



**MENCEGAH KREDIT MACET DENGAN ANALISA KELAYAKAN PEMBIAYAAN DENGAN METODE C4.5 DAN NAÏVE BAYES
(Studi Kasus : Koperasi BMT UGT Sidogiri Cabang Cikarang)**

Agung Nugroho¹, Dini Oktaviani²

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹agung@pelitabangsa.ac.id

Abstraksi

Kemajuan pertumbuhan UMKM (Usaha Mikro Kecil Menengah) di Indonesia dari waktu ke waktu yang semakin pesat mengakibatkan meningkatnya kebutuhan modal untuk mengembangkan usahanya. Hal ini terbukti dengan semakin banyaknya pengambilan kredit atau pembiayaan pada koperasi simpan pinjam maupun BPR (Bank Perkreditan Rakyat). Masalah yang dihadapi oleh Koperasi simpan pinjam, BPR, atau instansi keuangan lainnya saat ini dalam memberikan kredit adalah resiko terlambatnya pembayaran, pengembalian bahkan kegagalan pembayaran kredit. Masalah ini terjadi karena penyalahgunaan kredit dan lemahnya pengawasan baik dalam proses pemberian kredit maupun dalam tahap pelaksanaannya. Solusi yang tepat untuk memecahkan permasalahan yang ada yaitu dengan menggunakan algoritma data mining. Konsep Data Mining akan mempermudah mengatasi masalah yang belum optimal di koperasi, metode klasifikasi mampu menemukan model yang membedakan konsep atau kelas data dengan tujuan untuk mempermudah dalam memprediksi kelayakan kredit. Algoritma Naive Bayes dan algoritma C4.5 dinilai dapat memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Peneliti melakukan penelitian pada Koperasi BMT UGT Sidogiri dengan judul "Mencegah kredit macet dengan analisa kelayakan pembiayaan dengan metode Naive bayes dan C4.5". Dalam penelitian ini peneliti menggunakan 9 atribut sebagai penilaian yaitu : Nama, status tempat tinggal, akad pembiayaan, penghasilan, plafond, jangka pelunasan, jumlah tanggungan, jaminan. Pengujian dilakukan dengan menggunakan 520 data dan 104 data testing yang dipilih secara acak. Dari hasil pengujian yang dilakukan dengan menggunakan tools Rapid Miner dapat disimpulkan bahwa tingkat accuracy algoritma C4.5 lebih akurat yaitu sebesar 81,35% sedangkan algoritma Naive Bayes sebesar 78,85%.

Kata Kunci: Kredit, Klasifikasi, Akurasi, Naive Bayes, C4.5.

Abstract

The progress of the growth of MSMEs (Micro, Small and Medium Enterprises) in Indonesia from time to time which is increasingly rapid, resulting in an increase in the need for capital to develop their business. This is evidenced by the increasing number of credit or financing withdrawals from savings and loan cooperatives and BPRs (Rural Banks). The problem faced by savings and loan cooperatives, BPRs, or other financial institutions at this time in providing credit is the risk of late payments, repayments and even failure of credit payments. This problem occurs due to credit misuse and weak supervision both in the process of providing credit and in the implementation stage. The right solution to solve existing problems is by using data mining algorithms. The concept of data mining will make it easier to solve problems that are not optimal in cooperatives, the classification method is able to find

models that differentiate concepts or data classes with the aim of making it easier to predict creditworthiness. The Naive Bayes algorithm and the C4.5 algorithm are considered to be able to predict future opportunities based on previous experiences. The author conducted research on the BMT UGT Sidogiri Cooperative with the title "Preventing bad credit by analyzing the feasibility of financing with the Naive Bayes and C4.5 methods". In this study the authors used 9 attributes as an assessment, namely: name, residence status, financing contract, income, ceiling, term of repayment, number of dependents, collateral. Testing is done using 520 data and 104 randomly selected testing data. From the results of tests carried out using Rapid Miner tools, it can be concluded that the accuracy level of the C4.5 algorithm is more accurate at 81.35%, while the Naive Bayes algorithm is 78.85%.

Keywords : Credits, Classification, Accuracy, Naive Bayes, C4.5.

1. Pendahuluan

Usaha kecil menengah (UKM) dari waktu ke waktu mengalami perkembangan bagus. Usaha kecil menengah menjadi salah satu terobosan meningkatkan pertumbuhan ekonomi ditengah-tengah masyarakat untuk mencapai kesejahteraan hidup yang memadai. UKM ini diharapkan akan mampu mengurangi angka pengangguran jika melihat fakta lapangan pekerjaan yang semakin terbatas jumlah tenaga kerja yang belum terserap terus bertambah. Namun para pelaku UKM masih banyak ditemukan ketidaksempurnaan, salah satunya penyalahgunaan kredit. Berdasarkan data Direktorat Kredit, BPR dan UMKM Bank Indonesia (BI) mengungkapkan besaran kredit bermasalah (NPL) untuk sektor Usaha Mikro kecil Menengah (UMKM) pada awal 2019 mengalami peningkatan. Tingginya kredit macet di Koperasi/Bank Perkreditan Rakyat ini.

disebabkan oleh lemahnya pengawasan baik dalam proses pemberian kredit maupun dalam tahap pelaksanaannya. Koperasi harus mengurangi tingkat resiko dengan memperhatikan berbagai faktor yang bisa menyebabkan terjadinya kredit macet dalam pemberian pinjaman kepada nasabahnya. Berdasarkan masalah yang telah diuraikan diatas, maka peneliti mengangkat judul "Mencegah Kredit Macet Dengan Analisa Kelayakan Pembiayaan Dengan Metode algoritma C4.5 Dan Naive Bayes" yang diharapkan dapat memecahkan masalah yang dihadapi oleh sebagian besar Koperasi Maupun Bank Perkreditan Rakyat yang terdapat di Indonesia. Dipilihnya metode ini dianggap sesuai mengingat banyak penelitian terdahulu yang menggunakan metode ini untuk klasifikasi dan peneliti ingin membandingkan kedua metode tersebut untuk menganalisa kelayakan.

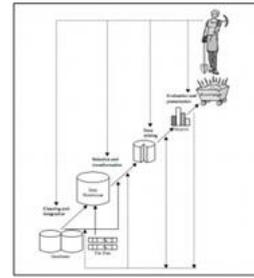
2. Tinjauan Studi

2.1. Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining merupakan teknologi yang digunakan dalam disiplin ilmu yang berbeda-beda untuk mencari hubungan yang signifikan data dengan mengubah data menjadi pohon keputusan (decision tree) dan aturan-aturan variabel dalam set data yang besar. Metode data mining dapat digunakan untuk melakukan klaturannya (rule), yang dapat digunakan untuk memprediksi atau mengklarifikasi kejadian.

2.2. Tahapan Data Mining

Istilah data mining dan knowledge Discovery in Database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Proses Knowledge Discovery in Database (KDD) secara garis besar dapat dijelaskan sebagai berikut:



Gambar 1. Tahapan Data Mining

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang akan diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file text. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada

1. Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja

2. Transformasi data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan, Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

3. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

4. Evaluasi Pola (Pattern Evaluation)

Untuk mengidentifikasi pola-pola knowledge based yang ditemukan. Dalam tahap ini hasil dalam teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila hasil yang dicapai tidak sesuai hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikan untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai sesuatu hasil yang diluar dugaan yang mungkin bermanfaat.

5. Presentasi pengetahuan (Knowledge Presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh

pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining.

2.3. Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [1]. Salah satu tugas yang dapat dilakukan dengan data mining adalah pengklasifikasian. Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasi spesies berdasarkan karakteristik fisik.

2.4. Naïve Bayes

Algoritma Naïve Bayes Classifier merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naïve bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh seorang ilmuwan inggris Thomas Bayes, yaitu memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema Naïve Bayes adalah :

$$P(H|X) = \frac{p(X|H) \times p(H)}{P(X)}$$

Keterangan :

X : Data dengan class yang belum diketahui
 H : Hipotesa data X merupakan suatu class spesifik
 P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
 P(H) : Probabilitas hipotesis H (prior probability)
 P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
 P(X) : Probabilitas X

2.5. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma Decision tree yang paling efektif untuk melakukan klasifikasi. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser) yang ditemukan oleh J Ross Quinlan. Pohon keputusan ini dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama. Bentuk pemecahan (split) yang digunakan untuk membagi data tergantung dari jenis atribut yang digunakan dalam split [2].

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

1. Pilih atribut sebagai akar
 Memilih atribut sebagai akar bisa dihitung dengan melihat nilai gain dari masing-masing atribut. Nilai gain tertinggi nantinya akan menjadi akar yang pertama.

Bentuk persamaan untuk mendapatkan nilai gain bisa dilihat pada persamaan(1)

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

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 nilai entropy sendiri bisa diperoleh dari persamaan (2).

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

Keterangan : S : Himpunan kasus n : Jumlah partisi s

Pi : Proporsi S_i terhadap S

2. Ulangi metode diatas hingga semua data terbagi.
3. Proses pengulangan pada metode decision tree ini akan berhenti jika :
 - a. Semua data telah terbagi rata
 - b. Tidak ada lagi atribut yang bisa dibagi lagi
 - c. Tidak ada data record dalam cabang yang kosong

2.6. Rapid Miner

Rapid miner adalah aplikasi data mining yang berbasis open source. Open source rapid miner berlisensi AGPL (GNU Affero General public License) versi 3. Penelitian mengenai tools ini dimulai sejak tahun ini dimulai sejak tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund yang kemudian diambil alih oleh SourceForge sejak tahun 2004. Rapid miner memperoleh peringkat satu sebagai tools data mining untuk proyek nyata pada poll oleh KDnuggets, sebuah Koran data mining pada 2010-2011.

2.7. Analisa Kelayakan Pembiayaan

Dalam menjalankan fungsinya sebagai lembaga penyalur dana, bank syariah perlu memperhatikan beberapa hal yang berkaitan dengan analisis kelayakan pembiayaan. Secara umum, analisis kelayakan pembiayaan tersebut terdiri atas beberapa tahapan, yaitu [3] :

1. Pendekatan analisis pembiayaan.
 Ada beberapa pendekatan analisis pembiayaan yang diterapkan oleh para pengelola bank syariah dalam kaitannya dengan pembiayaan yang akan dilakukan, yaitu :
 - a. Pendekatan Jaminan
 - b. Pendekatan karakter
 - c. Pendekatan kemampuan pelunasan
 - d. Pendekatan dengan studi kelayakan
2. Pendekatan fungsi-fungsi bank, Penerapan prinsip-prinsip analisis pembiayaan.
3. Prinsip analisis pembiayaan didasarkan pada rumus 5C, yaitu :
 - a. Character, yaitu sifat atau karakter nasabah pengambil pinjaman.
 - b. Capacity, yaitu kemampuan nasabah untuk menjalankan usaha dan mengembalikan pinjaman yang diambil.

- c. Capital, yaitu besarnya modal yang diperlukan peminjam.
 - d. Colateral, yaitu jaminan yang telah dimiliki yang diberikan peminjam kepada bank.
 - e. Condition, yaitu keadaan usaha atau nasabah prospek atau tidak.
4. Penerapan prosedur analisis pembiayaan. Aspek-aspek penting dalam analisis pembiayaan yang perlu dipahami oleh pengelola bank syariah adalah :
- a. Berkas dan Pencatatan.
 - b. Data pokok dan analisa pendahuluan
 - c. Penelitian data
 - d. Penelitian atas realisasi usaha
 - e. Penelitian atas rencana usaha
 - f. Penelitian dan penilaian barang jaminan
 - g. Laporan keuangan dan penelitiannya.

3. Metode Penelitian

3.1 Data yang digunakan

Pada penelitian ini data yang digunakan adalah data History pembiayaan periode Januari 2017 sampai Desember 2019, Data yang diperoleh dari koperasi BMT UGT Sidogiri cabang cikarang masih berupa raw data, yang berisi atribut nama, jenis kelamin, umur, status, status pekerjaan, gaji pokok/ penghasilan, status tempat tinggal, tanggungan beban keluarga, keperluan peminjaman. Kemudian data yang akan digali yaitu nilai dari setiap kriteria tersebut menjadi patokan untuk pemilihan kelayakan pembiayaan.

3.2 Pengolahan Data Awal

Pada tahap ini menjelaskan tentang tahap awal data mining. Data yang telah didapatkan akan diolah ke format yang dibutuhkan, pengelompokkan dan penentuan atribut, variabel.

1. Data Selection

Tahap data selection atau pemilihan data merupakan tahap pemilihan variabel dari data yang akan dianalisis, karena tidak semua data yang terdapat dalam data mentah akan digunakan. Sehingga didapat beberapa variabel yang akan dijadikan sebagai atribut.

Tabel 1. Pemilihan Variabel

Atribut/Variabel	Indikator	Detail Penggunaan
Nama	V	ID
Status tempat tinggal	V	Digunakan
Pendapatan rata-rata	V	Digunakan
Tanggungan keluarga	V	Digunakan
Jenis pembiayaan	V	Digunakan
Plafond	V	Digunakan
Waktu Pelunasan	V	Digunakan
Jaminan	V	Digunakan
Hasil	V	Label target

2. Pre-Processing (data cleaning)

Tahap *pre-processing* merupakan tahap dilakukannya pembersihan terhadap data *missing value* yaitu data yang tidak konsisten atau kosong, dan juga dilakukannya pemilihan variabel terhadap data yang ingin digunakan pada proses data mining.

Tabel 2. klasifikasi atribut penghasilan

PENGHASILAN	
<3 Juta	Sedikit
3 – 5 Juta	Cukup
>5 Juta	Banyak

Tabel 3. klasifikasi atribut Plafond

PLAFOND	
< 10 Juta	Rendah
10 – 25 Juta	Sedang
>25 Juta	Tinggi

Tabel 4. Klasifikasi Jangka waktu

Jangka Waktu	
<6 Bulan	Pendek
>6 Bulan	Panjang

3. Data Integration dan Transformation

Untuk meningkatkan akurasi dan efisiensi algoritma. Data dalam penelitian ini dikategorikan atau diklasifikasi. Setelah mendapatkan data selanjutnya adalah mengklasifikasi atribut-atribut karena berpengaruh terhadap hasil prediksi. Klasifikasi data ini dilakukan dengan bantuan tool Ms.Excel. Data bisa dilihat pada tabel dibawah ini:

Tabel 5. Sampel Data Klasifikasi

Nama	Status tempat tinggal	Pendapatan rata-rata	Akad pembiayaan	Plafond	Jkw (bulan)	aminan	Jumlah tanggungan	Status
awawi	Milik sendiri	Cukup	Bai bitsama nilaji	Rendah	Pendek	endaraan	Lebih dari 3	Layak
hammadriadi	Milik keluarga	Sedikit	Bai bitsama nilaji	Rendah	Panjang	endaraan	urang dari 3	Layak
ahimahapsari utami	Kontrak	Cukup	Bai bitsama nilaji	Rendah	Panjang	emas	urang dari 3	Layak

4. Pembahasan

4.1 Hasil Pengujian

Hasil pada penelitian ini dilakukan dengan pengolahan data history pembiayaan. Dengan menggunakan microsoft excel, kemudian diuji dengan tool dibantu aplikasi Rapid Miner dengan 520 data training dan 104 data yang telah di persiapkan untuk menentukan prediksi kelayakan pembiayaan.

4.2 Perhitungan Naïve Bayes

Terdapat seseorang yang akan mengajukan pembiayaan yang bernama Ibu Vira Nurifadiyah dengandata sebagai berikut :

Tabel 5. Kasus perhitungan Naive Bayes

Penghasilan	Akad Pembiayaan	Status tempat tinggal	Jangka waktu	Jaminan	Jumlah tanggungan	Plafond	Prediksi
Cukup	Murabahah	Miliksendiri	Panjang	kendaraan	Lebih dari 3	Rendah	?

Data Testing : X = (Penghasilan = "Cukup", Akad pembiayaan = "Murabahah", Status tempat tinggal = "Milik sendiri", jangka waktu = "Panjang", Jaminan = "Kendaraan bermotor", Jumlah tanggungan = "lebih dari 3", Plafond = " Rendah")

1. Tahap 1 menghitung jumlah kelas atau prediksi Dengan persamaan "P(C_i)" Keterangan :

P(C_i) = Probabilitas prior P(C₁) = Probabilitas Layak
P(C₀) = Probabilitas TidakLayak

Perhitungan :

P(C₁) = P(Layak) = 365/520=0.701 P(C₀) = P(Tidak) = 155/520=0.298

2. Tahap 2 menghitung jumlah kasus yang sama dengan kelas yang sama dengan persamaan "P(X|C_i)" Keterangan :

P(C_i) = Probabilitas prior P(C₁) =

Probabilitas Layak P(C₀) = Probabilitas TidakLayak

X = Kriteria Kejadian Perhitungan :

P(Penghasilan = "Cukup" | Layak) = 176/365 = 0.482

P(Penghasilan = "Cukup" | TidakLayak) = 56/155 =

0.361 P(Akad Pembiayaan = "Murabahah" | Layak) = 299/365 = 0.819

P(Akad Pembiayaan = "Murabahah" | Tidak Layak) = 111/155 = 0.716

P(Status tempat tinggal = "MilikSendiri" | Layak) = 172/365 = 0.471

P(Status tempat tinggal = "MilikSendiri" | Tidak Layak) = 31/155 = 0.2

P(Jangka waktu = "Panjang" | Layak) = 193/365 = 0.528

P(Jangka waktu = "Panjang" | TidakLayak) = 29/155 = 0.187

P(Jaminan = "Kendaraan bermotor" | Layak) = 263/365 = 0.720

P(Jaminan = "Kendaraan bermotor" | Tidak Layak) = 37/155 = 0.238

P(Jumlah Tanggungan = " Lebihdari 3" | Layak) = 161/365 = 0.441

P(Jumlah Tanggungan = " Lebihdari 3" | Tidak Layak) = 78/155 = 0.503

P(Plafond = " Rendah" | Layak) = 302/365 = 0.827

P(Plafond = " Rendah" | TidakLayak) = 86/155 = 0.554

3. Mengalikan semua hasil atribut

a. Mengalikan semua atribut "Layak" P(X|C₁) = P(X / Layak) P(X/Layak) = P(XPenghasilan = "Cukup" | Layak) * P(XAkad Pembiayaan = "Murabahah" | Layak) * P(XStatustempat tinggal = "Milik sendiri" | Layak)

* P(XJangka waktu = "Panjang" | Layak) * P(XJaminan = "Kendaraan bermotor" | Layak) * P(XJumlah Tanggungan = " lebihdari 3" | Layak) * P(XPlafond = "Rendah" | Layak) = 0.482 * 0.819 * 0.471 * 0.528 * 0.720 * 0.441 * 0.827 = 0.0257787803

b. Mengalikan semua atribut "Tidak" P(X|C₀) = P(X / Tidak)

P(X / Tidak) = P(XPenghasilan = "Cukup" | Tidak) * P(XAkad Pembiayaan = "Murabahah" | Tidak) * P(XStatus tempat tinggal = "Miliksendiri" | Tidak) * P(XJangka waktu = "Panjang" | Tidak) * P(XJaminan = "Kendaraan bermotor" | Tidak) * P(XJumlah Tanggungan = " Lebih dari 3" | Tidak) * P(XPlafond = " Rendah" | Tidak) = 0.361 * 0.716 * 0.2 * 0.187 * 0.238 * 0.503 * 0.554 = 0.000641130641

4. membandingkan nilai kelas "LAYAK" dan "TIDAK LAYAK"

a. Perhitungan pemaksimalan untuk memungkinkan klasifikasi ke dalam kelas Layak dalam penentuan perpanjangan kontrakkaryawan dengan mengalikannya hasil.

P(X|C₁) dengan P(C₁) :

P(X) | Layak * P(Layak) = 0.0257787803 * 0.701 = 0.018070925

0.018070925

b. Perhitungan pemaksimalan untuk memungkinkan klasifikasi ke dalam kelas Tidak dalam penentuan perpanjangan kontrakkaryawan dengan mengalikannya hasil.

P(X|C₀) dengan P(C₀) :

P(X) | Tidak Layak * P(TidakLayak) = 0.000641130641 * 0.298 = 0.000191056931

Dari perhitungan di atas dapat dihasilkan nilai P(C₁ | X) = 0.018070925

sedangkan nilai P(C₀ | X) = 0.000191056931.

Berdasarkan nilai tersebut dapat diambil kesimpulan bahwa P(C₁ | X) > P(C₀ | X) Maka data uji tersebut diklasifikasikan ke dalam kelas "Layak" dalam penentuan kelayakan Pembiayaan Perhitungan menggunakan algoritma Naive Bayes. algoritma C4.5.

4.3 Perhitungan Algoritma C4.5

Proses perhitungan klasifikasi kelayakan pengajuan

kredit menggunakan algoritma C4.5 adalah sebagai berikut :

- 1) Menghitung jumlah kasus, jumlah kasus untuk keputusan Tidak Layak, Jumlah kasus untuk keputusan Layak, dan *entropy* dari semua kasus.
- 2) Lakukan perhitungan *gain* untuk setiap atribut. Hasil dari *data training* diketahui jumlah kasus (S) adalah 520, jumlah keputusan tidak layak (Si) adalah 155, dan jumlah keputusan layak (S2) adalah 365. Perhitungan *entropy* total dapat dihitung dengan menggunakan persamaan sebagai berikut :

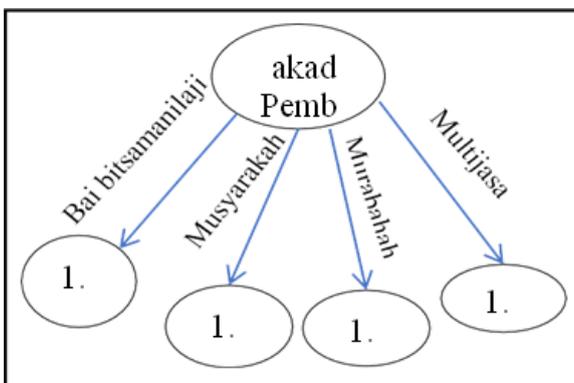
$$Entropy (Total) = \left(\frac{155}{520} * \log_2 \frac{155}{365}\right) + \left(\frac{365}{520} * \log_2 \frac{365}{520}\right)$$

Entropy(Total) = 0,878927426 Sementara itu, nilai gain pada baris atribut penghasilan dihitung dengan menggunakan persamaan sebagai berikut :

$$Gain (Total, Pendapatan) = Entropy (Total) - \sum_{i=1}^m \left(\frac{|penghasilan|}{|Total|} * Entropy (Penghasilan)\right)$$

$$Gain (Total, Pendapatan) = 0,878927426 - \left(\frac{120}{520} * 0,926819064\right) + \left(\frac{292}{520} * 0,79732651\right) + \left(\frac{98}{520} * 0,931304369\right) = 1,54837408$$

Untuk Perhitungan pada atribut- atribut berikutnya sama seperti perhitungan pada atribut pendapatan. Sehingga diperoleh hasil b diketahui bahwa atribut dengan gain tertinggi adalah Akad pembiayaan, yaitu 1,548374088 dengan demikian atribut akad pembiayaan menjadi node akar. Ada 4 nilai atribut dari pembiayaan yaitu Bai bitsamanilaji, musyarakah, murabahah, multi jasa. Dari keempat nilai tersebut dapat digambarkan pohon keputusan seperti gambar 2



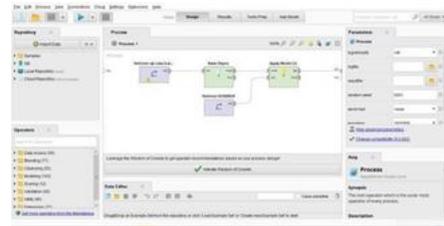
Gambar 2. Pohon keputusan node 1

4.4 Pengujian algoritma NaiveBayes

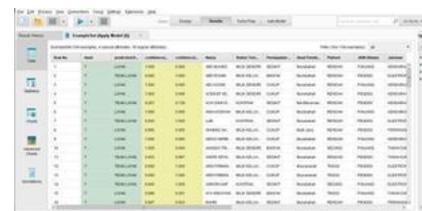
Pengujian model klasifikasi dengan *dataset* menggunakan *tools Rapid Miner* dengan 10-fold *Cross Validation*. Konfigurasi proses pengujian pada *Rapid Miner* seperti terlihat pada Gambar dibawah ini :



Gambar 3. Proses Validasi NaiveBayes



Gambar 4. Desain pengujian NaiveBayes pada Rapid Miner



Gambar 5. Hasil Prediksi Naive Bayes pada Rapid Miner

Hasil pengujian dengan menggunakan metode *Naive Bayes* Proses klasifikasi dengan *Rapid Miner* menggunakan metode *Naive Bayes* yang dilakukan untuk mengklasifikasi data kelayakan calon pembiayaan diperoleh *Accuracy, Precision, Recall. Performance vector*

1. Accuracy

Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil yaitu 78,85%. Untuk menghitung akurasinya sebagai berikut :

accuracy: 78.85% +/- 6.32% (micro average: 78.85%)			
	true LAYAK	true TIDAK LAYAK	class precision
pred LAYAK	289	34	89.47%
pred TIDAK LAYAK	76	121	61.42%
class recall	79.18%	78.06%	

Gambar 6. Accuracy algoritma Naive bayes

$$Accuracy = \frac{(TL + TT)}{(TL + FT + FL + TT)}$$

$$= \frac{(289 + 121)}{(289 + 34 + 76 + 121)} = \frac{410}{520}$$

$$= 0.788461538 * 100 = 78,84\%$$

2. Precision Naive Bayes

Precision adalah jumlah data yang True Positive (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif, sedangkan jumlah data yang True Negatif (jumlah data negatif yang dikenali secara benar sebagai negatif) dibagi dengan jumlah data yang dikenali sebagai negatif, dari hasil pengujian nilai precision yaitu 89,47% untuk class Layak dan 61,42 % untuk class Tidak.

precision: 61.72% +/- 5.18% (micro average: 61.42%) (positive class: TIDAK LAYAK)			
	True LAYAK	True TIDAK LAYAK	class precision
pred LAYAK	289	34	89.47%
pred TIDAK LAYAK	76	121	61.42%
class recall	79.18%	78.06%	

Gambar 7. Precision Naïve Bayes

$$\begin{aligned}
 \text{Precision Layak} &= \frac{(TL)}{(TL + FT)} \\
 &= \frac{289}{(289 + 34)} = \frac{289}{323} \\
 &= 0.894736842 * 100\% = 89,47\% \\
 \text{Precision Tidak layak} &= \frac{(TT)}{(TT + FL)} \\
 &= \frac{121}{(121 + 76)} = \frac{121}{197} \\
 &= 0.614213198 \\
 &* 100\% = 61,42\%
 \end{aligned}$$

3. Recall

Recall adalah jumlah data yang True Positive di bagi dengan jumlah data yang sebenarnya Layak (True Layak + False Tidak), sedangkan True Tidak di bagi dengan jumlah data yang sebenarnya Tidak (True Tidak + False Layak), untuk nilai Recall yaitu 79,18% pada Class Layak dan nilai Class Tidak yaitu 78,06%.

recall: 77.62% +/- 5.24% (micro average: 78.84%) (positive class: TIDAK LAYAK)			
	True LAYAK	True TIDAK LAYAK	class precision
pred LAYAK	289	34	89.47%
pred TIDAK LAYAK	76	121	61.42%
class recall	79.18%	78.06%	

Gambar 8. Recall Naïve Bayes

4. Performance Vector

```

PerformanceVector
confusionMatrix()
accuracy 78.85% +/- 6.32% (micro average: 78.85%)
confusionMatrix()
True LAYAK TIDAK LAYAK
LAYAK 289 34
TIDAK LAYAK 76 121
precision 89.47% +/- 5.18% (micro average: 61.42%) (positive class: TIDAK LAYAK)
confusionMatrix()
True LAYAK TIDAK LAYAK
LAYAK 289 34
TIDAK LAYAK 76 121
recall 79.18% +/- 5.24% (micro average: 78.84%) (positive class: TIDAK LAYAK)
confusionMatrix()
True LAYAK TIDAK LAYAK
LAYAK 289 34
TIDAK LAYAK 76 121
auc (optimal) 0.836 +/- 0.050 (micro average: 0.836) (positive class: TIDAK LAYAK)
auc 0.835 +/- 0.050 (micro average: 0.835) (positive class: TIDAK LAYAK)
auc (optimal) 0.834 +/- 0.051 (micro average: 0.834) (positive class: TIDAK LAYAK)
    
```

Gambar 9. Performance Vector C4.5

4.5 Pengujian algoritma C4.5

Pengujian model klasifikasi dengan dataset menggunakan tools Rapid

Miner dengan

10-fold

Cross Validation. Konfigurasi proses pengujian pada Rapid Miner seperti terlihat pada Gambar dibawah ini:



Gambar 10. Proses mining C4.5

Tabel 6. Hasil pengujian dua metode

	Pengujian Pertama	Pengujian Kedua
	Naive Bayes	Decisiontree C4.5
Accuracy	78,85%	81,35%
Precision	89,47%	83,50%
Recall	79,17%	91,50%

Berdasarkan hasil yang diperoleh nilai akurasi dengan menggunakan algoritma Decision tree lebih besardari metode Naive Bayes. Namun ada perbedaan pada nilai recall dan precisiom dari kedua algoritma seperti yang ada pada Tabel 4.2 Masih harus dilakukan penelitian yang lebih lanjut untuk dapat memudahkan dalam melakukan mining data. Dengan tingkat akurasi yang baru mencapai 81,35%.

5. Penutup

Berdasarkan hasil penelitian yang telah dilakukan dalam menganalisa kelayakan pembiayaan untuk mencegah timbulnya kredit macet dengan metode Naive Bayes dan algoritma C4.5 dengan menggunakan dataset History pembiayaan maka dapat diambil kesimpulan bahwa penggunaan Data mining dengan metode Algoritma C4.5 dapat menganalisa dan memprediksi kelayakan lebih cepat dan akurat. Tingkat akurasi pengujian menggunakan algoritma C4.5 lebih tinggi dari algoritma Naive Bayes, tingkat accuracy algoritma C4.5 sebesar 81,35% sedangkan Naive Bayes 78,85%.

Daftar Pustaka

- [1] A. Novandya and I. Oktria, "Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi," vol. 6, pp. 98–106, 2017.
- [2] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," J. Ilm. Ilmu-Ilmu Tek., vol. 8, no. 1, pp. 13–19, 2018.
- [3] N. Yahya et al., "Komparasi Kinerja Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaan mahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," Prosidng SENDI, no. 2014, pp. 978–979, 2019.
- [4] F. Tahani, M. Yunus, H. S. Dahlan, and B. Santoso, "SPK Pemilihan Calon Pendoron Darah Potensial," vol. 8, no. 1, pp. 47–54, 2014.
- [5] P. Bidang, K. Sains, Y. Mardi, J. Gajah, M. No, and S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD). Jurnal Edik Informatika."
- [6] C. Algoritma, "DATA MINING UNTUK

- MEMPREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN,” vol. 1, no. 2, pp. 32–37, 2017.
- [7] D. Yunita, “PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN DECISION TREE UNTUK PENENTUAN RISIKO KREDITKEPEMILIKAN MOBIL,” pp. 103–107.
- [8] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung, 2017.
- [9] D. Iskandar and Y. K. Suprpto, “Perbandingan Akurasi Klasifikasi Tingkat,” *J. Ilm. NERO*, vol. 2, no. 1, pp. 37–43, 2015.
- [10] R. Ilyas, “Konsep Pembiayaan Dalam Perbankan Syariah,” *J. Penelit.*, vol. 9, no. 1, pp. 183– 204, 2015.
- [11] K. Keuangan, “Undang-undang Republik Indonesia No 21 Tahun 2008,” Kementerian Keuangan, 2008. [Online]. Available: <https://jdih.kemenkeu.go.id/fullText/2008/21TAHUN2008UU.H TM>. [Accessed: 14-Jan-2020].
- [12] [Accessed: 14-Jan-2020].
- [13] S. Maiwati and R. Widayati, “Aktivitas Pemberian Kredit Komersil Pada Bank Nagari Cabang Sijunjung,” *J. Keuang. dan Perbank.*, p. Hal. 1- 12, 2019.
- [14] “Undang-Undang No 25 Tahun 1992,” Kementerian Keuangan. [Online]. Available: <https://jdih.kemenkeu.go.id/fullText/1992/25TAHUN~1992UU.h tm>. [Accessed: 14-Jan-2020].