	Agus Rifki	20	L	SMA	Kristen	Junior	0	Baik	A
2116	Subandriyo	30	L	SMA	Islam	Junior	0	Baik	A
2115	Nanik Sulis...	24	P	SMA	Kristen	Junior	3	Baik	B
2118	Siti Inayah	34	P	SMA	Islam	Junior	4	Tidak Baik	B
2119	Latifatun Ni...	23	P	SMK	Islam	Junior	2	Baik	B
2123	Nahari	22	L	SMA	Islam	Junior	1	Baik	A
2135	Muhamma...	21	L	SMA	Islam	Junior	3	Baik	A
2128	Tatar Kusni...	31	P	SMA	Islam	Senior	5	Tidak Baik	A
2154	Agus Suga...	36	L	SMA	Kristen	Senior	8	Baik	B

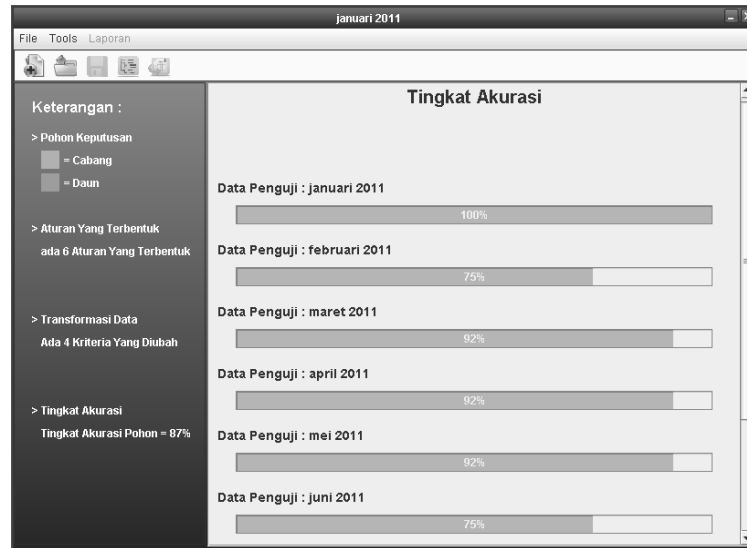
Gambar 2 Data-data yang digunakan untuk membentuk pohon keputusan

Setelah data diinputkan, data akan ditransformasi dan diproses untuk menghasilkan pohon keputusan. Pohon keputusan akan terbentuk jika data yang diinputkan adalah data yang memang mempunyai pola tertentu dan mempunyai jumlah yang cukup besar.



Gambar 3 Pohon keputusan yang terbentuk

Pohon keputusan di atas terdiri dari 4 atribut yang menjadi Node. Node dilambangkan dengan kotak berwarna coklat, sedangkan daun dilambangkan dengan kotak berwarna hijau. Atribut-atribut yang menjadi *Node* antara lain: Umur, JenisKelamin, Pendidikan, dan Level. Anggota atribut umur ditransformasi sehingga hanya menjadi 3 anggota, yaitu: Umur1, Umur2, dan Umur3. Pohon keputusan di atas akan menghasilkan *rule-rule* yang akan digunakan untuk membentuk sebuah jadwal kerja yang baru. Untuk mengetahui apakah *rule-rule* tersebut sudah sesuai, maka dilakukan pengujian dengan data-data asli. Berdasarkan pengujian tersebut, diperoleh nilai akurasi dari *rule-rule* yang telah terbentuk. Nilai akurasi inilah yang akan menentukan apakah *rule* yang dihasilkan akan digunakan atau tidak.



Gambar 3 Nilai akurasi yang terbentuk

Setelah pohon keputusan terbentuk, masukkan data-data yang belum mempunyai jadwal kerja. Kemudian proses pembentukan jadwal kerja baru akan dimulai. Jadwal kerja yang baru, dibentuk berdasarkan *rule-rule* yang mempunyai nilai akurasi tertinggi.

NP	Nama	Jadwal
2104	Gandung Arisal	A
2107	Hadi Purwanto	A
2103	Intan Nurhayati	A
2108	Fuad Hasyim	B
2106	Anjar Widhiasih	A
2110	Meida Nurkhotimah	B
2114	Ratih Rahmawati	A
2111	Agus Rifki	A
2116	Subandiyo	A
2115	Nanik Sulistyowati	B
2118	Siti Inayah	B
2119	Latifatun Nikmatina	B
2123	Nahani	A
2135	Muhammad Zahrani	A
2128	Tatar Kusniyati	A
2154	Agus Sugandi	B
2144	Siti Ngaisah	A

Gambar 4. Hasil pembentukan jadwal baru

Berikut ini *rule-rule* yang dipakai untuk membuat jadwal kerja yang baru di atas:

- Jika Umur = Umur3 maka Jadwal = B
- Jika Umur = Umur2 dan JenisKelamin = L maka Jadwal = A
- Jika Umur = Umur2 dan JenisKelamin = P maka Jadwal = B
- Jika Umur = Umur2 dan Pendidikan = SMA maka Jadwal = A
- Jika Umur = Umur2 dan Pendidikan = SMK dan Level = Junior maka Jadwal = A
- Jika Umur = Umur2 dan Pendidikan = SMA dan Level = Senior maka Jadwal = B

Rule-rule tersebut berasal dari pohon keputusan yang dibuat berdasarkan data bulan Januari 2011. Pohon keputusan Januari 2011 dipilih karena memiliki tingkat akurasi paling tinggi di antara pohon keputusan lainnya. Tingkat akurasi pohon keputusan Januari 2011 mencapai 87% dengan rincian sebagai berikut.

Tabel 1 Hasil pengujian pohon keputusan januari 2011

Data Penguji	Tingkat Akurasi
Januari 2011	100%
Februari 2011	75%
Maret 2011	92%
April 2011	92%
Mei 2011	92%
Juni 2011	75%

KESIMPULAN

Dari perancangan dan pembuatan Aplikasi Penjadwalan Kerja ini, dapat ditarik kesimpulan bahwa:

1. Aplikasi Penjadwalan Kerja ini mampu membuat jadwal kerja baru berdasarkan jadwal kerja lama dengan menggunakan metode *Decision Tree*.
2. Aplikasi Penjadwalan Kerja ini akan memudahkan *user* dalam membuat jadwal kerja karena aplikasi ini mampu membuat jadwal kerja dengan waktu yang singkat.
3. Rule yang digunakan adalah rule **januari 2011** karena rule tersebut memiliki nilai akurasi paling tinggi yaitu sebesar 87%.
4. Atribut hasil yang digunakan adalah atribut jadwal. Jika atribut lain yang digunakan, maka pohon keputusan tidak akan terbentuk karena kurangnya jumlah data.
5. Data baru yang diinputkan harus memiliki atribut pada pohon. Jika ada salah satu atribut pada pohon yang tidak disertakan pada data baru, maka jadwal kerja yang baru tidak akan terbentuk.

DAFTAR PUSTAKA

- [1]. **Larose, Daniel T.** (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.
- [2]. **Han, J. Kamber, M.** (2001). *Data Mining: Concept and Techniques*. San Fransisco: Morgan Kaufmann Publishers.
- [3]. **Fayyad, Usama.** (1996). *Advance in Knowledge Discovery and Data Mining*. MIT Press.
- [4]. **Turban, E., dkk.** (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: Andi Offset.
- [5]. **Kusrini.** (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset