

ALAT BANTU KLASIFIKASI DENGAN POHON KEPUTUSAN UNTUK SISTEM PENDUKUNG KEPUTUSAN

Tessy Badriyah¹⁾, Ria Rahmawati²⁾

¹⁾Politeknik Elektronika Negeri Surabaya - ITS, email: tessy@eepis-its.edu

²⁾Politeknik Elektronika Negeri Surabaya - ITS, email: rara63rk@yahoo.com

ABSTRAKSI

Penelitian ini bertujuan membuat alat bantu klasifikasi yang menerapkan model pohon keputusan dengan menggunakan algoritma ID3.

Tahapan yang dilakukan dalam pembuatan alat bantu atau tools klasifikasi data ini adalah membuat interface dengan melibatkan database, GUI (Graphical User Interface), dan program. Penggunaan antar muka dimulai dengan membuat tabel, menginputkan parameter masing-masing atribut, dan menginputkan data pada tabel. Kemudian digenerate ke dalam bentuk tree dan rule If-Then.

Dari pembuatan alat bantu klasifikasi ini dapat dihasilkan tree dan rule If-Then yang membentuk klasifikasi data secara benar

Kata kunci: klasifikasi, algoritma ID3

PENDAHULUAN

Penggunaan Model Pohon Keputusan (*Decision Tree*) adalah salah satu teknik klasifikasi sebagai bagian dari ilmu Data Mining. Data Mining melakukan penggalian pengetahuan (*knowledge*) terhadap data. Untuk teknik klasifikasi ini, digunakan algoritma ID3 yang pertama kali dikembangkan oleh Ross Quinlan di Universitas Sydney pada tahun 1975 dalam bukunya : *Machine Learning*, vol. 1, no. 1. ID3 is based off the Concept Learning System (CLS) algorithm.

Kelebihan dari penggunaan model Pohon Keputusan ini selain mudah dipahami juga dapat digunakan untuk menemukan aturan atau syarat-syarat yang dapat dijadikan sebagai kriteria yang berguna untuk keperluan analisa dalam suatu proses pengambilan keputusan.

Permasalahan yang diangkat dalam penelitian ini adalah:

1. Bagaimana menggunakan klasifikasi yang tepat dan sesuai dengan kasus yang diberikan sehingga didapatkan bentuk pohon yang dinamis.
2. Bagaimana mengimplementasikan algoritma ID3 untuk membangkitkan Pohon Keputusan.
3. Bagaimana mengubah suatu pohon keputusan yang menggunakan Algoritma ID3 ke dalam bentuk aturan keputusan berupa aturan *if-then*.

Tujuan dari penelitian ini adalah untuk membuat alat bantu klasifikasi dan identifikasi data dengan model Pohon Keputusan menggunakan algoritma ID3 dan mendapatkan hasil yang optimal dengan pemahaman akan teknik yang digunakan terhadap model permasalahan yang dihadapi.

METODE YANG DIGUNAKAN

Penelitian ini menggunakan pemodelan Pohon Keputusan. Struktur dari pohon keputusan dibangun dari 3 tipe simpul: simpul root, simpul perantara, dan

simpul leaf. Simpul leaf memuat suatu keputusan akhir atau kelas target untuk suatu pohon keputusan. Simpul root adalah titik awal dari suatu decision tree. Setiap simpul perantara berhubungan dengan suatu pertanyaan atau pengujian.

Dasar algoritma untuk menginduksi Pohon Keputusan:

- Pohon dibangun dalam suatu metoda rekursif topdown divide-and-conquer.
 1. Seluruh contoh pelatihan dimulai dari simpul root, lalu dilakukan pengujian
 2. Mencabang ke jalur yang benar berdasarkan hasil pengujian.
 3. Apakah simpul leaf ditemukan? Jika ya, masukkan contoh ini ke kelas target, jika tidak kembali ke langkah 1.
- Atribut-atribut berada dalam suatu kategori (jika bernilai kontinu, nilai-nilai tersebut didiskritkan terlebih dahulu)
- Contoh-contoh dipartisi secara rekursif berdasarkan atribut terpilih. Kondisi-kondisi penghentian partisi adalah seluruh sampel untuk suatu simpul yang diberikan masuk ke kelas yang sama.
- Atribut-atribut uji dipilih berdasarkan heuristik atau pengukuran statistik (misal, information gain).

Ukuran untuk melakukan seleksi atribut dilakukan dengan memilih atribut dengan information gain terbesar. Jika S memuat tupel-tupel si dari kelas Ci untuk i = {1,...,m}. Maka ukuran-ukuran informasi info diperlukan untuk mengklasifikasikan tuple sembarang, adalah:

$$I(s1, s2, \dots, sm) = -\sum_{i=1}^m \frac{si}{s} \log 2 \frac{si}{s}$$

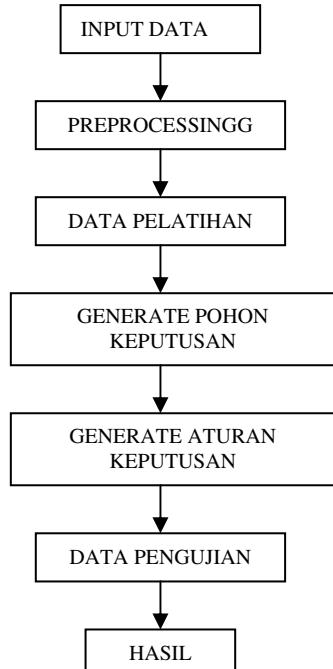
Entropy dari atribut A dengan nilai-nilai {a1,a2,...,av}

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj})$$

Informasi yang diperoleh melalui percabangan pada atribut A:

$$\text{Gain}(A) = I(S_1, S_2, \dots, S_m) - E(A)$$

Perencanaan pembuatan sistem untuk klasifikasi data ini dapat digambarkan secara umum pada block diagram berikut ini :



Gambar 1. Blok Diagram Sistem

Adapun untuk pembuatan sistem yang akan dilakukan dalam pengklasifikasian data menjadi tiga elemen yaitu database, GUI, dan program. Database yang akan dibuat menggunakan MS.Access sedangkan dalam pembuatan program dan Graphical User Interface menggunakan java programming.

Dalam pengklasifikasian data terdapat parameter atribut yang harus dimasukkan yaitu atribut non-kategori dan atribut target.

Agar data training yang telah disimpan dalam database aplikasi microsoft access dapat diakses maka perlu dibangun sebuah koneksi menggunakan Java Database Connectivity (JDBC). JDBC adalah Application Programming Interface (API) yang dirancang untuk mengakses database berdasarkan SQL (Structure Query Language). Untuk mengakses database dengan DBMS yang berbeda maka JDBC memerlukan Driver. Untuk mengakses database MS-Access JDBC memerlukan driver yang menjembatani aplikasi dengan ODBC (Open Database Connectivity) yang ada dilingkungan Microsoft Windows.

Aplikasi Java → JDBC → ODBC → MS-Access

Untuk mengaktifkan hubungan antar aplikasi dan database, maka diperlukan inisialisasi driver menggunakan ODBC Driver. Koneksi dibentuk melalui 1 class yaitu `java.sql.DriverManager` dan 2 interface yaitu `java.sql.Driver` dan `java.sql.Connection`. Class untuk JDBC diakses melalui `java.sql.*`.

```

Connection c;
String url = "jdbc:odbc:decisiontree";
String myLogin = "db2admin";
String myPassword = "db2admin";
try {
    Class.forName(
        "sun.jdbc.odbc.JdbcOdbcDriver");
    Connection c =
        DriverManager.getConnection(
            url, myLogin, myPassword);
    ....
}
  
```

- Class.forName(nama-driver)**, merupakan loading driver, artinya koneksi akan dibuat dengan menggunakan jalur driver tersebut.
- DriverManager.getConnection(url,myLogin, myPassword)**, menghubungkan aplikasi dengan nama database yang akan diakses, dilengkapi dengan otentifikasi berupa nama user dan password.

HASIL DAN PEMBAHASAN

Hasil dari pembuatan alat bantu klasifikasi menggunakan data pelatihan yang nantinya akan digenerate ke dalam model pohon keputusan. Data pelatihan memiliki atribut non-kategori dan atribut kategori (target). Contoh data pelatihan yang digunakan sebagai berikut:

Tabel 1. Contoh Data Pelatihan

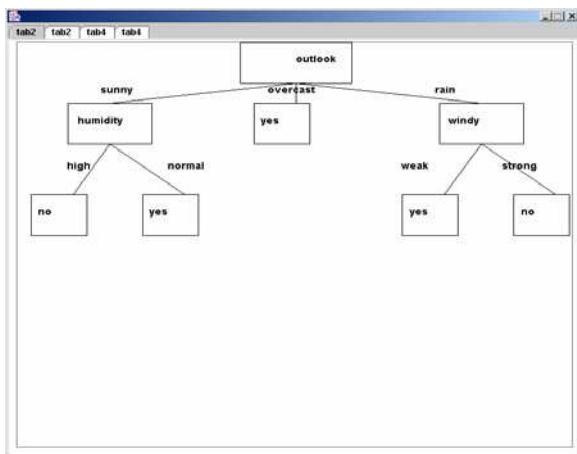
Outlook	Temperature	Humidity	Windy	Play/Don't play
Sunny	hot	high	false	don't play
Sunny	hot	high	true	don't play
overcast	hot	high	false	play
rain	mild	high	false	play
rain	cool	normal	false	play
rain	cool	normal	true	don't play
overcast	cool	normal	true	play
sunny	mild	high	false	don't play
sunny	cool	normal	false	play
rain	mild	normal	false	play
sunny	mild	normal	true	play
overcast	mild	high	true	play
overcast	hot	normal	false	play
rain	mild	high	true	don't play

Untuk memasukkan data pelatihan tersebut, pada proses pembuatan GUI ada pada menu File yang terdiri dari terdiri dari 3 menu item yaitu:

1. **baru**, digunakan untuk membuat data baru dengan menginputkan parameter masing-masing atribut berserta nilainya.
2. **input**, digunakan untuk menginputkan dan mengisi tabel dengan parameter yang telah dibuat sebelumnya.
3. **exit**, digunakan untuk keluar dari window.

Setelah data pelatihan dimasukkan, proses berikutnya adalah memilih menu *Proses* untuk menggenerate Pohon Keputusan.

Hasil eksekusi menu *Proses* dapat dilihat pada menu *Report* yang terdiri dari output berupa pohon keputusan, output data testing, output rule *If-Then*, dan output nilai average entropi masing-masing atribut. Untuk contoh data pelatihan seperti yang ada pada table sebelumnya akan menghasilkan output pohon keputusan seperti gambar berikut:



Gambar 2. Window output tree

Dari output pohon keputusan pada Gambar 2, didapatkan aturan if-then sebagai berikut:

```
If attribute1 = sunny and attribute3 = high
    Then Classification = don't play
If attribute1 = sunny and attribute3 = normal
    Then Classification = play
If attribute1 = overcast
    Then Classification = play
If attribute1 = rain and attribute4 = false
    Then Classification = play
If attribute1 = rain and attribute4 = true
    Then Classification = don't play
```

Hasil data testing pada tabel merupakan hasil klasifikasi dari nilai atribut non-kategori (target) yang tidak diketahui. Dengan menggunakan model Pohon Keputusan dan aturan If-Then yang dihasilkan, nilai dari atribut target ini bisa diketahui.

Pembahasan untuk keakuratan klasifikasi data yang dihasilkan dilakukan dengan beberapa tabel yang ingin di uji coba untuk mengetahui seberapa benar pengklasifikasian data dengan menggunakan alat bantu yang telah dibuat. Dengan menggunakan data yang sama tetapi beda nilai atribut target untuk setiap barisnya menghasilkan *undefined* pada tree. *Undefined* terjadi jika pada tree terdapat nilai yes dan

no pada leaf node terakhir. Sehingga tidak bisa menentukan apakah akhir cabang itu bernilai yes atau no. Cabang yang terdapat nilai yes dan no pada leaf node tidak dapat dicari leaf node lagi karena tidak adanya atribut non-kategori yang harus sudah habis terpakai semuanya. Keakuratan data dapat dilihat dari data testing sebelum dan sesudah pengujian.

KESIMPULAN DAN SARAN

Kesimpulan

1. Teknologi data mining dengan menggunakan pemodelan Pohon Keputusan dapat digunakan untuk melakukan klasifikasi data. Dan hasil dari klasifikasi yang didapatkan dengan menggunakan alat bantu ini adalah dengan ditemukannya aturan-aturan yang dapat digunakan sebagai kriteria-kriteria untuk mengidentifikasi dan mengetahui apakah suatu tindakan atau kegiatan harus dilakukan atau tidak.
2. Dalam beberapa kali percobaan dengan menggunakan data pelatihan yang berbeda, ditemukan *undefined* pada tree. *Undefined* terjadi jika pada tree terdapat nilai yes dan no pada leaf node terakhir sehingga tidak bisa ditentukan nilai dari akhir cabang tersebut atau karena adanya atribut non-kategori yang sudah habis terpakai semuanya.

Saran

1. Pengklasifikasian data umum yang menggunakan teknik datamining ini tidak sepenuhnya benar dan akurat. Oleh karena itu sebagai kepastian perlu adanya penganalisaan yang lebih tajam dan sewaktu-waktu diperlukan campur tangan seorang pakar.
2. Perlu dikembangkan teknik yang baru dalam klasifikasi data dengan mengembangkan teknik baru yang dapat melakukan klasifikasi dengan lebih akurat dan dapat mengatasi terjadinya *undefined* pada pembentukan klasifikasi dengan model Pohon Keputusan.

DAFTAR PUSTAKA

- [1] Blake, C.L. & Merz, C.J. (1998) *UCI Repository of machine learning databases*. Irvine, CA: University of California, Department of Information and Computer Science.
- [2] Fayyad, Usama. "Advances in Knowledge Discovery and Data Mining". MIT Press. 1996.
- [3] J.Gehrke, R. Ramakrishnan and V. Ganti. "Rainforest: A Framework for fastdecision tree construction of large datasets". InProc. 1998 Int.conf Very Large Data Bases (VLDB), 1998.
- [4] Kononenko, I. (1991) Semi-naive Bayesian classifier. *Proceedings Sixth European Working Session on Learning*, 206-219. Berlin: Springer-Verlag.
- [5] R. Agrawal, et.al., "Data mining: A Performance Perspective", *IEEE Transaction on Knowledge and Data Engineering*, 1993, pp.914-925.

