



Prediksi Cabang Primagama Bermasalah Menggunakan Algoritma C4.5

Andin Eka Safitri^{1*}

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: dosen01122@unpam.ac.id

(* : coresponding author)

Abstrak– Primagama adalah salah satu situs bimbingan belajar terbesar di Indonesia dengan ratusan cabang di seluruh Indonesia. Karena banyaknya cabang, tim audit harus menginvestigasi semua cabang dan mengidentifikasi cabang mana yang bermasalah. Cabang bermasalah adalah hal yang harus dihindari oleh perusahaan demi mencapai target keuntungan perusahaan serta menjaga agar cabang selalu taat peraturan yang dibuat perusahaan. Akan tetapi cabang bermasalah masih banyak, dan perlu waktu untuk tim audit memilih cabang dari masalah ada. Penelitian ini menerapkan C4.5 untuk mencari pohon keputusan serta nilai akurasi dari dataset yang didapat. Hasil dari akurasi C4.5 dalam penelitian ini mencapai 98,60%. Dengan hasil akurasi tersebut diharapkan dapat dibuat kebijakan-kebijakan baru terkait prediksi cabang bermasalah.

Kata Kunci: Bimbingan Belajar, Algoritma C4.5, Decision Tree, Data Mining

Abstract– Primagama is one of indonesia largest site preparation hundreds of branches across the country. The number of branches, the audit should investigate the branches and identify troubled bermasalah.cabang branches which are to be avoided by companies to reach its profit goal the company and preventing the regulation made the turn. But the problem is a lot, and it takes time to choose a branch of the audit problems there. This study apply c4.5 to find the decision and the accuracy of a unique dataset obtained. The result of accuracy c4.5 % 98,60 reached in this research. The accuracy is expected to made new predictions kebijakan-kebijakan related to the problem.

Keywords: Tutoring, Algorithm C4.5, Decision Tree, Data Mining

1. PENDAHULUAN

Hingga saat ini tercatat tidak kurang dari 630 cabang atau franchise Primagama, dan yang aktif beroperasi 443 cabang (staff 2017). Semakin banyak cabang yang dimiliki, semakin banyak juga data yang didapatkan, sedangkan keberhasilan dari suatu perusahaan franchise sangat ditentukan dari perkembangan cabang. Cabang yang memiliki kinerja yang baik, dari sisi marketing serta finance dapat menunjang tercapainya tujuan perusahaan.

Kinerja sendiri adalah hasil kerja secara kualitas dan kuantitas yang dicapai oleh cabang dalam melaksanakan tugasnya sesuai dengan tanggung jawab yang diberikan kepadanya. Kinerja yang baik adalah kinerja yang optimal, yaitu kinerja yang sesuai standar organisasi dan mendukung tercapainya tujuan organisasi. Terdapat faktor negatif yang dapat menurunkan kinerja cabang, diantaranya adalah menurunnya jumlah siswa dikarenakan tempat yang tidak strategis atau marketing yang kurang, dan tidak patuh pada sistem yang ada.

PT Prima Edu Pendamping Belajar membagi cabang berdasarkan area, ada 11 area yang dimiliki, Jakarta 1 terdiri dari 67 cabang, Jakarta 2 terdiri dari 38 cabang, Sumatra 1 terdiri dari 32 cabang, Sumatra 2 terdiri dari 37 siswa, Jawa Barat terdiri dari 40 cabang, Jawa Tengah 1 terdiri dari 36 cabang, Jawa Tengah 2 terdiri dari 33 cabang, Jawa Timur 1 terdiri dari 90 cabang, Jawa Timur 2 terdiri dari 25 cabang, Bali Nusra terdiri dari 19 cabang, dan Kalimantan terdiri dari 19 cabang. Area-area tersebut dipimpin oleh 1 manager area, yang tidak setiap hari mengunjungi cabang untuk mengontrol, dan tidak semua cabang terkontrol dengan baik oleh manager area dikarenakan komunikasi atau jarak cabang yang cukup jauh.

Decision Tree adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (leaf) mempresentasikan kelompok kelas tertentu. Level node tertas dari sebuah Decision Tree adalah node akar (root) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya Decision Tree melakukan strategi pencarian secara top-down untuk solusinya.

Algoritma C4.5 merupakan generasi baru dari algoritma ID3 yang dikembangkan. Untuk membuat sebuah pohon keputusan, algoritma ini dimulai dengan memisahkan training samples kedalam sampel akar pada pohon keputusan. Training samples adalah sampel yang digunakan untuk membangun model classifier dalam hal ini pohon keputusan. Kemudian sebuah atribut dipilih untuk mempartisi sampel ini. Untuk tiap nilai yang dimiliki atribut ini, sebuah cabang dibentuk. Setelah cabang terbentuk maka subset dari himpunan data yang atributnya memiliki nilai yang sesuai dengan cabang tersebut dimasukkan ke dalam sampel yang baru.

2. METODOLOGI PENELITIAN

Berdasarkan sumber daya yang digunakan dalam penelitian ini berupa data cabang dan siswa. Karena data dan informasinya diperoleh untuk memperkuat analisis data juga dilakukan wawancara dengan pihak-pihak yang terkait yaitu tim audit, dimana data tersebut diperoleh dengan teknik pengumpulan data yaitu wawancara, observasi dan studi pustaka.

Setelah data yang diperlukan diperoleh, kemudian ditentukan variable-variabel yang akan digunakan dalam penelitian. Variabel respon yang digunakan dalam penelitian ini adalah kinerja, cabang yang disiplin dan tanggung jawab terhadap penginputan data ke sistem primaedu.

2.1 Analisa Kebutuhan

Analisa kebutuhan sistem pada tugas akhir ini merupakan tahapan yang sangat vital mengingat cakupan yang harus ada dalam sistem ini komplek serta ada banyak hal yang harus diperhatikan dalam menentukan kebutuhan sistem itu sendiri. Data dan informasi yang diperlukan tersebut diperoleh dari berbagai sumber terkait untuk memberikan masukan yang lengkap bagi pengembangan sistem informasi ini. Data dan informasi tersebut antara lain:

a. Mengumpulkan Data Mengenai Sistem Rekam Medis dan Tipe Penyedia Layanan Kesehatan

Pada tahap ini dilakukan pengumpulan data-data tertulis yang berkaitan dengan proses yang ada dan dokumen-dokumen yang berkaitan dengan cabang serta hasil data dari divisi audit. Selanjutnya data-data yang diperoleh akan digunakan ke dalam rancangan sistem untuk diolah.

b. Study literatur, Penelitian, dan Teori yang Berkaitan dengan Obyek Penelitian

Tahap pengumpulan pustaka dipergunakan untuk memperkuat dasar teori guna mendukung pengerjaan dan penelitian tugas akhir, untuk itu di dalam pencarian pustaka dipergunakan:

1. Jurnal
Paper-paper yang diterbitkan oleh suatu organisasi nasional maupun internasional dalam bentuk majalah fisik maupun elektronik. Paper-paper yang diambil adalah yang berhubungan dengan algoritma C4.5
2. Prosiding
Paper-paper yang dibutuhkan adalah paper-paper yang diterbitkan di suatu seminar dalam bentuk buku yang berhubungan dengan bidang ilmu kesehatan khususnya rekam medis, rekam medis baik manual maupun elektronik, perancangan sistem informasi dan penggunaan basis data terpusat.
3. Text Book
Text book yang dibutuhkan adalah materi-materi yang berhubungan dengan bidang sistem rekam medis manual maupun elektronik dan perancangan sistem informasi.
- c. **Menentukan Kebutuhan dari Sistem Informasi Rekam kesehatan Elektronik Terpusat**

Tahap ini dilakukan sebelum melakukan pemodelan yaitu menentukan kebutuhan rancangan penelitian sesuai dengan masing-masing cabang. Data-data yang dibutuhkan untuk penelitian antara lain:

1. Data pembayaran franchise
2. Data pembayaran siswa
3. Jumlah siswa

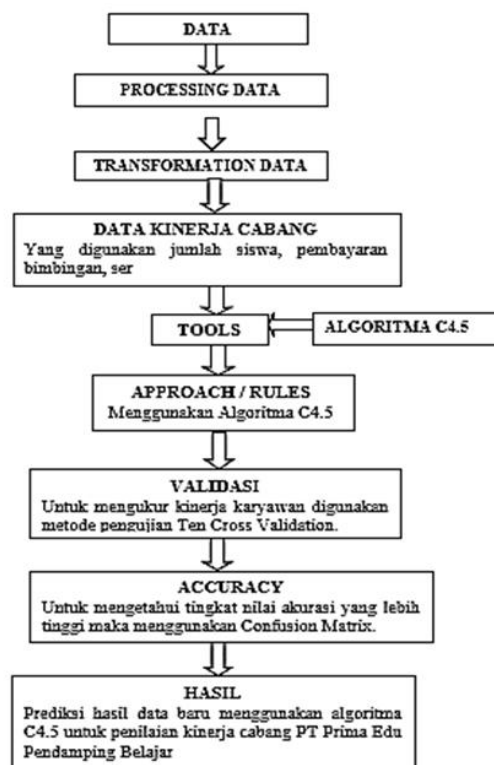
Setelah itu berdasarkan data yang ada akan diproses sehingga menghasilkan output yang berguna untuk divisi audit dalam mengetahui kondisi cabang.

2.2 Perancangan Penelitian

Langkah-langkah yang dilakukan untuk menganalisis penilaian kinerja cabang ialah :

1. Prapengolahan data kinerja cabang
2. Menghitung data secara manual data tersebut untuk menentukan nilai kinerja yang baik dan nilai yang kurang baik pada cabang.
3. Hasil dari perhitungan tersebut dimasukan kedalam rumus Igoritma C4.5 kemudian dicari nilai gain yang tertinggi dan cabangnya untuk membuat pohon keputusan.
4. Dalam pengujian data peneliti menggunakan aplikasi Rapidminer
5. Kemudian hasil dari analisa tersebut dapat digunakan untuk mengetahui sejauh mana kinerja cabang yang berpotensi bermasalah.

Pada gambar 1 dibawah ini merupakan usulan kerangka kerja dari penelitian ini untuk menerapkan model serta data yang gunakan serta mengukur sebuah kinerja terhadap suatu model.



Gambar 1. Kerangka Pemikiran

Pengolahan data cabang dengan menentukan atribut-atribut apa saja yang akan digunakan untuk proses perhitungan yang akan digunakan untuk menentukan kondisi cabang.

Tabel 1. Atribut Data Penelitian

Kode Cabang	Atribut yang menginformasikan no identitas cabang
Nama Cabang	Atribut yang menginformasikan nama-nama cabang Primagama
Jumlah Siswa	Atribut yang menginformasikan berapa jumlah siswa pada cabang
Total Biaya Bimbingan	Atribut yang menginformasikan keseluruhan biaya bimbingan pada cabang
Total Pembayaran	Atribut yang menginformasikan keseluruhan pembayaran biaya bimbingan yang dilakukan oleh siswa per cabang
Sisa Hutang	Atribut yang menginformasikan hutang siswa ke cabang
Siswa MD	Atribut yang menginformasikan siswa yang mengundurkan diri

Tanggal Aktif Cabang	Atribut yang menginformasikan Tanggal aktif cabang sesuai dengan berlakunya MOU
Biaya Franchise	Atribut yang menginformasikan biaya franchise pada masing-masing cabang
Pembayaran Franchise	Atribut yang menginformasikan pembayaran yang dilakukan cabang untuk franchise

2.3 Teknik Analisa

Langkah-langkah yang dilakukan untuk menganalisis penilaian kinerja cabang ialah :

Pada teknik analisa akan mencari nilai akurasi, sensitivitas dan spesifitas. Sebelum mendapatkan nilai tersebut pada penelitian ini dibagi menjadi enam tahapan, yaitu:

1. Domain Understanding

Pada tahap ini adalah mencari persoalan-persoalan yang berkaitan dengan klasifikasi cabang yang dilanjutkan dengan menentukan rumusan masalah serta menganalisa kebutuhan-kebutuhan yang dapat digunakan untuk memecahkan masalah tersebut.

2. Selection and Addition

Pada tahap Selection and Addition data yang didapatkan dari database PT Primagama Pendamping Belajar dianalisa agar dapat ditentukan atribut-atribut apa saja yang akan diperlukan berdasarkan analisa kebutuhan. Adapun atribut data set yang digunakan berada pada tabel 3.3

3. Preprocessing Data Cleaning

Pada tahap ini akan dilakukan pembersihan data

4. Transformation

Data yang digunakan dalam penelitian ini sebagian besar masih berupa numerik sehingga data ditransformasikan ke dalam klasifikasi bilangan polynomia, serta dibuat kedalam beberapa sampling. Nilai asli dapat dilihat pada tabel 3.12 selanjutnya nilai yang telah dikonversi dapat dilihat pada tabel 2.

Tabel 2. Data Penelitian

Atribut	Nilai Angka	Kategori
Jumlah Siswa	< 100	Sedikit
	> 100	Banyak
Total Biaya Bimbingan	$\geq 521.000.000$	Sedikit
	$\leq 521.000.001$	Banyak
Sisa Hutang	0	Lunas
	$\leq 50\%$ Biaya Bimbingan	Normal
	$> 50\%$ Biaya Bimbingan	Sangat Besar
Total Pembayaran	= Total Biaya Bimbingan	Lunas
	< Total Biaya Bimbingan	Kurang
Siswa MD	< 15%	Normal
	> 15%	Banyak
Biaya Franchise	$> 100.000.000$	Cabang Besar
	$< 100.000.000$	Cabang Kecil
Pembayaran Franchise	< biaya franchise	Kurang Bayar
	= biaya franchise	Lunas

5. Evaluation and Interpretation

Tahap ini adalah tahap pengujian terhadap data training dan data testig serta mencari nilai akurasi C4.5.

3. ANALISA DAN PEMBAHASAN

Penelitian ini dilakukan di PT Prima Edu Pendamping Belajar, yang mempunyai merk PRIMAGAMA. Data dan informasi diperoleh untuk memperkuat analisis data juga dilakukan wawancara dengan pihak-pihak yang terkait yaitu tim audit, dimana data tersebut diperoleh dengan teknik pengumpulan data yaitu wawancara, observasi dan studi pustaka. Selanjutnya data dimasukan atau diolah, dengan mengambil Jumlah Siswa, Total Biaya Bimbingan, Sisa Hutang, Total Pembayaran, Siswa Mengundurkan Diri, Biaya Franchise, Pembayaran Franchise. Adapun hasil *preprocessing* data akumulasi yang didapat adalah sebagai berikut:

Tabel 3. Hasil *Preprocessing* Data Akumulasi

Jumlah Siswa	Total Biaya Bimbingan	Sisa Hutang	Total Pembayaran	Siswa MD	Biaya Franchise	Pembayaran Franchise
Banyak	Sedikit	Normal	Kurang	Normal	Cabang Kecil	Kurang bayar
Banyak	Banyak	Normal	Kurang	Normal	Cabang Besar	Kurang bayar
Sedikit	Sedikit	Normal	Kurang	Normal	Cabang Besar	Kurang bayar
Banyak	Banyak	Banyak	Kurang	Normal	Cabang Kecil	Kurang bayar
Banyak	Banyak	Normal	Kurang	Normal	Cabang Kecil	Lunas
Banyak	Sedikit	Normal	Kurang	Normal	Cabang Besar	Kurang bayar
Banyak	Banyak	Banyak	Kurang	Normal	Cabang Kecil	Kurang bayar
Sedikit	Sedikit	Banyak	Kurang	Normal	Cabang Kecil	Lunas
Banyak	Sedikit	Banyak	Kurang	Normal	Cabang Kecil	Lunas
Banyak	Banyak	Banyak	Kurang	Normal	Cabang Kecil	Kurang bayar
Banyak	Sedikit	Normal	Kurang	Normal	Cabang Kecil	Kurang bayar
Banyak	Banyak	Banyak	Kurang	Normal	Cabang Besar	Kurang bayar
Sedikit	Sedikit	Banyak	Kurang	Normal	Cabang Besar	Kurang bayar
Banyak	Sedikit	Banyak	Kurang	Normal	Cabang Besar	Lunas

Pada tahapan algoritma *decision tree* C4.5 memiliki tahapan pencarian *entropy* dan *gain*. Berikut adalah contoh pembuktian perhitungan *entropy* dan *gain* pada *node* ke-1.

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

1. Entropy Total

$$Entropy(Total) = \left(-\left(\frac{380}{833}\right) \cdot \log_2\left(\frac{380}{833}\right) \right) + \left(-\left(\frac{453}{833}\right) \cdot \log_2\left(\frac{453}{833}\right) \right) \\ = 0,994453$$

accuracy: 98.60% +/- 1.02% (mikro: 98.60%)

	true Cabang Bermasalah	true Cabang Sehat	class precision
pred. Cabang Bermasalah	224	2	99.12%
pred. Cabang Sehat	5	268	98.17%
class recall	97.82%	99.26%	

Gambar 3. Hasil Perhitungan Dengan Menggunakan *Software Rapidminer*

5. KESIMPULAN

Berdasarkan hasil perancangan, implementasi, dan pengujian yang dilakukan, dapat diambil kesimpulan bahwa Implementasi algoritma *C4.5* untuk prediksi cabang bermasalah dimulai dengan melakukan analisis kebutuhan data kemudian dilakukan penelitian, setelah didapatkan data-data yang diperlukan kemudian dilakukan pembuatan sistem dengan mengimplementasikan algoritma *C4.5* kedalam sistem tersebut dan melakukan pengujian terhadap beberapa sampel data untuk mendapatkan akurasi dari hasil prediksi yang dilakukan oleh sistem yang telah dibuat.

Berdasarkan hasil pengujian akurasi dari pengujian menghasilkan tingkat akurasi 98,60%.

REFERENCES

- Adriana, I. A. (2007). *Penalaran Komputer Berbasis Kasus (Case Based Reasoning)*. Yogyakarta: Ardana Media.
- Al-Kabi, M. N., Kanaan, G., Al-Shalabi, R., Al-Sinjilawi, S., & Al-Mustafa, R. (2005). Al-Hadith Text Classifier. *Journal of Applied Sciences* 5, 584-587.
- Andina Kusumaningrum, Said Al-Faraby, & Adiwijaya. (2017). Klasifikasi Informasi, Anjuran dan Larangan pada Hadits Shahih Bukhari menggunakan Metode Support Vector Machine. *e-Proceeding of Engineering*, 5014.
- Asriyanti Indah Pratiwi, & Adiwijaya. (2018). On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis. *Applied Computational Intelligence and Soft Computing*, 2018, 5. doi:https://doi.org/10.1155/2018/1407817
- Desiani, A., & Arhami, M. (2006). *Konsep Kecerdasan Buatan* (1 ed.). (D. Hardjono, Penyunt.) Yogyakarta, Indonesia: Penerbit ANDI.
- Desiani, A., & Arhami, M. (2006). *Konsep Kecerdasan Buatan* (1 ed.). (D. Hardjono, Penyunt.) Yogyakarta, Indonesia: Penerbit ANDI.
- Eliza Riviera Rachmawati Jasin, Said Al-Faraby, & Adiwijaya. (2017). Klasifikasi Anjuran, Larangan dan Informasi pada Hadis Sahih Al-Bukhari. *e-Proceeding of Engineering*, 4683.
- Faza Akmal, S. W. (Februari 2014). SISTEM PPAKAR UNTUK MENDIAGNOSA PENYAKIT LAMBUNG DENGAN IMPLEMENTASI METODE CBR (CASE BASED REASONING) BERBASIS WEB. *Jurnal Sarjana Teknik Informatika*, 2 (1).
- Haza Nuzly Abdull Hamed, Siti Mariyam Shamsuddin, & Naomie Salim. (2008). Particle Swarm Optimization For Neural Network Learning Enhancement. *Jurnal Teknologi*, 13-26.
- Kusrini. (2006). *Sistem Pakar Teori dan Aplikasi* (1 ed.). (f. Suyantoro, Penyunt.) Yogyakarta, Indonesia: Penerbit ANDI.
- M.L. Zhang, & Z. H. Zhou. (2006). Multilabel neural networks with applications to functional genomics and text. *IEEE transactions on Knowledge and Data Engineering*, 1338-1351.
- Nurchahyo, S. a. (2014). Rainfall Prediction in Kemayoran Jakarta Using Hybrid Genetic Algorithm (GA) and Partially Connected Feedforward Neural Network (PCFNN). *Information and Communication Technology (ICoICT)*, (pp. 166-171).
- Raharjo, J. S. (2013). Model Artificial Neural Network berbasis Particle Swarm Optimization untuk Prediksi Laju Inflasi. *Sistem Komputer*.
- Reynaldi Ananda Pane, M. S. (2018). A Multi-lable Classification on Topics of Quranic Verses in English Translation using Multinomial Naive Bayes. *6th International Conference on Information and Communication Technology (ICoICT)*.
- staff, I. (2017). *Cabang Aktif*. Jakarta: Prmagama.
- T.Sutojo, E. M. (2011). *Kecerdasan Buatan*. Yogyakarta: Andi.