

PENERAPAN ALGORITMA DATA MINING UNTUK MENENTUKAN PINJAMAN DI KOPERASI PEMULIHAN RAKYAT

**Diajukan untuk Memenuhi Salah Tugas Akhir dari Program
Studi Teknik Informatika**



Oleh:

Ahmad Avrizal Widiarso

A11.2019.11851

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO
SEMARANG
2022**

KATA PENGANTAR

Alhamdulillahillâhi rabbill ‘aalamiin. Segenap puji syukur kami panjatkan ke hadirat Allah swt, yang telah melimpahkan petunjuk, bimbingan dan kekuatan lahir batin kepada kami, sehingga tugas akhir yang sederhana ini guna menyelesaikan tugas mata kuliah metodologi penelitian terselesaikan dengan sebagaimana mestinya. Sholawat dan salam semoga dilimpahkan oleh-Nya kepada junjungan kita Nabi Besar Muhammad saw, sosok historis yang membawa proses transformasi dari masa yang gelap gulita ke zaman yang penuh peradaban ini, juga kepada para keluarga, sahabat serta semua pengikutnya yang setia disepanjang zaman.

Karya tulis yang berjudul **“TUGAS AKHIR PENILAIAN ULANGAN TENGAH SEMESTER”** pada dasarnya merupakan persyaratan dan tugas di mata kuliah **“METODOLOGI PENELITIAN”** Tahun ajaran 2021/2022 ini Karya tulis ini berisikan suatu pembetulan ulang dalam pembangunan suatu sistem koperasi yang akan mengangkat dana pinjaman maupun data tarik dari koperasi, dan juga membuat sebuah sistem yang dapat mempermudah koperasi berjalan dalam kurun waktu jangka panjang.

Usaha dalam menyelesaikan tugas akhir ini memang tidak bisa lepas dari berbagai kendala dan hambatan, tetapi dapat kami selesaikan juga walaupun masih banyak kekurangan yang ada. Oleh karena itu izinkan kami mengucapkan terima kasih kepada hamba-hamba Allah yang membantu kami sehingga karya sederhana ini bisa menjadi kenyataan, bukan hanya angan dan keinginan semata, diantaranya kepada:

1. Dr. Khafiizh Hastuti, S.Kom., M.Kom.selaku Kepala mata kuliah Metodologi Penelitian yang telah memberi saran.
2. Dr. Khafiizh Hastuti, S.Kom., M.Kom.selaku dosen pembimbing yang telah memberikan pengarahan dan bimbingan dalam membuat tugas akhir ini, sehingga dapat terselesaikan dengan baik.
3. Dr. Khafiizh Hastuti, S.Kom., M.Kom.selaku guru Mata Kuliah Metodologi Penelitian.
4. Seluruh bapak ibu dosen serta karyawan Universitas Dian Nuswantoro.
5. Ayah dan ibu tercinta.

Akhirnya, penulis sadar bahwa dalam penyusunan tugas akhir ini masih terdapat banyak kekurangan. Namun, terlepas dari kekurangan yang ada, kritik dan saran yang konstruktif sangat kami harapkan untuk perbaikan di masa yang akan datang. Besar harapan penulis, karya tulis ini dapat bermanfaat bagi diri sendiri maupun orang lain.

Semarang, 22 April 2022.

penulis

Daftar Isi

PENERAPAN ALGORITMA DATA MINING UNTUK MENENTUKAN PINJAMAN DI KOPERASI PEMULIHAN RAKYAT	1
KATA PENGANTAR.....	2
BAB I	4
PENDAHULUAN.....	4
1.1. Latar Belakang	4
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Maksud dan Tujuan.....	5
1.5 Manfaat Penelitian	5
DAFTAR PUSTAKA.....	20

BAB I

PENDAHULUAN

1.1. Latar Belakang

Tugas akhir adalah tugas menghasilkan pengetahuan atau sesuatu yang dapat dilakukan menurut aturan atau prosedur tertentu yang dijelaskan secara ilmiah dan diterima secara luas oleh para profesional sebagai metode ilmiah (Soedjono, 1992). Penetapan judul tugas akhir perlu mendapat persetujuan fakultas melalui tahapan-tahapan tertentu. Di Universitas Dian Nuswantoro khususnya Program Studi Teknik Informatika salah satu tahapannya adalah Seminar Kolokium. Kolokium adalah fase di mana instruktur memutuskan apakah proposal risalah perlu dinyatakan layak atau diganti dengan judul baru.

Koperasi Pemulihan Rakyat adalah lembaga keuangan yang menyediakan layanan simpan pinjam bagi nasabah. Koperasi Rekonstruksi Rakyat adalah bagian dari Koperasi Pegawai Negeri Sipil Indonesia (KPRI) dan bertujuan untuk berkontribusi pada kesejahteraan finansial pelanggannya. Koperasi ini dapat menangani dua pinjaman, pinjaman modal atau pinjaman komersial. Pada tahun 2021, hampir 600 pelanggan terdaftar di Koperasi Rekonstruksi Rakyat.

Wawancara dengan Ketua Koperasi Rekonstruksi Rakyat KPRI menunjukkan bahwa ketua KPR KPR sempat kesulitan dalam memutuskan pinjaman baru. Hal ini didasarkan pada kenyataan bahwa klasifikasi yang sedang berjalan hanya didasarkan pada tinggi, tinggi, atau durasi angsuran pinjaman. Akibatnya, lebih dari 600 orang mengalami kredit macet. Akibatnya, KPR akan mengalami penurunan pendapatan koperasi yang sangat tajam, arus kas yang stagnan dan penurunan modal koperasi. Oleh karena itu, pada saat pengambilan keputusan pinjaman baru dilakukan proses data mining pada data history pinjaman. KPR dapat menggunakan data riwayat pinjaman untuk menentukan pinjaman baru.

Data mining adalah sekumpulan proses untuk mengambil informasi dari sekumpulan data yang dapat digunakan secara khusus untuk mendukung pengambilan keputusan. Dalam data mining terdapat banyak metode klasifikasi dalam proses ini, dan metode klasifikasi data mining yang digunakan dalam penelitian ini adalah pohon keputusan. Pengambilan keputusan menguraikan masalah karena pengambilan keputusan adalah metode klasifikasi yang menggunakan pengambilan keputusan untuk memodelkan data dari perspektif prediktif. Saat membangun pohon keputusan, penelitian ini menggunakan algoritma C4.5. Algoritma C4.5 dapat membangun pohon keputusan sederhana yang spesifik.

Berdasarkan permasalahan yang telah dipaparkan sebelumnya, maka dilakukan penelitian yang berjudul **"Penerapan Algoritma C4.5 Data Mining Untuk Menentukan Pinjaman di Koperasi Pemulihan Rakyat"**

1.2 Rumusan Masalah

Berdasarkan yang telah kita dapatkan/disampaikan pada latar belakang, maka permasalahan yang akan dibahas dan diteliti yaitu apakah pengelompokan kasifikasi dapat membantu untuk menentukan pinjaman lancar dan macet terhadap pinjaman baru

1.3 Batasan Masalah

Agar dalam enelitian ini tidak meluasnya permasalahan yang akan dibahas, maka diterapkan batasan masalah agar tercapainya sasaran yang diinginkan.

Batasan masalah tersebut diantaranya:

1. Data yang akan digunakan dalam penelitian ini adalah data riwayat simpan pinjam tahun 2019 dan 2020.
2. Metode yang digunakan metode klasifikasi yang bertujuan untuk pengelompokan pinjaman nasabah lancar dan macet dapat melakukan pinjaman.
3. Sistem yang akan dibangun berbasis website.
4. Lama angsuran untuk pinjaman baru maksimal 24 bulan
5. Besar pinjaman untuk pinjaman baru maksimal Rp 20.000.000
6. Penggunaanya hanya untuk ketua koperasi.
7. Keluaran dari sistem merupakan hasil dari algoritma C4.5, yang akan menampilkan aturan dalam mengambil keputusan.
8. Bahasa pemrograman yang digunakan adalah PHP dan database management system MySQL

1.4 Maksud dan Tujuan

Maksud penelitian ini yaitu menerapkan data mining kasifikasi untuk menentukan pinjaman nasabah lancar dan macet terhadap pinjaman baru. Adapula tujuan yang akan dicapai dalam penelitian ini adalah untuk membantu ketua koperasi dalam melakukan pemilahan pinjaman lancar dan macet berdasarkan riwayat pinjaman terhadap pinjaman baru.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah :

- 1) Ketua koperasi dapat mudah mengatu sistem pinjaman.
- 2) Penggunaan sistem ini akan mengurangi tingkat macetnya pinjaman baru.
- 3) Memudahkan dalam mendata para nasabah baru.
- 4) Memepermudah ketua mengelompokan data berdasarkan riwayat yang sudah ada.
- 5) Memudahkan para pengguna pinjaman baru untuk melakukan simpan pinjam.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Studi

Beberapa hasil dari penelitian yang saya lakukan adalah sebagai berikut :

1. "Penerapan Data Mining Untuk Koperasi Simpan Pinjam "

Penelitian yang membahas tentang bagaimana caranya merancang sebuah aplikasi yang dimana menerapkan beberapa teknik data mining dalam suatu pendataan pinjaman di suatu Koperasi Simpan Pinjam di Kecamatan Boja. Tujuan dilakukannya penelitian ini yaitu untuk menyelesaikan permasalahan penentuan besar maupun kecilnya pinjaman yang akan di pinjam dari koperasi tersebut dengan data mining, karena metode yang digunakan pada umumnya masih menganalisis data nasabah dengan total berapa lama waktu yang dibutuhkan untuk para peminjam tersebut akan mengembalikan uangnya ke dalam target awal yang sudah dijanjikan. Teknik yang akan diterapkan dalam aplikasi yang dibuat adalah dengan cara pohon keputusan dan algoritma yang di pakai adalah algoritma C4.5.

2. "Klasifikasi Menggunakan Algoritma C4.5."

Penelitian ini menjelaskan tentang klasifikasi menggunakan algoritma C4.5. Penyelidikan dilakukan dengan menggunakan metode pohon keputusan dan algoritma C4.5. Data yang digunakan adalah 9.993 dari institusi medis bersama dari total 10.533 data yang tersedia. Atribut yang digunakan terdiri dari jumlah negara bagian, nama kota, nama fasilitas, dan nilai fasilitas.

3. "Penerapan Algoritma C4.5. Untuk Penentuan Jumlah Provinsi dari Fasilitas Kesehatan"

Penelitian ini membahas tentang penggunaan pendataan teknik kasifikasi data mining dengan menggunakan algortima C4.5. yang ditetapkan dalam menentukan nama provinsi dalam bidang kesehatan yang akan diperoleh saat melakukan pendataan terhadap nama fasilitas sesuai dengan nilai fasilitas yang terdapat dalam fasilitas tersebut. Parameter digunakan sebgai data pemilihan kota/provinsi pada tahun 2019/2020. Hasil eskperimen dan evaluasi menunjukan tingkat cross validation, yaitu 97,5% dan tanpa cross validation, yaitu 91,17%

4. "Klasifikasi Data Nasabah Sebuah Asuransi menggunakan Algoritma C4.5."

Penelitian ini membahas tentang pengklasifikasian data nasabah pada sebuah pinjaman yang akan disalurkan sehingga dapat menemukan pola status peminjam sebagai bahan analisis dalam menentukan jumlah waktu atau tenggat waktu yang akan diambil untuk masa mendatang. Atribut yang digunakan dalam penelitian ini adalah penghasilan, jumlah waktu, tenggat waktu, nominal pinjaman, dan cara pembayaran. Hasil yang didapat dalam penelitian ini adalah aplikasi yang dapat menyimpan beberapa data yang dimana pinjaman yang sudah dilakukan kepada seseorang, agar bukti sewaktu peminjaman masih ada dengan menggunakan bunga sebesar 2% dalam peminjaman, dan supaya bisa mengetahui berapa pembayaran yang harus dilakukan setiap bulannya ketika sudah mendapatkan bunga 2% dengan jumlah rata - rata pinjaman yang diberikan.

Dari beberapa penelitian terkait yang telah disebutkan diatas, maka didapatkan *state of the art* sebagai berikut :

Tabel 2.1 State Of The Art

No	Judul	Penulis	Tahun	Ringkasan
1.	Penerapan Data Mining Untuk Menentukan Besar Pinjaman Koperasi Simpan Pinjam	Indi Nadya Syahfitri	2017	Perancangan aplikasi dengan teknik data mining menggunakan algoritma C4.5. untuk pendataan criteria besar kecil
2.	Klasifikasi Masa Studi Mahasiswa Fakultas Teknik Informatika Universitas Gunadarma Menggunakan Algoritma C4.5.	Kemal Ade Sekarwati	2019	Metode pohon keputusan dan algoritma C4.5. untuk menghitung total fasilitas kesehatan di Indonesia
3.	Penerapan Algoritma C4.5. Untuk Penentuan Jumlah Provinsi dari Fasilitas Kesehatan	Kemal Ade Sekarwati	2019	Menerapkan algoritma C4.5 untuk menunjukan berapa persen total fasilitas kesehatan yang sudah lulus uji tahap
4.	Klasifikasi Data Nasabah Sebuah Asuransi menggunakan Algoritma C4.5.	Larissa Navia Rani	2016	Terapkan algoritme C4.5 untuk menunjukkan persentase semua fasilitas medis yang lulus uji langkah

2.2 Tinjauan Pustaka

2.2.1 Koperasi Simpan Pinjam

Koperasi simpan pinjam adalah lembaga keuangan non bank yang misinya melayani masyarakat dalam bentuk pinjaman, simpanan sukarela, simpanan wajib, simpanan inti, dan berbagai instansi pemerintah yang kelebihan dana.

Jenis - jenis produk pinjaman yang ada di koperasi simpan pinjam :

1. Pinjaman Musiman

Pinjaman musiman adalah pinjaman yang dipinjamkan setiap bulan oleh anggota atau calon anggota sebagai modal kerja. Setiap bulan, debitur hanya membayar bunga dan membayar produk pinjaman utama pada akhir periode atau pada saat jatuh tempo.

2. Pinjaman Angsuran

Pinjaman angsuran adalah pinjaman yang diberikan kepada anggota atau pemangku kepentingan sebagai modal kerja sampai dengan 84 bulan atau 7 tahun secara angsuran atau angsuran.

2.2.2 Data

Sebuah Dataset adalah kumpulan dari berbagai objek dan atribut. Atribut juga dikenal sebagai variabel, bidang, karakteristik, atau fitur. Meskipun objek adalah kumpulan dari berbagai atribut. Objek juga dikenal sebagai kumpulan data, titik, kasus, sampel, entitas, atau instance.

Tipe - Tipe dari himpunan data (data-set) anantara lain :

1. Data Matrix

Jika objek data memiliki kumpulan atribut numerik yang sama, objek data dapat direpresentasikan sebagai titik dalam ruang multidimensi, dengan setiap dimensi mewakili atribut yang berbeda.

No. Sampel	Kuota Internet (GB)	Penghasilan (juta Rp)	Penggunaan Bensin (liter)
1	150	10	20
2	80	8	16
3	200	12	32
4	50	6	20

Gambar 2.1 Contoh Data Matrix

2. Data Dokumen

Untuk tipe data dokumen, setiap dokumen adalah "konsep" vektor. Setiap suku merupakan komponen dari vektor. Nilai setiap komponen menunjukkan berapa kali istilah muncul dalam dokumen.

	team	coach	play	ball	score	game	win	lost	timeout	season
Document 1	3	0	5	0	2	6	0	2	0	2
Document 2	0	7	0	2	1	0	0	3	0	0
Document 3	0	1	0	0	1	2	2	0	3	0

Gambar 2.2 Contoh Data Dokumen

3. Data Transaksi

Data transaksi adalah jenis catatan khusus di mana setiap catatan (transaksi) berisi satu set item.

Tanggal	Keterangan	Ref.	Debit	Kredit
5 Jan 2020	Kas Modal awal		500.000.000	500.000.000
11 Jan 2020	Sewa dibayar dimuka Kas		20.000.000	20.000.000
15 Jan 2020	Peralatan Perlengkapan Kas		10.000.000 5.000.000	15.000.000
20 Jan 2020	Kas Pendapatan		10.000.000	10.000.000
31 Jan 2020	Beban gaji Kas		25.000.000	25.000.000
	TOTAL		570.000.000	570.000.000

Gambar 2.3 Contoh Data Transaksi

4. Data Graf

Data graf adalah data graf yang terdiri dari simpul (node) dan ridge (tepi).

2.2.3 Data Mining

Data mining adalah proses menganalisis dan mengekstrak pengetahuan secara otomatis menggunakan satu atau lebih teknik pembelajaran komputer.

Data mining adalah proses interaktif dan iteratif untuk menemukan pola atau model yang sama sekali baru yang mudah dimengerti dari database yang sangat besar.

Data mining mencari database untuk data, tren, atau pola minat dan membuat keputusan di masa depan. Pola ini dikenali oleh alat tertentu yang membantu Anda menganalisis data dengan alat pendukung lainnya.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

1. Penjelasan

Peneliti mungkin perlu mencari data, menganalisis seluruh data, atau menemukan gambaran lengkap tentang pola dan tren yang unik untuk data yang diperoleh. Penjelasan pola dan tren seringkali memiliki kekuatan penjas yang mengarah pada penemuan pola dan tren dalam data hasil analisis.

2. Estimasi

Estimasi mirip dengan klasifikasi, kecuali bahwa respons ditentukan secara numerik daripada kategoris. Model dibangun dengan dataset lengkap yang menetapkan nilai variabel sebagai nilai target yang diprediksi. Pada penelitian selanjutnya, kami memperoleh estimasi variabel target dari data yang kami peroleh, berdasarkan nilai variabel prediktor.

3. Prakiraan

Prakiraan mirip dengan klasifikasi dan perkiraan, kecuali bahwa nilai dalam hasil data diprediksi di masa depan. Anda juga dapat memprediksi hasil dari data Anda menggunakan beberapa metode dan teknik yang digunakan untuk klasifikasi dan estimasi.

4. Klasifikasi

Mendapatkan variabel target untuk kategori.

5. Pengelompokan

Clustering adalah proses pengelompokan, pemantauan, atau memperhatikan data untuk membentuk kelas objek yang memiliki kesamaan. Cluster adalah kumpulan record yang memiliki kesamaan.

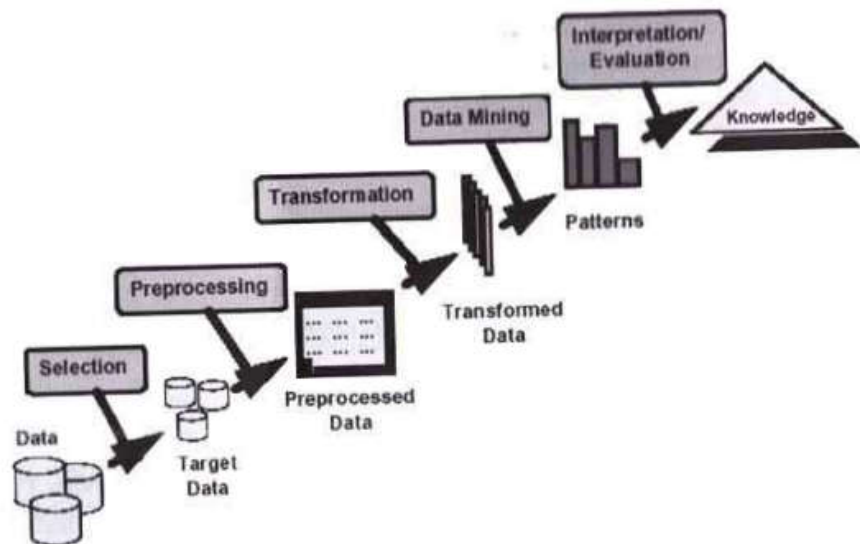
6. Asosiasi

Tugas yang terkait dengan data mining adalah menemukan atribut yang terjadi pada saat yang bersamaan. Dalam dunia bisnis lebih sering disebut market basket analysis.

Proses Knowledge Discovery (KDD) dalam database yang merupakan tahapan proses saat menggunakan data mining adalah sebagai berikut:

1. Ketahui domain aplikasi dan pahami untuk menelepon pengetahuan sebelumnya dan apa target pengguna.
2. Buat kumpulan data target yang berisi pemilihan data dan fokus pada subset data.
3. Pembersihan konversi data mencakup opsi penghapusan kebisingan, outlier, nilai yang hilang, dan pemilihan dan pengurangan fitur ukuran.

4. Penggunaan algoritma data mining yang terdiri dari asosiasi, sekuensial, klasifikasi, pengelompokan, dll.
5. Menafsirkan, mengevaluasi, dan memvisualisasikan pola ada yang baru dan menarik



Gambar 2.4 Proses Data Mining

Faktor-faktor yang mendorong kemajuan luar biasa dalam penambangan data adalah:

1. Pertumbuhan arsip yang cepat.
2. Penyimpanan data di gudang data untuk memberikan akses seluruh perusahaan ke database terpercaya.
3. Akses data melalui navigasi web dan internet semakin meningkat.
4. Pengembangan teknologi perangkat lunak untuk data mining.

2.2.4 Klasifikasi

Klasifikasi adalah contoh model atau proses penambangan data, dan klasifikasi diberikan sekumpulan data yang bertujuan untuk mengklasifikasikan model ke dalam kategori atau kelas yang sesuai. Kemudian gunakan model ini untuk mengklasifikasikan record yang model kelasnya sebelumnya tidak diketahui.

Komponen-komponen utama dari proses klasifikasi antara lain :

1. Kelas, merupakan variabel terikat yang merepresentasikan label hasil klasifikasi data.
2. Prediktor adalah variabel independen dari model berdasarkan karakteristik atribut data yang diklasifikasikan.

3. Dataset pelatihan adalah dataset lengkap yang berisi kelas prediktor yang melatih model sehingga dapat dikelompokkan ke dalam kelas yang benar.
4. Dataset eksperimental adalah plot data yang berisi beberapa dataset yang dikelompokkan menjadi sebuah model dengan mengetahui akurasi dan model yang dibuat.

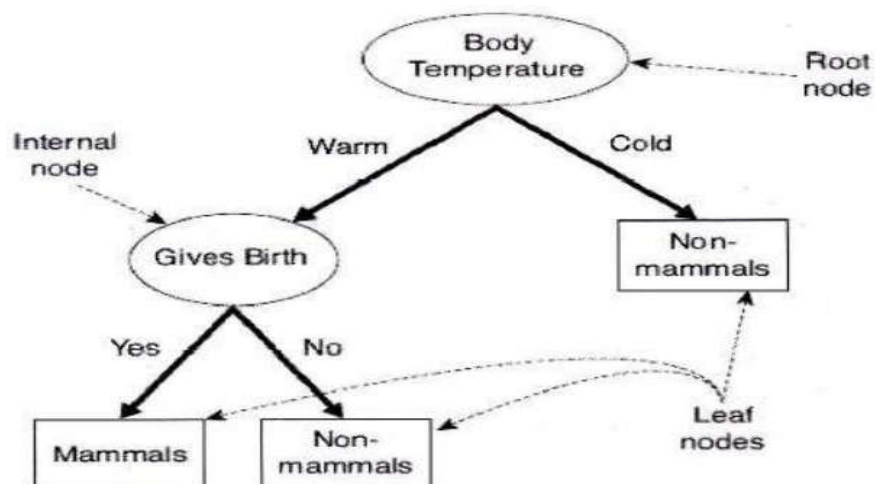
2.2.5 Pohon Keputusan

Decision Tree adalah klasifikasi dan teknik prediksi yang sangat kuat dan terkenal untuk mengubah fakta yang sangat besar menjadi aturan yang berlaku untuk Decision Tree ini. Pohon keputusan juga dapat membantu Anda memeriksa data Anda untuk menemukan hubungan tersembunyi antara berbagai variabel input potensial yang merupakan variabel target. Pohon keputusan dapat menggabungkan eksplorasi data dan pemodelan data, sehingga berguna untuk melakukan langkah pertama dari proses pemodelan. Dari langkah ini, dipilih dari awal hingga akhir, berbeda dengan teknik lainnya.

Pohon keputusan adalah struktur yang dapat digunakan untuk membagi catatan besar menjadi catatan yang lebih kecil dengan menerapkan beberapa aturan keputusan. Setiap baris split, anggota dataset akan menghasilkan dengan anggota lainnya.

Algoritma yang digunakan untuk membangun pohon keputusan adalah algoritma ID3, CART, dan C4.5 (pengembangan algoritma ID3).

Proses pembuatan pohon keputusan dengan mengubah bentuk data menjadi model pohon, mentransformasikan model menjadi aturan, dan menyederhanakan aturan.



Pada pohon keputusan terdapat 3 jenis node, yaitu :

1. Root Node, adalah simpul atas. Node ini tidak memiliki entri dan tidak dapat memiliki jalan keluar atau beberapa jalan keluar.

2. Internal Node, adalah simpul cabang. Node ini memiliki satu input dan setidaknya dua output.
3. Leaf Node atau Terminal Node, adalah simpul terakhir. Node ini hanya memiliki satu input dan tidak ada output.

2.2.6 Algoritma C4.5

Algoritma C4.5 merupakan sekumpulan algoritma pohon keputusan (decision tree). Algoritma ini memiliki input berupa contoh dan contoh pelatihan. Contoh Pelatihan Contoh data yang digunakan untuk membangun pohon yang divalidasi. Sampel, di sisi lain, adalah bagian dari data yang nantinya akan digunakan sebagai analisis parametrik ketika mengklasifikasikan data.

Algoritma C4.5 merupakan evolusi dari algoritma ID3 dan dirancang untuk memperbaiki data yang hilang atau rusak dalam file data yang disiapkan. Algoritma C4.5 memiliki kelebihan yaitu mudah dipahami, fleksibel, dan menarik karena dapat dijelaskan dalam hal pengambilan keputusan panggilan telepon (gambar).

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai root, didasarkan pada nilai boost tertinggi dari atribut yang ada. Rumus berikut digunakan untuk menghitung keuntungan.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots \dots \dots (2.1)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S_i|: jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Sedangkan untuk menghitung entropy digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots \dots \dots (2.2)$$

Keterangan:

S: himpunan kasus

A: fitur

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

2.2.7 Tahapan Data Mining Dalam Penelitian

Beberapa tahapan data mining pada penelitian diantaranya :

1. Mempersiapkan data-data yang akan digunakan dalam penelitian, Misal, data nominal pinjaman dan data pengajuan pinjaman.
2. Melakukan pengolahan data awal atau pembersihan data, yaitu dengan menentukan atribut apa saja yang akan digunakan dan diperlukan dalam penelitian, menghilangkan atribut yang tidak diperlukan dan membersihkan data yang kosong maupun data lama yang sudah tidak digunakan kembali, Misal, atribut bank pinjaman, data pinjaman uang lama, dan sebagainya.
3. Membagi data menjadi dua bagian yaitu 75% data training dan 25% data testing dari keseluruhan data.
4. Penerapan algoritma C4.5 pada data dengan menghitung nilai pinjaman atau nominal pinjaman dari entropy dan gain. Nilai gain tertinggi akan digunakan sebagai akar dari awal pinjaman.
5. Pembuatan pohon keputusan dari hasil entropy dan gain.
6. Pembuatan data dari pohon keputusan untuk melengkapi data pinjaman yang masih kosong.
7. pembuatan database untuk penyimpanan data yang sudah terkumpul dalam satu file.
8. penghitungan nilai total pinjaman dengan total bunga yang akan didapatkan dari hasil berapa lama akan meminjam.

BAB III

METODE PENELITIAN

3.1 Dasar Penelitian

Penelitian ini dilakukan berdasarkan rumusan masalah yang telah dijabarkan pada bab sebelumnya yaitu mengatur aturan pinjaman dana atau mengatur jumlah anggaran dana yang diberikan kepada para peminjam di “KPR” Boja – Kendal 2020.

3.2 Instrumen Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, maka alat dan bahan yang diperlukan dalam penelitian ini adalah :

3.2.1 Bahan

Dalam penelitian ini menggunakan data pinjaman di “KPR” Boja Kendal

3.2.2 Peralatan

Pada penelitian yang akan dilakukan untuk menyelesaikan tugas akhir ini, instrument peralatannya yang berupa software dan hardware meliputi :

1. Perangkat Lunak(Software)

a) System Operasi windows 7

Menggunakan Windows 7 32/64 bit sebagai sistem operasi pada penelitian ini

b) Rapid Miner

Rapid Miner adalah perangkat lunak yang akan digunakan penulis untuk mendata atau menguraikan metode apa saja yang akan digunakan dalam penelitian ini.

c) Database(MySQL)

Database ini bertujuan untuk menyimpan beberapa data yang akan dijalankan melalui penelitian ini sebagai media penyimpanan beberapa data yang kemungkinan akan digunakan.

2. Perangkat Keras(Hardware)

Perangkat keras yang digunakan untuk menyelesaikan penelitian ini berupa Personal Computer dengan spesifikasi

- CPU : Intel® Core™ i3-3110M 2,40 Ghz
- VGA : Nvidia GT 740 4GB
- RAM 4GB

3.3 Sumber Data dan Jenis Data

Penelitian ini menggunakan data dari sejarah “KPR” Boja-Kendal yang berjumlah 467 record. Terdiri dari 15 atribut : NO, Cabang, Nama Debitur, No Kredit, Alamat, Jenis Kelamin, Umur, Agama, Status, Pekerjaan, Jumlah Tanggungan, Nilai Tanggungan, Jangka Waktu, Anguran Per-Bulan, Keterangan.

3.3.1 Jenis Data

Pada Penelitian ini akan dilakukan menggunakan dua jenis data yaitu data kualitatif dan kuantitatif.

a) Data Kualitatif

Data kualitatif adalah data berformat kata, dan data yang digunakan dalam penelitian ini adalah data kualitatif seperti nama industri, nama debitur, alamat, agama, dan pekerjaan.

b) Data Kuantitatif

Data kuantitatif adalah data berupa angka dan nilai yang dapat dihitung secara andal menggunakan statistik atau matematika. Fungsi data kuantitatif adalah untuk menentukan jumlah objek data yang akan diselidiki. Data kuantitatif adalah data yang menunjukkan kemungkinan data yang digunakan dalam penelitian yaitu No Kredit, Umur, Jumlah Tanggungan, Nilai Tanggungan, Jangka Waktu, Angsuran Per-Bulan.

3.3.2 Sumber Data

Selama penelitian di “KPR” Boja-Kendal penulis dapat mengumpulkan beberapa data, antara lain :

a) Data Primer

Data primer adalah data yang secara langsung diambil dari berbagai penelitian atau beberapa interview dari pihak KPR itu sendiri, dan juga bisa didapatkan dari database “KPR” Boja-Kendal.

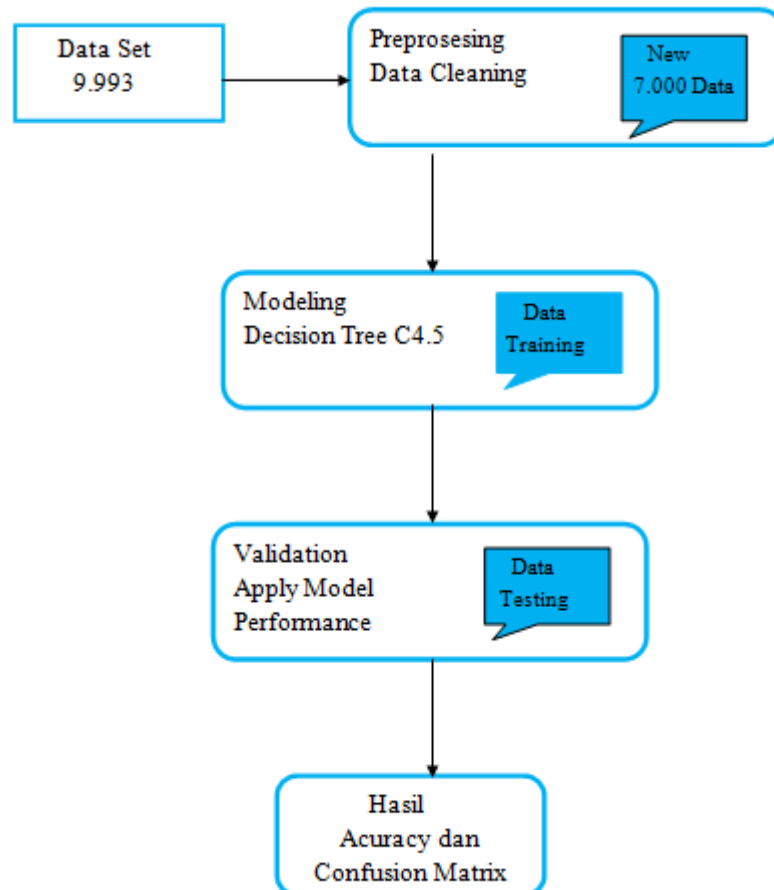
b) Data Sekunder

Data Yang didapatkan secara tidak langsung dari objek tertentu disebut data sekunder. Dengan memperoleh data dari studi pustaka diantara buku ataupun penelitian dari berbagai pembahasan yang mengenai data mining pada perbankan, dan juga membahas pengajuan angsuran kredit dengan menggunakan algoritma C4.5.

3.4 Model Atau Metode yang Diusulkan

1.1 Permodelan (Modeling)

Metode yang digunakan pada penelitian adalah Algoritma C4.5, Dalam permodelan ini algoritma C4.5 akan dicari performanya yaitu performa vector dan confusion matrix. Data yang digunakan sudah melalui proses cleaning data, untuk melakukan pengukuran penelitian ini pada aplikasi Rapid Miner.

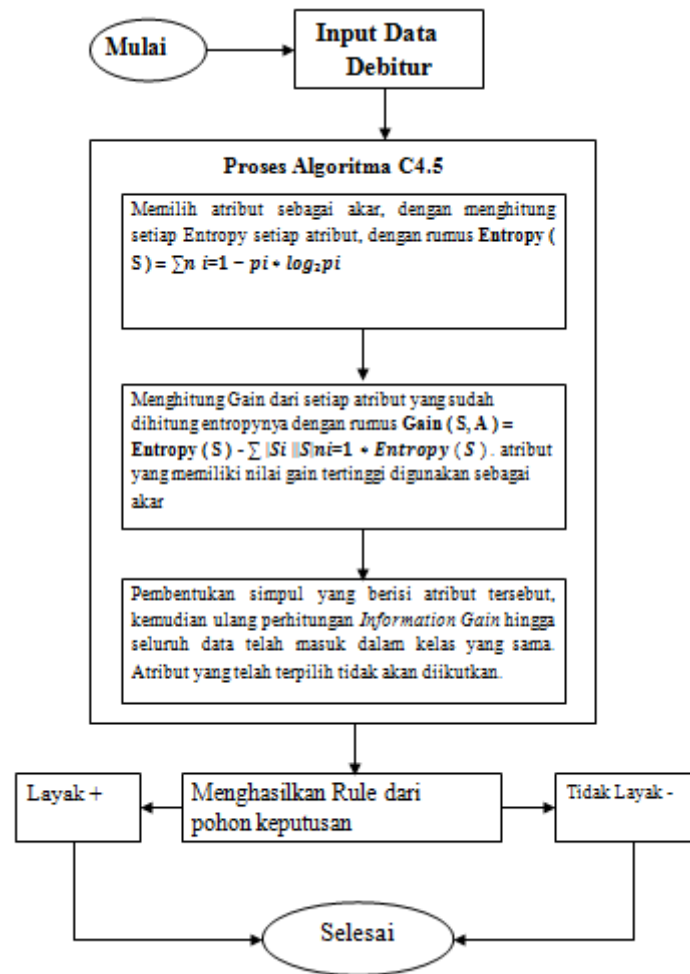


Gambar 3.1 Model tahap menemukan Accuracy dan Confusion Matrix

Dalam pengujian data set pertama dilakukan adalah melakukan proses cleaning data, dari 9993 data setelah dilakukan data cleaning menghasilkan 7000 data set baru.

Uji validasi menggunakan cross-validation dengan 2, 5, 7, dan 10 pengujian, dan uji validasi split dengan empat uji coba dengan data latih 60%, 70%, 80%, dan 90%. RapidMiner menjalankan proses pengujian validasi mutual dan split. Hal ini dilakukan dengan pemodelan menggunakan metode pohon keputusan C4.5 dan memasukkan model aplikasi dan kinerja. Validasi adalah proses mengevaluasi akurasi prediksi suatu model. Teknik validasi membagi data secara acak menjadi k buah yang disebut validasi silang. Cross-validation hanya melakukan konvolusi data atau jumlah partisi. Validasi silang menghasilkan matriks kebingungan dan akurasi model.

2.1 Pemodelan Algoritma C4.5



Gambar 3.2 Permodelan Algoritma C4.5

Keterangan :

Entropy (s)

S = Himpunan Kasus

A = Atribut

N= Jumlah Partisi S

|Si| = jumlah Jumlah kasus pada partisi ke i

pi = Proporsi dari Si terhadap S

Kumpulan data dibaca Identifikasi sampel data. Kemudian hitung entropi (S) dari semua atribut, kemudian gain tertinggi dari semua atribut, dan kemudian dapatkan atribut yang dijadikan root/node. Buat cabang untuk setiap nilai, pisahkan kasus di cabang, dan ulangi perhitungan keuntungan sampai semua data berada di kelas yang sama. Atribut yang dipilih tidak akan lagi disertakan dalam perhitungan. Untuk dataset dalam

penelitian ini terdiri dari dua grade, yaitu debitur yang memenuhi syarat yang ditandai dengan (+) dan debitur yang tidak memenuhi syarat yang ditunjukkan dengan (-).

3.5 Cara Pengujian Model/ Metode

Saat menghitung algoritma C4.5, dataset dibagi menjadi 70 data latih dan satu data uji.

3.5.1 Model Algoritma C4.5

1. Ini menghitung jumlah kasus yang diklasifikasikan ke dalam dua kelas, memenuhi syarat dan tidak memenuhi syarat, dan nilai entropi semua atribut. Usia, kepribadian, pendapatan, jumlah tanggungan, agunan. Kemudian cari nilai boost untuk setiap atribut.
2. Sebelum melakukan perhitungan gain, hitung entropy dari tiap atribut dengan rumus sebagai berikut $Entropy = \sum_{i=1}^n - p_i * \log_2 p_i$ Dengan perhitungan yang sama, dilakukan pada atribut berdasarkan pengelompokan jumlah kasus.
3. Kemudian menghitung nilai gain dan menentukan nilai maksimum untuk menentukan node atau rute pertama. Rumus untuk menghitung keuntungan adalah sebagai berikut.

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

4. Ulangi langkah Ke- 3 hingga semua atribut dan subset terhitung.
5. Proses pembentukan pohon keputusan berhenti jika sudah tidak ada atribut yang dipartisi dan semua tuple dalam node N telah memiliki kelas yang sama.

3.5.2 Tahapan Evaluasi Perhitungan Akurasi dan Confusion Matrix

Pada tahap ini, mengevaluasi keakuratan hasil yang diperoleh dari model yang digunakan. Gunakan matriks kebingungan yang disediakan oleh kerangka kerja RapidMiner.

Tabel 3.3 Confusion Matrix

Classification C4.5	Layak (+)	Tidak Layak(-)
Layak(+)	TN	FN
Tidak Layak(-)	FP	TP

Dari tabel di atas dapat hitung akurasinya sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

DAFTAR PUSTAKA

- Abdullah Umar. (t.thn.). *RapidMiner, Definisi dan Fitur-Fiturnya*. Dipetik July 6, 2022, dari Abd.Umar: <https://www.abdumar.com/2021/03/rapidminer-definisi-dan-fitur-fiturnya.html?m=1>
- Indi Nadya Syahfitri. (2018). Koperasi Simpan Pinjam Metode Algoritma C4.5. *Penerapan Data Mining untuk Menentukan Besar Pinjaman pada Koperasi Simpan Pinjam dengan Algoritma C4.5 (Studi Kasus: Koperasi Wanita XYZ)* , 1-2.
- IYKRA. (2018, July 23). *Mengenal Decision Tree dan Manfaatnya*. Dipetik July 6, 2022, dari IKRYA Mengenal Decision Tree dan Manfaatnya: <https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d>
- Kemal Ade Sekarwati. (2019, October 10). *Penerapan Data Mining Dengan Menggunakan Algoritma C4.5 Pada Klasifikasi Fasilitas Kesehatan Provinsi Di Indonesia*. Dipetik July 7, 2022, dari Library Gunadarma: <https://library.gunadarma.ac.id/repository/penerapan-data-mining-dengan-menggunakan-algoritma-c4-5-pada-klasifikasi-fasilitas-kesehatan-provinsi-di-indonesia-skripsi>
- Rani, Larissa Navia. (2016). Algoritma C4.5. *Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit* , 1.
- Sunjana. (2010, Juni 19). *KLASIFIKASI DATA NASABAH SEBUAH ASURANSI MENGGUNAKAN ALGORITMA C4.5*. Dipetik Juni 19, 2010

Feedback Studio - Google Chrome
ev.turnitin.com/app/carta/en_us/?s=8&o=1867356562&u=1133016846&lang=en_us&student_user=1

feedback studio Ahmad Avrizal Tes Palgiasi

**PENERAPAN ALGORITMA DATA MINING UNTUK
MENENTUKAN PINJAMAN DI KOPERASI PEMULIHAN
RAKYAT**

Diajukan untuk Memenuhi Salah Tugas Akhir dari Program
Studi Teknik Informatika

IVERSITAS DIAN NUSWANTC

Page: 1 of 20 Word Count: 3935 Text-Only Report High Resolution On 6:19 08/07/2022

Match Overview

49%

1	123dok.com Internet Source	15%
2	docplayer.info Internet Source	8%
3	eprints.dinus.ac.id Internet Source	7%
4	roudotunnafingah.wor... Internet Source	2%
5	reniwulansari.wordpres... Internet Source	2%
6	repository.uin-suska.ac... Internet Source	2%
7	Docplayer.Info Internet Source	2%

Itu Hasil Turnitin nya bu, Susah sekali bu mau nurunin persennanya, padahal udah di parafrase ulang dan udah diubah – ubah terus tetap aja segitu bu persennanya