



Perbandingan Implementasi Algoritma CT-PRO dan Algoritma C45 Untuk Menentukan Pola Nasabah

Ade Sukma Wati¹, Pipin Octavia², Azhar Andika Putra³, Mariana Purba⁴, M. Bayu Wibisono^{5*}

^{1,2,3,4,5} Sistem Informasi, Fakultas Ilmu Komputer

^{1,2,3,4} Universitas Sjakhyakirti Palembang

⁵ Universitas Pembangunan Nasional Veteran Jakarta

email: ^{1,2,3,4}fasillomsjakhyakirti@gmail.com, ⁵bayu.wibisono@upnvj.ac.id

^{1,2,3,4}Jl. Sultan Muhammad Mansyur Kb. Gede, Kota Palembang

⁵Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

Abstrak

Bank merupakan salah satu instansi pemerintah atau perusahaan untuk mengelola keuangan. Sebagai upaya untuk memberikan kinerja yang baik kepada nasabah diperlukan tata kerja yang tertib, rapi dan teliti sehingga akan menghasilkan informasi yang cepat, akurat dan tepat waktu sesuai kebutuhan. Dalam instansi bank banyak data yang setiap tahunnya bertambah. Salah satunya yaitu data nasabah baru. Akan tetapi dengan sekian banyaknya data nasabah baru maka semakin sulit juga data pola nasabah baru tersebut dipelajari lebih lanjut dan umumnya hanya digunakan sebagai arsip saja. Pada penelitian ini, telah diterapkan metode Algoritma CT-PRO serta Algoritma c45 suatu studi kasus, yaitu kasus menentukan "Pola Nasabah". menentukan pola nasabah dari ratusan atau ribuan field. Dari hasil implementasi yang dilakukan, ada sebuah aplikasi yang dapat menerapkan algoritma CT-PRO dan algoritma C45 pada kasus tersebut.

Kata kunci: Pola Nasabah, Algoritma CT Pro, Algoritma C45

1. PENDAHULUAN

Bank merupakan salah satu instansi pemerintah atau perusahaan untuk mengelola keuangan. Sebagai upaya untuk memberikan kinerja yang baik kepada nasabah diperlukan tata kerja yang tertib, rapid an teliti sehingga akan menghasilkan informasi yang cepat, akurat dan tepat waktu sesuai kebutuhan. Dalam instansi bank banyak data yang setiap tahunnya bertambah. Salah satunya yaitu data nasabah baru. Akan tetapi dengan sekian banyaknya data nasabah baru maka semakin sulit juga data pola nasabah baru tersebut dipelajari lebih lanjut dan umumnya hanya digunakan sebagai arsip saja.



Pada penelitian ini, telah diterapkan metode Algoritma CT-PRO serta Algoritma C45 suatu studi kasus, yaitu kasus menentukan “Pola Nasabah”. menentukan pola nasabah dari ratusan atau ribuan field. Dari hasil implementasi yang dilakukan, ada sebuah aplikasi yang dapat menerapkan algoritma CT-PRO dan algoritma C45 pada kasus tersebut. Akan diuji dengan CT-PRO dan algoritma C45 dengan hasil dapat mengimplementasikan kaduah buah algoritma dengan benar. Sehingga untuk menentukan algoritma terbaik yang akan dipakai disebuah kasus, harus melihat kriteria, variabel maupun jumlah data dikasus tersebut.

Dalam permasalahan yang ada di penelitian ini untuk mengetahui perbandingan antara algoritma CT-Pro dan algoritma C45 dalam menentukan pola nasabah distudi kasus salah satu bank di Palembang. Dalam perbandingan tersebut memerlukan suatu system yang bisa memandingkan dua algoritma tersebut. Dimana Algoritma CT-PRO merupakan pengembangan suatu informasi dari sebuah struktur yang lebih kecil atau ringan, sehingga baik pembentukan menjadi lebih cepat. Algoritma C45 adalah salah satu metode algoritma yang digunakan oleh beberapa penelitian guna membuat suatu pohon keputusan.

2. METODOLOGI PENELITIAN

2.1 Data Penelitian

Untuk mendapatkan data penelitian, maka penulis mengumpulkan data-data berupa data nasabah pada tahun 2020-2021 dan data nasabah yang masih aktif. Data yang di ambil ini merupakan data yang memenuhi kriteria untuk di set. Data yang digunakan seperti data ID Nasabah / No Nasabah, Jenis Kelamin, Status, Umur, Penghasilan, Angsuran/Bulan dan Label. Data itu sendiri diambil di bagian administrasi Bank. Data yang didapatkan sebanyak 299 data dan yang digunakan sebanyak 100 data.

2.2 Pengolahan Data Mining

Adapun pengolahan *data mining*, mengikuti tahapan dalam *Knowledge Discovery in Database* (KDD), sebagai berikut (Kusmini & Taufiq, 2009) :

1. *Data Selection* : Data-data nasabah yang masih aktif yaitu tahun 2018 - 2019 pada Bank Palembang, yang digunakan sebanyak 100 data. Untuk data Training dan Testing yaitu data nasabah yang masih aktif. Data yang diambil merupakan nasabah yang memiliki angsuran 1000000/bulan. Data yang dipilih merupakan data nasabah dengan angsuran kategori tinggi karena dianggap layak.
2. Pra Pemrosesan Data : Data yang digunakan berasal dari data nasabah 2018-2019 sebagai data set yang akan di implementasikan menggunakan aplikasi Rapid Miner dengan metode C4.5 dan CT-Pro. Dari proses pengumpulan data didapatkan *record* sebanyak 100 data nasabah yang lancar dan macet dalam mengangsur.



3. Pembersihan data, untuk mengidentifikasi dan menghapus data yang tidak konsisten dan data yang tidak lengkap. Dalam penelitian ini record nasabah dengan status non aktif dihapuskan karena tidak mengandung nilai untuk dilabeli.
4. Reduksi data, Dalam penelitian ini atribut Nama dan jenis kelamin tidak digunakan dalam proses mining. Data yang di butuhkan yaitu data ID Nasabah atau No Nasabah, Penghasilan, Angsuran, dan Label.

Tabel 1. Attribute Data Set Klasifikasi Label Angsuran

1	ID Nasabah	(No Urut)
2	Status	Menikah / Belum Menikah
3	Umur	<50 Tahun dan > 50 Tahun
4	Penghasilan	Rendah : <2500000 Sedang : >2500000 Tinggi : >4500000
5	Angsuran/Bulan	Rendah : <1000000 Tinggi : >1000000
6	Label	Lancar Macet

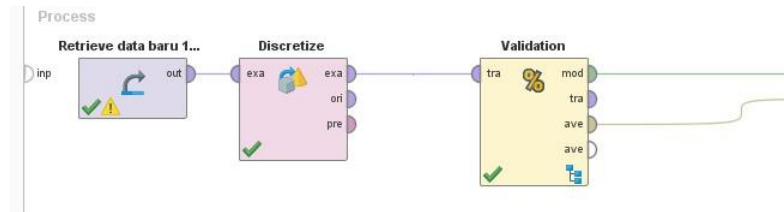
5. *Transformation* : Pada tahapan ini yaitu merubah data kedalam bentuk yang sesuai di data mining. Teknik data transformation yaitu descritize yang digunakan untuk merubah data dengan tipe continuous menjadi data dengan tipe diskrit. *Descrctize* adalah salah satu metode yang digunakan saat *preprocessing input*. Teknik diskritisasi data dapat digunakan untuk mengurangi jumlah nilai *atribute numeric* dengan cara membagi *range* atribut ke dalam interval. Label interval kemudian dapat digunakan untuk menggantikan nilai-nilai aktual data.
6. *Data Mining* : Setelah melakukan proses *preprocessing* dan transformasi data yang sesuai untuk penggunaan teknik data mining klasifikasi dengan algoritma C4.5 dan CT-Pro maka tahapan selanjutnya adalah melakukan proses analisis data mining yang telah dikerjakan pada tahap sebelumnya. Proses analisis *data mining* ini hanya dilakukan untuk mempermudah dan mengertipenggunaan teknik *klasifikasi* dengan metode algoritma C4.5 dan CT- Pro menggunakan aplikasi *Rapid Miner*.

3 HASIL DAN PEMBAHASAN

3.1 Implementasi Metode C4.5 dengan RapidMiner

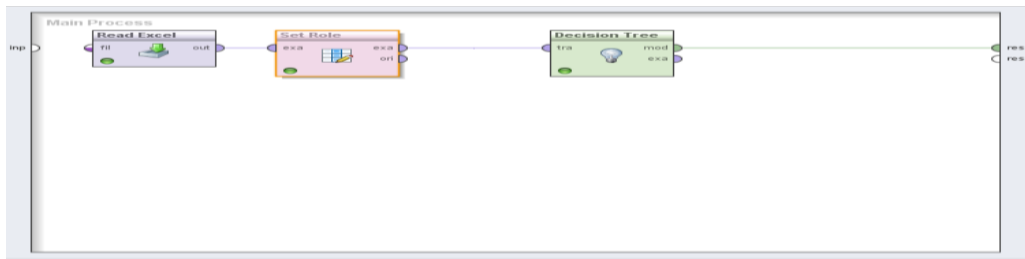
Penelitian ini di lakukan untuk mengetahui klasifikasi pola nasabah. Penggunaan parameter berpengaruh pada hasil akurasi dan model yang akan dihasilkan oleh algoritma C4.5 dan CT-Pro. Adapun tampilan dari struktur model data mining yang di gunakan terdiri dari beberapa objek atau operator antara lain read file XLSX, selected attribute discritisasi dan Cross Validation yang tampil pada proses utama untuk menjalankan proses data

mining. Untuk menentukan akurasi dari algoritma, penelitian ini menggunakan Cross Validation. Desain model C4.5 yang terdapat pada RapidMiner yang ditunjukkan pada gambar 1.



Gambar 1. Cross Validation

Pada objek Discretize attribute yang dipilih yaitu subnet yang memiliki nilai continuous saja seperti label lancar atau macet. Pada objek atau operator Cross Validation terdapat sub proses yang terdiri dari operator Decision Tree sebagai algoritma yang digunakan pada proses data mining Classification yang akan dilakukan, operator Apply Model dan operator Performance sebagai operator untuk menghasilkan pengolahan data mining berupa Decision Tree. Adapun model sub proses dapat di lihat pada gambar berikut:

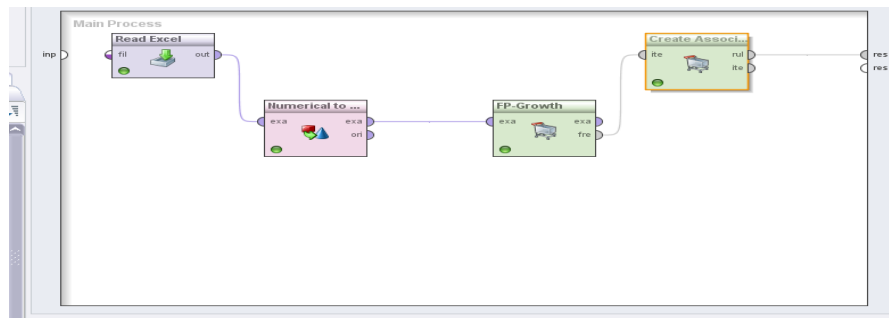


Gambar 2. Rapidminer Model Sub Proses Clasification Decision Tree Salah

satu tujuan penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma Decision tree yang digunakan untuk mengklasifikasi Label. Didalam kolom training terdapat algoritma klasifikasi yang diterapkan yaitu decision tree, sedangkan didalam kolom set role terdapat pengauran lbel yang digunakan untuk menjalankan model decision tree dan Performance untuk mengukur performa dari model decision Tree Tersebut.

3.2 Implementasi Metode CT-Pro dengan RapidMiner

Pada implementasi dengan algoritma CT-Pro, adapun tampilan dari struktur model data mining yang di gunakan terdiri dari beberapa objek atau operator antara lain read file XLSX, selected attribute dan Cross Validation yang tampil pada proses utama untuk menjalankan proses data mining. Untuk menentukan akurasi dari algoritma, penelitian ini menggunakan Assosiation Desain model algoritma CT-Pro yang terdapat pada Rapid Miner.

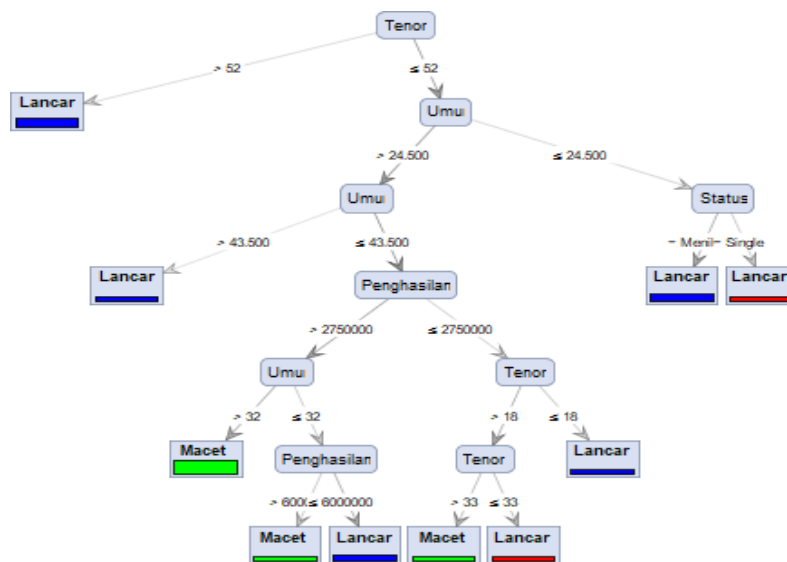


Gambar 3. Rapidminer Model Sub Proses KlasifikasiCT-Pro

Pada objek atau operator Cross Validation terdapat sub proses yang terdiri dari operator algoritma CT-Pro sebagai algoritma yang digunakan pada proses data mining Classification yang akan dilakukan, operator FP Growth dan operator Association sebagai operator untuk menghasilkan pengolahan data mining berupa algoritma CT-Pro.

3.3 Hasil Implementasi Klasifikasi Menggunakan Metode C4.5

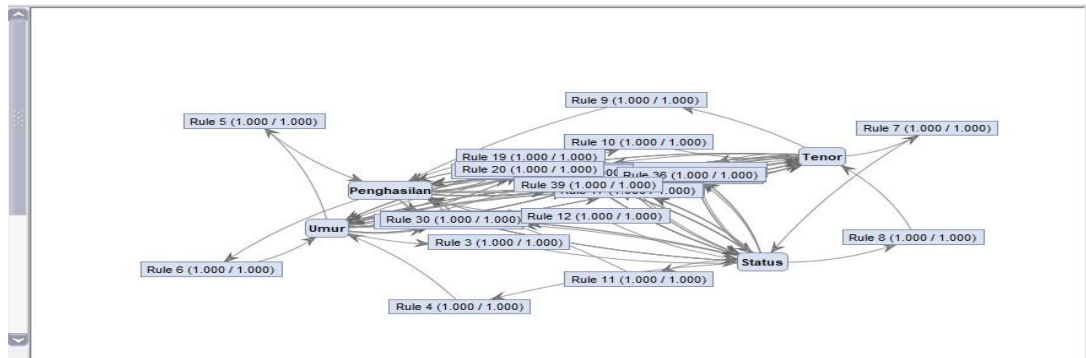
Implementasi proses pengolahan data menggunakan diskritisasi data yang digunakan untuk mengurangi jumlah nilai *atribute numeric* dengan cara membagi *range* atribut ke dalam interval. Label interval kemudian dapat digunakan untuk menggantikan nilai-nilai aktual data. Hasil pemodelan yang telah di proses oleh *Tool RapidMiner* selain menghasilkan dalam bentuk pola pemodelan ini juga dapat mengetahui tingkat akurasi, *recall*, dan *precision*. Hasil pohon keputusan dalam implementasi C4.5 yang dapat dilihat pada gambar sebagai berikut :



Gambar 4 Pohon Keputusan Algoritma C45

3.4 Implementasi Klasifikasi Menggunakan Metode Algoritma CT-Pro Pada Rapid Miner

Hasil permodelan yang telah di proses oleh Tool RapidMiner dengan metode algoritma CT-Pro dapat diketahui nilai akurasi data tersebut seperti terlihat pada gambar di bawah ini.



Gambar 5 Implementasi Algoritma CT-Pro

Pada gambar diatas, kolom pertama adalah kode id nasabah atau nomor nasabah, kolom kedua adalah index, kolom ketiga adalah frekuensi kemunculan. Gambar diatas menunjukkan hasil CFP-Tree dari kelima informasi. Setiap simpul pada Tree mengandung nama sebuah *item* dan frekuensi kemunculannya dalam tiap lintasan transaksi. CFP-Tree dibentuk dengan cara berikut:

1. Pembacaan proses pertama pada tabel 5, sehingga diperoleh lintasan 123456789101
2. Pembacaan proses kedua yaitu Angsuran memiliki node 1 an 2, Penghasilan 3,4 dan 5, Tenor 6 dan 7, umur 8 dan 9, Status 10 dan 11, Pembayaran 12 dan 13.
3. Pembacaan transaksi terus dilanjutkan sesuai dengan ketentuan sebelumnya. Apabila transaksi diawali dengan node baru yang bukan root, maka node berada pada index baru. Tahap ini terus berlangsung sampai pembacaan semua transaksi dan CFP-Tree berhasil dibangun.

4 KESIMPULAN

Setelah diproses analisis dan diperoleh hasil pengujian data dari algoritma C45 dan algoritma CT Pro, maka kedua algoritma tersebut memiliki hasil yang berbanding sehingga dapat ditarik kesimpulan mana algoritma yang akurat. Adapun hasil dari kedua algoritma tersebut dapat dilihat sebagai berikut:

**Tabel 4.1** Perbandingan Algoritma C45 dan Algoritma CT Pro

No.	Algoritma C45	Algoritma CT Pro
1.	Menampilkan pohon decision tree yang menunjukkan pola nasabah yang dapat dibaca dengan jelas.	Menampilkan pola nasabah yang sulit dibaca karena menggunakan node-node pada CT Pro.
2.	Algoritma C45 menghasilkan pohon keputusan berupa notasi dan angka.	Algoritma CT PRO menghasilkan pohon keputusan berupa node.

5 Referensi

- Anam, C., & Santoso, H. B. Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Klasifikasi Penerimaan Beasiswa. *ISSN : 2088-4591*. 2018.
- Dewi, S. Komparasi Metode Algoritma Data Mining Pada Prediksi Uji Kelayakan Credit Approval Pada Calon Nasabah Kredit Perbankan. *Jurnal Khatulistiwa Informatika, Vol VII, No. 1, ISSN : 2579 - 633X59*. 2019.
- Haryati, S., Sudarsono, A., & Suryana, E. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Media Infotama* , 130. 2015.
- Hermawati, F. A. *Data Mining*. Yogyakarta: Andi. 2013.
- Hijriana, N., & Muttaqin, R. Penerapan Metode Decision Tree Algoritma C4.5 Untuk Klasifikasi Mahasiswa Berprestasi. *Al Ulum Sains dan Teknologi* , 39-40. 2016.
- Kusmini, L., & Taufiq, E. *Algoritma Data Mining*. Yogyakarta: Andi. 2009.
- Sulianta, F., & Juju, D. *Rapid Miner (YALE)*. Jakarta: PT. Elex Media Komputindo. 2010.
- Suyanto. *Data Mining Untuk Klasifikasi dan Klustering Data*. Bandung: Informatika. 2017.