

KAJIAN PENERAPAN ALGORITMA DATA MINING TERHADAP PEMILIHAN MITRA KERJA PENYEDIA JASA TRANSPORTASI DI JAKARTA

Harry Dhika¹⁾, Tri Yani Akhirina²⁾, Surajiyo³⁾

^{1), 2), 3)} Teknik Informatika Universitas Indraprasta PGRI
Jl Nangka No. 58B, Jakarta, DKI Jakarta 12530

dhikatr@yahoo.com¹⁾, azizahputriku@gmail.com²⁾, drssurajiyo@yahoo.co.id³⁾

Abstrak

Penelitian ini membahas tentang cara menerapkan algoritma C4.5 yang berupa pohon keputusan terhadap pemilihan mitra kerja jasa transportasi di DKI Jakarta. Analisa dilakukan terhadap 10 perusahaan di DKI Jakarta dengan luaran metode atau rule algoritma C4.5 terhadap pemilihan mitra kerja jasa transportasi. Data dianalisis menggunakan perangkat lunak Rapidminer yakni salah satu software berbasis opensource. Tingkat akurasi data yang diperoleh adalah 94,37% dengan class recall mencapai 90.91% terhadap nilai True Very Good. Untuk nilai class recall true good nilai akurasi mencapai 94.12% sedangkan untuk class recall true bad nilainya mencapai 100%. Penelitian menggunakan konsep data mining dengan framework KDD (Knowledge Discovery in Databases) sehingga dalam menentukan pemilihan mitra kerja melalui beberapa tahapan atau tingkat. KDD memiliki tahapan diantaranya tahap pertama yakni tujuan utama KDD, seleksi dan audisi, tahap preprosesan awal:perbaikan data, transformasi data, menerjemahkan data mining, evaluasi dan presentasi data serta penyimpulan akhir data.

Kata kunci: Algoritma C4.5, Data Mining, Sistem Informasi, Jasa Transportasi

1. Pendahuluan

Indonesia merupakan Negara kepulauan, yakni berarti suatu Negara yang seluruhnya terdiri dari satu gugus besar atau lebih kepulauan dan dapat mencakup pulau-pulau lain. Indonesia terdiri dari lima pulau besar, untuk itu dibutuhkan sarana atau angkutan laut yang cukup memadai sebagai sarana transportasi khususnya pada bidang jasa pengiriman barang.

Dewasa ini semakin banyak berkembang perusahaan yang bergerak pada bidang jasa pengiriman barang untuk menjangkau pulau-pulau tersebut. Perusahaan transportasi merupakan sebuah perusahaan yang bergerak dibidang jasa pengiriman barang. Perusahaan ini selalu melayani pelanggannya ketika pelanggan tersebut membutuhkan jasa untuk pengiriman barang ke daerah atau wilayah lain atau pulau-pulau lain. Dalam melakukan pengiriman barang, perusahaan transportasi membutuhkan sarana yang dapat menjangkau wilayah-

wilayah pengiriman tujuan barang. Oleh karena itu, perusahaan ini melakukan kerja sama dengan perusahaan-perusahaan penyedia jasa transportasi baik darat, laut maupun udara.

Karena pada tahun-tahun sebelumnya, perusahaan tersebut selalu menggunakan cara konvensional dalam menggunakan alat transportasi terhadap semua pemesanan jasa pengiriman barang dan selalu menimbulkan masalah yaitu kendala kecelakaan atau kurang layaknya alat transportasi yang dapat menyebabkan barang hilang bahkan rusak.

Setelah melihat penerapan dan penggunaan perbandingan mitra penyedia jasa pengiriman barang yang ada saat ini, ternyata masih ditemukan beberapa masalah misalnya kerusakan dalam pengiriman juga masih ditemukan khususnya ketika menerima atau mengirim barang bahkan hilangnya barang saat dikirim.

Meskipun penyedia jasa pengiriman barang tersebut sudah memiliki banyak pengalaman, tetapi masih saja ada beberapa penyedia jasa pengiriman barang yang tidak konsisten dalam melakukan tugasnya tersebut secara baik dan benar. Pada umumnya alasan buruknya pelayanan adalah kurangnya pengetahuan terhadap koordinasi dalam pengiriman barang.

Melihat masalah-masalah yang ada maka dalam penelitian ini digunakan metode pengambilan keputusan dengan algoritma C4.5 untuk menentukan pemilihan mitra kerja penyedia jasa transportasi jasa pengiriman barang yang baik pada perusahaan transportasi di Jakarta.

Pemilihan mitra kerja penyedia jasa transportasi barang yang baik berdasarkan pengalaman kerja, harga jasa pengiriman, legalitas perusahaan mitra, jumlah armada yang dimiliki, dan intensitas pelayanan kerja mitra dalam melayani jasa pengiriman barang.

Pada penelitian ini pengambilan keputusan menggunakan menggunakan metode algoritma C4.5 dalam pengolahan data akan dilakukan menggunakan software pendukung yaitu Rapidminer versi 5.2.008.

Menerapkan algoritma data mining C4.5 pada data mitra kerja penyedia jasa transportasi. Pada tahapan ini dilakukan kajian mendalam terhadap data dari mitra kerja penyedia jasa transportasi, pengumpulan data

dilakukan pada sepuluh perusahaan yang menyediakan fasilitas pengiriman barang. Dilakukan pengolahan data, menyusun data yang diperoleh dalam bentuk table dan menggunakan algoritma sebagai prinsip pengolahan data.

Membuat model atau rule algoritma C4.5. Semua data yang sudah di cleaning, dikelompokkan sesuai dengan tahapan algoritma C4.5 disini akan ditentukan nilai gain dan nilai entropy dari seluruh data yang diolah dan akan menghasilkan model atau rule C4.5, rule yang dibuat akan di uji menggunakan data training sehingga dapat dimasukan data dari perusahaan untuk di uji dan dari hasil uji data penyedia mitra kerja perusahaan maka langsung akan diperoleh bahwa mitra tersebut layak atau tidak layak untuk menjadi mitra yang dibutuhkan oleh perusahaan.

Setiap lembaga atau perusahaan bekerja dengan data, perusahaan yang sudah cukup lama tentunya juga memiliki data yang banyak, data dalam perusahaan tersebut semakin lama akan semakin menggunung atau semakin menumpuk dan disimpan tiap harinya, tiap bulannya dan tiap tahunnya. Data tersebut merupakan aset berharga bagi perusahaan, sangat disayangkan jika data tersebut tidak digunakan sama sekali, untuk itu data yang jumlahnya sudah cukup banyak dan menggunung dapat digunakan kembali dengan cara menambangnya.

Data mining atau menambang data didefinisikan sebagai metode yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada database, ini adalah teknologi yang sangat potensial bagi perusahaan dalam memberdayakan data *warehouse*. Data mining disebut sebagai proses ekstraksi pengetahuan dari data yang besar, sesuai fungsinya data mining adalah proses pengambilan keputusan dari volume data yang besar yang disimpan dalam basis data, data warehouse, atau informasi yang disimpan dalam repositori [2]. Istilah data mining berasal dari kemiripan antara pencarian informasi yang bernilai dari database yang besar dengan menambang sebuah gunung untuk memperoleh sesuatu yang bernilai [8]. Data mining juga dapat didefinisikan sebagai sebuah proses penemuan pola dalam data [9]. Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang bernilai. Berdasarkan polanya data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering, asosiasi [5]. Banyak yang menggunakan data mining sebagai istilah populer dari Knowledge Discovery in Database (KDD). Data mining merupakan inti dari proses Knowledge Discovery in Database (KDD) [7]. KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah data set yang besar atau menggunung [6].

Langkah-langkah dalam KDD [5]:

1. Pembentukan pemahaman domain aplikasi. Pada tahap ini menentukan tujuan dari end-user dan bagian terkait dimana KDD dilakukan. Mengembangkan pemahaman tentang domain aplikasi ini adalah awal langkah persiapan. Ini

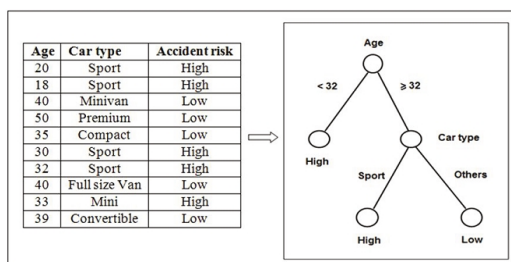
mempersiapkan adegan untuk memahami apa yang harus dilakukan dengan banyak keputusan (tentang transformasi, algoritma, representasi, dll). Orang-orang yang bertanggung jawab atas proyek KDD perlu memahami dan menentukan tujuan dari pengguna akhir dan lingkungan di mana pengetahuan penemuan proses akan berlangsung (termasuk pengetahuan awal yang relevan). Sebagai KDD hasil proses, mungkin ada revisi dan perbaikan dari langkah ini. Memiliki memahami tujuan KDD, preprocessing data dimulai, sebagaimana didefinisikan dalam berikutnya tiga langkah (perhatikan bahwa beberapa metode di sini mirip dengan algoritma Data Mining, tetapi digunakan dalam konteks preprocessing)

2. Memilih dan menciptakan satu data set untuk mendukung proses penemuan knowledge akan dilakukan. Penentuan data yang akan digunakan untuk proses KDD dilakukan pada tahap ini. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data untuk KDD ke dalam sebuah data set, termasuk atribut yang diperlukan dalam proses KDD. Terdapat interaktif dan iteratif dari KDD tersebut. Dimulai dengan data yang tersedia baik mengatur dan kemudian mengembang dan mengamati efeknya dalam KDD
3. Preprocessing dan cleansing. Dalam tahap ini kehandalan data ditingkatkan. Termasuk data clearing, seperti menangani data yang tidak lengkap, menghilangkan gangguan atau outlier. Termasuk menggunakan metode statistik yang kompleks, atau melakukan penambangan spesifik data dengan algoritma dalam KDD
4. Transformasi data. Pada tahap ini, generasi data yang lebih baik untuk data mining dipersiapkan dan dikembangkan, membuat data menjadi lebih baik menggunakan metode reduksi dimensi dan transformasi atribut. Sebagai contoh, dalam pemeriksaan medis, hasil bagi atribut mungkin sering menjadi faktor yang paling penting, dan tidak satu persatu. Di pemasaran, kita mungkin perlu mempertimbangkan efek di luar kendali kita serta upaya dan temporal isu (seperti mempelajari pengaruh akumulasi iklan). Namun, bahkan jika kita tidak menggunakan transformasi yang tepat di awal, kita dapat memperoleh efek mengejutkan bahwa petunjuk kepada kita tentang transformasi diperlukan (Pada iterasi berikutnya). Dengan demikian proses KDD mencerminkan kepada dirinya sendiri dan menyebabkan pemahaman tentang transformasi yang dibutuhkan (seperti pengetahuan ringkas dari sebuah ahli dalam bidang tertentu mengenai indikator terkemuka kunci).
5. Memilih tugas data mining yang cocok. Pada tahap ini ditentukan tipe data mining yang akan digunakan, apakah klasifikasi, regresi, atau

clustering, tergantung pada tujuan KDD dan tahap sebelumnya.

6. Memilih algoritma data mining. Pemilihan algoritma yang paling tepat untuk menemukan pola dilakukan pada tahap ini. Ada dua tujuan utama dalam Data Mining: prediksi dan deskripsi. Prediksi sering disebut sebagai Supervised Data Mining, sementara deskriptif Data Mining meliputi aspek-aspek unSupervised dan visualisasi Data Mining. Sebagian besar data teknik pertambangan didasarkan pada pembelajaran induktif, dimana model yang dibangun secara eksplisit maupun implisit oleh generalisasi dari jumlah yang memadai pelatihan data training. Asumsi yang mendasari pendekatan induktif adalah bahwa model terlatih ini berlaku untuk kasus masa depan. Strategi ini juga memperhitungkan tingkat meta-learning untuk set tertentu dari data yang tersedia.
7. Penggunaan algoritma data mining. Pada tahap ini dilakukan implementasi dari algoritma data mining yang telah ditentukan pada tahap sebelumnya. Misalnya dengan menseting parameter kontrol algoritma, seperti jumlah minimum kasus dalam daun tunggal dari pohon keputusan
8. Evaluasi. Pada tahap ini dilakukan evaluasi dan penerjemahan dari pola yang diperoleh, sehubungan dengan tujuan yang ditetapkan pada langkah pertama. Langkah ini berfokus pada komprehensibilitas dan kegunaan dari model induksi. Pada langkah ini pengetahuan ditemukan juga terdokumentasi untuk penggunaan lebih lanjut. Langkah terakhir adalah penggunaan dan umpan balik secara keseluruhan pada pola dan hasil penemuan diperoleh dengan Data Mining
9. Penggunaan pengetahuan yang ditemukan yakni memasukkan pengetahuan ke dalam sistem lain untuk ditindaklanjuti. Pengetahuan menjadi aktif dalam arti bahwa kita dapat membuat perubahan ke sistem dan mengukur dampak. Sebenarnya keberhasilan langkah ini menentukan efektivitas proses KDD secara keseluruhan. Ada banyak tantangan dalam langkah ini, seperti kehilangan "kondisi laboratorium". Misalnya, pengetahuan itu ditemukan dari sebuah snapshot statis tertentu (biasanya sampel) dari data, tapi sekarang data menjadi dinamis

Algoritma C4.5



Gambar 1. Skema Algoritma C4.5 [1]

Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang. Dan juga skalabilitas yaitu kemampuan untuk membangun model secara efisien untuk data berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir interpretabilitas yaitu model yang dihasilkan mudah dipahami. Selain itu dapat juga dikatakan bahwa Algoritma merupakan kumpulan perintah yang tertulis secara sistematis guna menyelesaikan permasalahan logika dari matematika. Pengertian Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Sedang pohon keputusan dapat diartikan suatu cara untuk memprediksi atau mengklarifikasi yang sangat kuat. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [4], yaitu:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus:

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

Keterangan:

- S = himpunan kasus
- n = jumlah partisi S
- p_i = proporsi S_i terhadap S

3. Kemudian hitung nilai gain menggunakan rumus:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots (2)$$

Keterangan:

- S = himpunan kasus
- A = fitur
- n = jumlah partisi atribut A

$$\left| \frac{S_i}{S} \right| = \text{proporsi } S_i \text{ terhadap } S$$

$$S = \text{jumlah kasus dalam } S$$

4. Ulangi langkah ke-2 hingga semua record terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua record dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam record yang dipartisi lagi.
 - c. Tidak ada record di dalam cabang yang kosong.

Penelitian Yang relevan:

1. Penelitian yang dilakukan oleh Lan Yu [6] dan kawan-kawan yang berjudul Application and Comparison of Classification Techniques in Controlling Credit Risk, membahas tentang komparasi teknik klasifikasi model statistik tradisional (LDA, QDA dan logistic regression), k-nearest neighbors, Bayesian networks (Naïve Bayes dan TAN), decision trees (C4.5), associative classification (CBA), neural network dan support vector machines (SVM), dan mengaplikasikannya ke dalam pengendalian credit risk. Eksperimen dilakukan pada 244 perusahaan terutama bidang industri dan bank komersial di Cina. Untuk verifikasi dan komparasi kinerja metode, digunakan kurva ROC dan metode Delong-Pearson. Dari pengukuran kinerja yang dilakukan, diketahui bahwa model statistik tradisional menghasilkan nilai paling kecil, C4.5 atau SVM kinerjanya kurang, dan CBA merupakan yang terbaik
2. Modeling Consumer Credit Scoring Through Bayes Network 2011 dilakukan oleh Abbas Heiat, membahas tentang prediksi dan komparasi kinerja dari metode data mining TAN Bayes, Markov Bayes, Dan Markov-FS dan mengidentifikasi prediktor untuk menentukan mana yang kredit yang bermasalah dan mana yang tidak bermasalah. Hasilnya, model Markov sedikit lebih baik dalam memprediksi dengan [11].
3. Komparasi Algoritma Klasifikasi Data Mining dalam Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor dilakukan oleh Henny Leidiyana komparasi algoritma C4.5, naïve bayes dan neural network yang diaplikasikan terhadap data konsumen yang mendapat pembiayaan kredit motor baik yang konsumen yang bermasalah maupun tidak. Dari hasil pengujian dengan mengukur kinerja ketiga algoritma tersebut menggunakan metode pengujian Cross Validation, Confusion Matrix dan Kurva ROC, diketahui bahwa algoritma C4.5 memiliki nilai accuracy dan AUC paling tinggi, diikuti oleh metode neural network, dan yang paling rendah adalah metode naïve bayes. [11].

Metode Penelitian

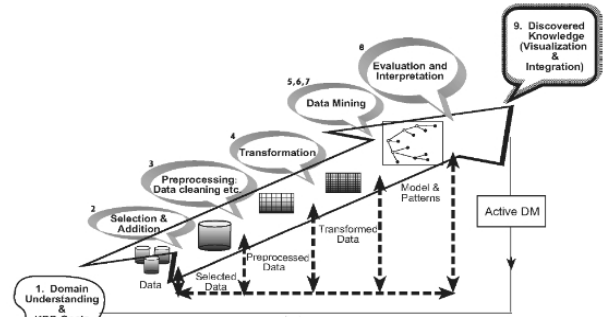
Obyek Penelitian ini adalah pada perusahaan pengiriman barang di Indonesia khususnya di DKI Jakarta. Dari perusahaan yang ada di DKI Jakarta penelitian ini akan mengambil sample pada 10 (sepuluh)

perusahaan di wilayah DKI Jakarta yang membutuhkan mitra kerja penyedia jasa transportasi.

Tempat dan Waktu Penelitian

Penelitian ini dilakukan pada perusahaan di wilayah DKI Jakarta dan berlangsung selama 2 (dua) tahun sejak tahun 2013 hingga tahun 2014.

Metode penelitian



Gambar 2. Proses KDD (Knowledge Discovery In Databases)

Penelitian dilakukan dengan menggunakan konsep dari data mining menggunakan Knowledge Discovery in Database (KDD) yakni sebuah kerangka kerja sebagai bahan acuan dalam pengkajian dibidang data mining [5].

2. Pembahasan

Berikut kriteria penggolongan data yang dilakukan terhadap perusahaan dalam menentukan mitra kerja penyedia jasa transportasi.

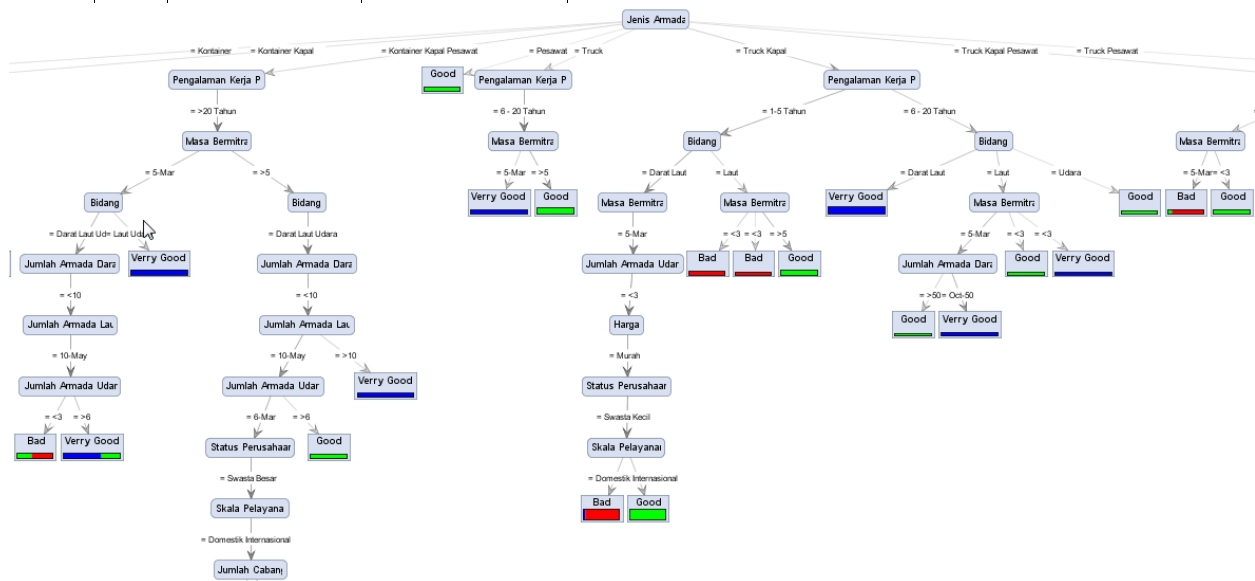
Tabel 1. Atribut dan Nilai Kategori

No.	Atribut	Nilai
1.	Pengalaman Kerja PT.	1-5 Tahun 6 – 20 Tahun > 20 tahun
2.	Masa Bermitra	< 3 tahun 3 - 5 tahun > 5 tahun
3.	Bidang	Darat Laut Udara Darat Laut Darat Udara Laut Udara Darat Laut Udara
4.	Jenis Armada	Kontainer Kontainer Kapal Kontainer Kapal Pesawat Truck Truck Kapal Pesawat

		Kapal Truck Kapal Pesawat
		Truck Pesawat
5.	Jumlah Armada Darat	< 10 10-50 > 50
6.	Jumlah Armada Laut	< 5 5-10 > 10
7.	Jumlah Armada Udara	< 3 3-6 > 6
8.	Harga Pelayanan	Murah Sedang Mahal
9.	Status Perusahaan	Swasta Kecil Swasta Sedang Swasta Besar

10.	Skala Pelayanan	Domestik Domestik Internasional Internasional
11.	Jumlah Cabang	< 5 Kota Wilayah 5 - 10 Kota Wilayah > 10 Kota Wilayah
12.	Remark	Very Good Good Bad

Setelah dilakukan pengolahan dan penyesuaian data berdasarkan algoritma C4.5 maka diperoleh data seperti pada gambar 3, jenis armada merupakan simpul tertinggi yang diperoleh dari nilai gain index kemudian simpul pengalaman kerja hingga selanjutnya remark *Verry Good*, *Good* dan *Bad*.



Gambar 3. Rule Pengolahan data dengan Algoritma C4.5

Dari pohon keputusan pada gambar 3 setelah diuji coba kembali menghasilkan tingkat akurasi yang cukup tinggi seperti pada Tabel 2 berikut ini:

Pred. Bad	7	6	76	85.39%
Class Recall	90.91%	94.12%	100.00%	

Tabel 2. Pengukuran Akurasi dengan Algoritma C4.5

accuracy: 94.37% +/- 4.40% (mikro: 94.38%)

	True Verry Good	True Good	True Bad	Class Precision
Pred. Verry Good	100	4	0	96.15%
Pred. Good	3	160	0	98.16%

Pada tabel 2 diperoleh nilai akurasi sebesar 94.37% melalui pengukuran tingkat akurasi dengan *confusion matrix* dengan *True positives* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam

confusion matrix, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity* (*recall*), *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif.

$$sensitivity = \frac{TP}{P} \dots\dots\dots(3)$$

$$specificity = \frac{TN}{N} \dots\dots\dots(4)$$

$$precision = \frac{TP}{TP+FP} \dots\dots\dots$$

$$\dots\dots\dots(5)$$

$$accuracy = sensitivity \frac{P}{(P+N)} + specificity \frac{N}{(P+N)} \dots\dots(6)$$

3. Kesimpulan

Penerapan algoritma C4.5 dilakukan pada pemilihan mitra kerja jasa transportasi untuk memilih mitra kerja yang baik, dari data tersebut terbentuk pohon keputusan seperti pada gambar 3. Tingkat akurasi yang dihasilkan dengan pengolahan data tersebut mencapai 94.37% dari gambar 6. Penerapan algoritma C4.5 dalam memilih mitra kerja dilakukan untuk mencari mitra kerja terbaik dan terpercaya. Penelitian ini baru memasuki tahap 1 dan akan terus dikembangkan pada tahap 2.

Daftar Pustaka

[1] Contoh Kasus Prediksi layak tidaknya Indonesian menjadi Juara AFF, <http://dennistorio.com/DM/bola.php> diakses tanggal 29 juni 2014.

[2] Han, J., & Kamber, M. 2006. Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman.

[3] Karismariyanti, Magdalena. 2011. Jurnal: Simulasi Pendukung Keputusan Beasiswa menggunakan Metode Composite Performance Index.

[4] Kusriani & Luthfi, E. T. 2009. Algoritma Data Mining. Yogyakarta: Andi Publishing.

[5] Larose, D. T. 2005. Discovering Knowledge in Data. New Jersey: John Willey & Sons, Inc.

[6] Liao. 2007. Recent Advances in Data Mining of Enterprise Data: Algorithms and Application. Singapore: World Scientific Publishing

[7] Maimon, Oded & Rokach, Lior. 2005. Data Mining and Knowledge Discovey Handbook. New York: Springer

[8] Sumathi, & S., Sivanandam, S.N. 2006. Introduction to Data Mining and its Applications. Berlin Heidelberg New York: Springer

[9] Witten, I. H., Frank, E., & Hall, M. A. 2011. Data Mining: Practical Machine Learning and Tools. Burlington: Morgan Kaufmann Publisher.

[10] Wu, Xindong & Kumar, Vipin. 2009. The Top Ten Algorithms in Data Mining. Boca Raton: CRC Press

[11] Heiat, Abbas. (2011). Modeling Consumer Credit Scoring Through Bayes Network. World Journal of Social Sciences, Volume 1, Number 3, pp.132-141, July 2011

[12] Leidiyana, Henny. (2011). Tesis: Komparasi Algoritma Data Mining Dalam Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. Juli 2011.

Biodata Penulis

Harry Dhika, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Indraprasta PGRI, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika STMIK Nusa Mandiri, lulus tahun 2012. Saat ini menjadi Dosen di Univ. Indraprassta PGRI.

Tri Yani Akhirina, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Indraprasta PGRI, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika STMIK Nusa Mandiri, lulus tahun 2011. Saat ini menjadi Dosen di Univ. Indraprassta PGRI.

Surajiyo, memperoleh gelar Sarjana Sosial (S.Pd), Jurusan Ilmu Filsafat, Universitas Gadjah Mada, lulus tahun 1987. Memperoleh gelar Magister Komunikasi (M.Si) Program Pasca Sarjana Magister Ilmu Komunikasi, lulus tahun 2009. Saat ini menjadi Dosen di Univ. Indraprassta PGRI.