

APLIKASI DATA MINING DENGAN METODE CLASSIFICATION BERBASIS ALGORITMA C4.5

Rizky Tahara Shita¹⁾, Nita Marliani²⁾

^{1, 2)}Universitas Budi Luhur, Fakultas Teknologi Informasi,
Jl. Ciledug Raya. Petukangan Utara. Jakarta Selatan, Jakarta, 12260
HP: +62 857 1648 3190
E-mail: rizky.tahara@gmail.com¹⁾, nita.marliani@gmail.com²⁾

Abstrak

PT. Biro Klasifikasi Indonesia telah memiliki kumpulan data yang besar; baik data pegawai, keuangan, kapal, surveyor maupun laporan survey yang sebenarnya data tersebut berpotensi untuk menghasilkan informasi baru yang berguna dalam menyusun berbagai strategi maupun sebagai penunjang keputusan. Dengan memanfaatkan teknik Classification pada data mining yang berbasis algoritma C4.5 (pembentukan pohon keputusan) untuk membagi kumpulan data yang besar menjadi himpunan record yang lebih kecil, maka dapat memperkirakan kelas dari suatu atribut dan diharapkan dapat membentuk pohon keputusan yang merepresentasikan aturan dalam data yang digunakan agar dapat dipahami dan disajikan dalam bentuk yang lebih informatif. Pohon keputusan yang ada dapat digunakan untuk memperkirakan terlambat atau tidaknya laporan survey pada divisi survey dengan melihatnya berdasarkan data cabang, jenis kapal, jabatan surveyor maupun jenis survey.

Kata kunci: data mining, classification, C4.5, survey

Abstract

PT. Biro Klasifikasi Indonesia already has lots of data; from employee, finance, ship, surveyor and the survey report which potentially to generate new useful information to form strategy as decision support. By using the Classification technique in data mining with C4.5 algorithm (build the decision tree) to divide big data into smaller records, then predicting a class in attribute and building the decision tree that represent rules in the data that used to understand and served in more informative way. Existing decision tree can be used to predict how late or not the survey report at the survey division by seeing it based on branch, ship type, surveyor position nor the survey type.

Keywords: data mining, classification, C4.5, survey

1.PENDAHULUAN

PT. Biro Klasifikasi Indonesia yang memiliki kumpulan data yang besar, sangat berpotensi untuk digunakan dalam penyampaian informasi yang lebih baik sebagai penunjang keputusan. Memanfaatkan konsep data *mining* dengan algoritma C4.5 maka diharapkan penyampaian data menjadi informasi sebagai penunjang keputusan dapat tercapai.

1.1 Tujuan

- Menganalisis dan merancang aplikasi berbasis *web* yang dijadikan dasar untuk membangun sistem data *mining* perkiraan laporan survey dengan memanfaatkan metode *Classification* berbasis algoritma C4.5.
- Meningkatkan akses informasi dari mana saja dan kapan saja secara cepat dan mudah.
- Mempermudah pengambilan keputusan.

1. LANDASAN TEORI

1.1 Pengertian Data Mining

Merupakan serangkaian proses untuk menggali nilai tambah yang berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan proses ekstraksi dan mengenali pola

penting dari data yang ada [1]. Berdasarkan tugas yang dapat dilakukan, data mining dibagi menjadi beberapa kelompok, yaitu: Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklusteran dan Asosiasi.

1.2 Classification Rule

Sebuah proses untuk menemukan model atau fungsi yang menjelaskan / membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Hal ini juga dapat dikatakan sebagai pembelajaran (klasifikasi) yang memetakan sebuah unsur (*item*) data kedalam salah satu dari beberapa kelas yang sudah didefinisikan [2].

1.3 C4.5 Algorithm

Pembuatan pohon keputusan menggunakan algoritma C4.5 yang merupakan pengembangan dari algoritma ID3; dimana pengembangan dilakukan dalam hal mengatasi *missing* data, data *continue*, *pruning* [3]. Secara umum, algoritma C4.5 untuk membangun pohon keputusan dimulai dari pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai; membagi kasus dalam cabang dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2. PROSES BISNIS

2.1 Analisa Kebutuhan

Kebutuhan PT. Biro Klasifikasi Indonesia adalah mengurangi tumpukan data yang ada agar dapat diolah menjadi informasi yang lebih bermanfaat. Data yang dibutuhkan dari sistem adalah data yang berhubungan dengan laporan survey, nilai atribut dan kemungkinannya. Variabel yang digunakan yaitu:

- Jenis survey; berisi seluruh kemungkinan jenis survey yang dilakukan pada setiap permohonan.
- Cabang; berisi seluruh kemungkinan cabang yang melakukan survey pada setiap permohonan.
- Kategori kapal; berisi seluruh kemungkinan jenis kapal yang di survey pada setiap permohonan.
- Jabatan; variabel tentang kemungkinan jabatan yang dimiliki oleh surveyor.
- Perkiraan; berfungsi untuk menentukan hasil keputusan yang pengelompokannya sudah ditentukan secara tetap agar tidak terjadi kesalahan dalam proses dan data perkiraan ini memiliki 2 nilai; yaitu: “tepat” dan “terlambat”.

2.2 Rancangan Basisdata

Rancangan basisdata dapat dilihat pada *class diagram* gambar 1 .

3. HASIL

a. Implikasi Penelitian

3.1.1 Aspek Manajerial

Pihak manager dan bagian survey dapat melihat secara detil mengenai keterlambatan laporan survey.

3.1.2 Aspek Sistem

Aplikasi tidak mengganggu sistem yang berjalan (*legacy system*) karena merupakan pendukung (*add on*) dari aplikasi yang sudah ada. Dengan dukungan aplikasi *web based*, maka data serta laporan dapat diakses dan disajikan dengan mudah dan cepat oleh pihak manager maupun bagian survey sebagai penunjang keputusan.

3.1.3 Aspek Penelitian Lanjut

Dapat dikembangkan agar proses lebih cepat dengan memanfaatkan metode lain yang disesuaikan dengan data yang ada untuk jumlah yang lebih besar.

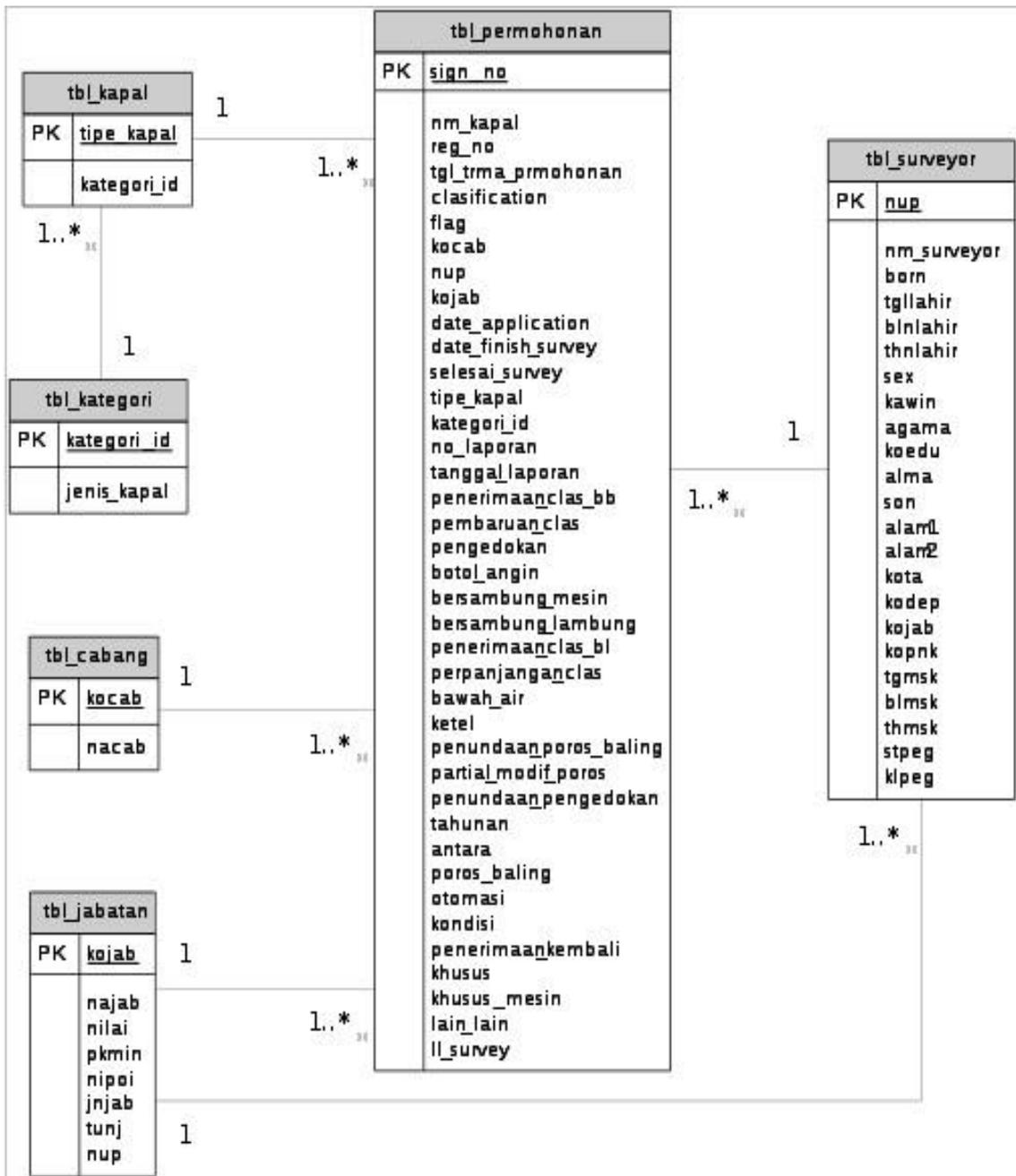
4. PENUTUP

4.1 Kesimpulan

Implementasi *Classification* berbasis algoritma C4.5 pada aplikasi sangat membantu dalam menganalisis perkiraan keterlambatan laporan survey dengan cara membentuk pohon keputusan yang memanfaatkan data permohonan yang ditunjang dengan aplikasi berbasis *web*, sehingga pihak terkait dapat lebih cepat dan mudah dalam mendapatkan informasi untuk menunjang keputusan secara *mobile*.

4.2 Saran

Agar lebih baik, aplikasi dapat dikembangkan agar dapat bekerja lebih cepat; mengembangkan sisi basisdata agar dapat memproses data yang jumlahnya sangat besar serta pemilihan proses *import* data selain format .xls.



Gambar 1: Class Diagram

4.3 Tampilan Aplikasi

4.3.1 Import Data



Gambar 2: Import Data

4.3.2 Laporan



Gambar 3 Laporan

4.3.3 Data Mining

Data Permohonan Survey

No. Survei	no kapal	reg	no tgl	nama perusahaan	klasifikasi	flag	tanggal input data	aplikasi
1	OKZENDE GALANGGA-41	1041	2010-01-24	000000	001	INDONESIA	01	2010-03-08 08:00:00
2	OKKOGED ZY 20	2715	2010-02-03	000000	001	INDONESIA	02	2010-03-27 08:00:00
3	OKLIGED ANANDA SEACE	2347	2010-06-03	000000	001	INDONESIA	03	2010-03-28 08:00:00
4	OKHPSGD SRIHATI 2 II	2181	2010-05-25	000000	001	INDONESIA	04	2010-03-29 08:00:00
5	OKHUGED ANANDA - XE	1467	2010-07-10	000000	001	INDONESIA	05	2010-07-14 08:00:00
6	OKHUSGD SUKARNO	1465	2010-04-10	000000	001	INDONESIA	06	2010-04-06 08:00:00
7	OKDUGED SUKUMA SILAS	6086	2010-06-03	000000	001	INDONESIA	07	2010-07-01 08:00:00
8	OKLIGED WANGITI SATU	2182	2010-04-27	000000	001	INDONESIA	08	2010-03-28 08:00:00
9	OKKPSGD SUMBERTO	6031	2010-07-03	000000	001	INDONESIA	09	2010-07-15 08:00:00
10	OKDPSGD SRI SULAWESI	1448	2010-05-25	000000	001	INDONESIA	10	2010-03-29 08:00:00
11	OKDUGED SONARA C		2010-04-03	000000		INDONESIA	01	2010-04-24 08:00:00
12	OKHUGED SARAWAN I	2213	2010-07-10	000000	001	INDONESIA	12	2010-07-16 08:00:00
13	OKLUSGD ANINDI	3463	2010-06-10	000000	001	INDONESIA	13	2010-05-18 08:00:00
14	OKKUSGD SISAKA	3072	2010-06-22	000000	001	INDONESIA	14	2010-05-22 08:00:00

Filter Data

No	nama atribut	nilai	nilai atribut	jumlah kasus	total	jumlah kasus tepat	jumlah
1	jenis	Tidak	Tidak	30	12	8	
2	jenis	jenis_survey perikanan_aku	1	1	1	1	
3	jenis	jenis_survey perikanan_aku	4	1	3		
4	jenis	jenis_survey perikanan_aku	9	7	2		

Pohon Keputusan

```

jenis = 00 ITyagk = 0. Terlamban = 01. Hasil
jenis = 01 ITyagk = 0. Terlamban = 01. Terlamban
jenis = 02 ITyagk = 0. Terlamban = 01. Hasil
jenis = 03 ITyagk = 0. Terlamban = 01. Hasil
jenis = 04 ITyagk = 0. Terlamban = 01. Terlamban
jenis = 05 ITyagk = 0. Terlamban = 01. Hasil

```

Gambar 4 Proses Data Mining

5. DAFTAR PUSTAKA

- [1] Huda, Nugson Masykur, 2010. *Aplikasi Data Mining untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*. Universitas Diponegoro.
- [2] Kusriani, dan Luthfi, Emha Taufiq, 2009. *Algoritma Data Mining*. Yogyakarta: C.V Andi Offset.
- [3] Moertini, Veronica Sri, 2007. *Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 dengan Pendekatan Konsep Operator Relasi*. Universitas Pajajaran.