



PENERAPAN DATA MINING UNTUK MENGGKLASIFIKASI NASABAH BANK SUMSEL BABEL MENGGUNAKAN ALGORITMA C4.5

Reliya Indah Okta Limabri¹, Ferry Putrawansyah², Alfis Arif³

¹²³Program Studi Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam

e-mail: *¹reliyaindah0512@gmail.com, ²feypuawansyah@gmail.com, ³alfisarif@yahoo.com

*Corresponding Author

Abstrak

Penelitian bertujuan untuk mengetahui penerapan data mining untuk mengklasifikasi pola nasabah debitur menggunakan algoritma C4.5 pada Bank Sumsel Babel Cabang Pagar Alam. Penelitian ini dilatarbelakangi dengan tingginya minat masyarakat terhadap pinjaman kredit usaha rakyat (KUR), maka pengajuan berkas pinjaman KUR yang banyak membuat account officer menghabiskan banyak waktu melihat keadaan calon debitur ke lapangan yang belum tentu memenuhi prediksi nasabah lancar. Maka dinilai perlu adanya suatu sistem yang akan membantu *account officer* dalam meningkatkan pelayanan pinjaman KUR dengan data mining. Pada penelitian ini akan digunakan teknik klasifikasi data mining dengan model algoritma C4.5 dalam melakukan proses analisis. Data calon debitur diolah menggunakan aplikasi *RapidMiner* dan bahasa pemrograman *Python*, metode pengembangan sistem dalam penelitian ini menggunakan metode *Rapid Application Development* (RAD). Untuk metode pengujian sistem menggunakan *Blackbox Testing*. Pada algoritma C4.5 akan dihitung *entropy* dan *gain ratio* yang akan menjadi analisis klasifikasi nasabah lancar dan nasabah macet yang berbentuk pohon keputusan (*decision tree*). Hasil dari penelitian ini adalah pola nasabah debitur KUR pada tahun 2022 dari total seluruh data 1391 nasabah dapat disimpulkan bahwa jumlah nasabah terbanyak berada di plafon kredit > Rp. 9.000.000 (diatas sembilan juta) dengan jangka waktu 12 bulan terdapat sebanyak 834 nasabah (sampel). Berdasarkan hasil pengolahan data *python* didapatkan bahwa jangka waktu 12 bulan mendapatkan *precision* sebanyak 80%, *recall* 100% dan *f1-score* 89% sedangkan untuk jangka waktu 24 bulan mendapatkan *precision* sebanyak 100%, *recall* 1% dan *f1-score* 2%. Nilai akurasi yang didapatkan sebesar 80%.

Kata kunci : Data Mining, Klasifikasi, Algoritma C4.5, Kredit.

Abstract

The study aims to determine the application of data mining to classify patterns of debtor customers using the C4.5 algorithm at the Pagar Alam Branch of the Sumsel Babel Bank. The background of this research is the high public interest in people's business credit loans (KUR), so that many submissions of KUR loan files make account officers spend a lot of time looking at the condition of prospective debtors to the field who do not necessarily meet current customer predictions. So it is considered necessary to have a system that will assist account officers in improving KUR loan services with data mining. In this study, data mining classification techniques will be used with the C4.5 algorithm model in carrying out the analysis process. Prospective debtor data is processed using the *RapidMiner* application and the *Python* programming language, the system development method in this study uses the *Rapid Application Development* (RAD) method. For system testing methods using *Blackbox Testing*.



In the C4.5 algorithm, the entropy and gain ratio will be calculated which will be an analysis of the classification of current customers and non-performing customers in the form of a decision tree. The results of this study are the pattern of KUR debtor customers in 2022. Out of a total of 1391 customer data, it can be concluded that the largest number of customers are in the credit ceiling > Rp. 9,000,000 (above nine million) with a 12-month period, there are 831 customers (sampel). Based on the results of python management, the data for a period of 12 months gets a precision of 80%, recall of 100% and f1-score of 89% while for a period of 24 months it gets a precision of 100%, recall of 1% and f1-score of 2% and an accuracy of 80%.

Keywords : Data Mining, Classification, Algorithm C4.5, Credit.

I. PENDAHULUAN

Perusahaan perbankan menerapkan *data mining* adalah untuk memecahkan masalah keuangan, ataupun korelasi antara lapangan dan internal yang informasinya tidak cepat diketahui oleh manajer karena data yang sangat besar. Penerapan data mining yang sering digunakan perbankan adalah tentang mengklasifikasi pola nasabah debitur yang akan melakukan pinjaman. Perbankan memiliki data riwayat nasabah debitur di tahun-tahun sebelumnya dengan berbagai kriteria dan bisa mengklasifikasi nasabah lancar atau macet setelah debitur menyelesaikan kewajibannya membayar cicilan tiap bulan sesuai kesepakatan,

Berdasarkan penelitian yang dilakukan oleh (Parida, 2018) dalam judul “Penerapan Data Mining Klasifikasi Nasabah Menggunakan Algoritma C4.5 pada PT. BPR Syariah Kotabumi” bahwa penerapan data mining klasifikasi nasabah menggunakan algoritma C4.5 pada PT. BPR Syariah Kotabumi menghasilkan akurasi yang tinggi yaitu sebesar 77,86%. Dengan menggunakan teknik ID3 Klasifikasi Algoritma C4.5 tersebut menghasilkan *rules* yang dapat diterapkan kedalam aplikasi untuk mengetahui pola kredit nasabah apakah tergolong klasifikasi lancar ataupun macet sehingga mempermudah pihak bank dalam memprediksi pola kredit nasabah. Penelitian tersebut selaras dengan peneliti

yaitu objek perbankan dan dengan algoritma yang sama yaitu algoritma C4.5.

Data Mining merupakan proses analisa data dari sudut yang berbeda dan mengolahnya menjadi informasi-informasi penting yang bisa digunakan untuk meningkatkan keuntungan. Secara teknis, *data mining* dapat disebut juga sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field*. Pada data mining dapat digunakan metode *decision tree* untuk melakukan klasifikasi. (Donny & Raden, 2019)

Klasifikasi merupakan jenis teknik pengolahan data yang digunakan untuk memisahkan sebuah objek menjadi sejumlah kelas yang berbeda sesuai dengan jumlah kelas yang diharapkan. Selain itu, teknik ini juga digunakan untuk menemukan pola yang dapat memisahkan kelas data satu sama yang lain untuk menentukan objek yang termasuk dalam kategori tertentu, yaitu membandingkannya dengan ciri-ciri variabel kelompok yang telah ditentukan. (Romli & Zy, 2020)

Pohon keputusan atau yang lebih dikenal dengan istilah *Decision Tree* ini merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah-masalah tersebut dengan memperhitungkan berbagai macam faktor yang berkaitan di dalam lingkup masalah tersebut. (D. Aprilia, 2013)

Algoritma C4.5 pada dasarnya adalah model yang digunakan untuk membantu

menyusun struktur pohon keputusan sesuai dengan pemilihan atribut dengan tingkat prioritas paling tinggi atau nilai dari gain tertinggi sesuai dengan nilai entropy atribut yang dijadikan sebagai proses atribut klasifikasi. Berikut adalah penjelasan tentang tahapan proses umum dalam membangun sebuah struktur dari pohon keputusan yaitu (David, 2022):

1. Tahapan awal yang perlu dilakukan yaitu melakukan pememilihan atribut atau deskripsi data yang akan dijadikan sebagai akar dari pohon keputusan.
2. Setelah itu dilanjutkan dengan membuat struktur cabang dari pohon keputusan dengan nilai.
3. Tahapan langkah ketiga yaitu lakukan proses pembagian atau pemberian kasus permasalahan pada struktur cabang yang dimiliki oleh pohon keputusan.
4. Dan pada langkah terakhir yang perlu dilakukan adalah mengulangi proses dari langkah pertama sampai akhir bagi setiap cabang hingga kasus permasalahan pada cabang mempunyai kelas atau kategori yang sama.

Kredit merupakan dana yang diberikan oleh bank kepada pihak lain berdasarkan perjanjian pinjam-meminjam, yang mewajibkan peminjam melunasi pinjamannya setelah jangka waktu tertentu. (Ibnu Ubaedi, 2022)

Berdasarkan penelitian yang dilakukan oleh (Asmira, 2019) dalam judul “ Penerapan *Data Mining* untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4.5 pada Bank Bri Unit Anduonohu Kendari” yaitu Seringkali pembayaran kredit yang macet membuat bank kerepotan, karena semakin banyaknya nasabah yang membayar secara macet dapat berdampak buruk untuk kesehatan bank. Oleh karena itu, data nasabah dari Bank BRI Unit Anduonohu Kendari menjadi bahan acuan untuk

menganalisa pola nasabah pemohon kredit. Pemohon kredit termasuk dalam kategori lancar, atau macet. Untuk mengatasi masalah tersebut, dapat digunakan metode *decision tree*. Sehingga penelitian ini bisa dijadikan acuan pihak Bank untuk menilai nasabah dengan *record* data yang ada untuk pengambilan kredit selanjutnya. Informasi yang ditampilkan berupa tingkat akurasi data nasabah lancar dan macet. Dalam penelitian ini, *decision tree* menggunakan bahasa pemrograman java. Kemudian hasil akurasi dari aplikasi yang diimplementasikan akan dibandingkan dengan hasil menggunakan *software rapidminer*. Sehingga diperoleh akurasi dengan *decision tree* sebesar 89,5%. Penelitian ini relevan dengan peneliti karena objek yang dipakai di perbankan, menggunakan algoritma C4.5 dan juga aplikasi *RapidMiner*.

Dari penelitian di atas akan menjadi referensi bagi peneliti untuk bisa membuat sistem penerapan algoritma C4.5 dalam menentukan klasifikasi pola nasabah debitur macet dan debitur lancar yang terkomputerisasi dan akan diterapkan khususnya di Bank Sumsel Babel Cabang Pagar Alam dengan menggunakan *Data Mining*. Tujuannya adalah untuk membantu mempermudah analisa *Account Officer* karena metode algoritma C4.5 menghasilkan pohon keputusan dengan dua pilihan sehingga terbentuk pola nasabah debitur lancar dan debitur macet, maka lebih efektif dan efisien bagi *Account Officer* langsung bisa membuat keputusan debitur tersebut akan disurvei atau tidak. Tentunya hal tersebut berdasarkan perhitungan komputerisasi berdasarkan indikator-indikator yang dianalisa antara lain produktivitas usaha ke depan bagus atau tidak, kondisi usaha yang sedang dijalankan lancar atau tidak, jaminan harus melebihi dari *plafond* pinjaman yang diajukan, status kreditnya tidak pernah

macet dipinjam manapun, jumlah pinjaman berdasarkan analisa laba yang dihasilkan dan jangka waktu pinjaman sesuai dengan kemampuan debitur serta status pernikahan dan jumlah tanggungan.

II. METODOLOGI PENELITIAN

Penelitian dilakukan di Bank Sumsel Babel Cabang Pagar Alam Jalan Kapten Sanap No. 41 Kelurahan Beringin Jaya, Kecamatan Pagar Alam Utara, Kota Pagar Alam.

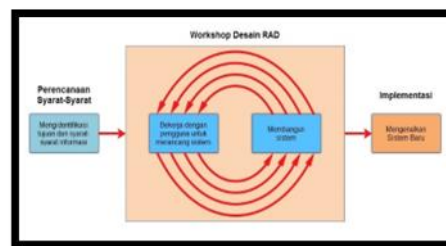
1. Metode pengumpulan data

- Observasi yang dilakukan yaitu dengan pengamatan secara langsung pada objek penelitian yang dalam hal ini berkaitan dengan Unit Kredit pada Bank Sumsel Babel Cabang Pagar Alam.
- Wawancara Merupakan metode pengumpulan data yang dilakukan secara langsung dari *Account Officer* yaitu pelaksana dari KUR yang tepat dan diteliti oleh Penyelia Unit Kredit Bank Sumsel Babel Cabang Pagar Alam.
- Dokumentasi. Penelitian dilakukan dengan menggunakan data yang telah diperoleh dari arsip ataupun yang lainnya.
- Studi Pustaka yaitu dimana peneliti mempelajari teori-teori baik dari buku cetak maupun non cetak, dan jurnal penelitian yang berhubungan dengan objek penelitian.

2. Model Pengembangan Sistem

Penelitian ini menggunakan metode *Rapid application development* (RAD) yang merupakan model proses pengembangan perangkat lunak yang bersifat incremental terutama untuk waktu pengerjaan yang pendek. Model RAD adalah adaptasi dari model air terjun versi kecepatan tinggi dengan menggunakan model air terjun untuk

pengembangan setiap komponen perangkat lunak (Rosa A.S, 2018).



Gambar 1. Model Pengembangan Sistem

1. *Requirements Planning* (Perencanaan Persyaratan)

Untuk mengidentifikasi tentang kebutuhan, batasan dan objektifitas dari sistem yang akan dibangun, dengan cara mengumpulkan data dari *stakeholder*. Aktivitas yang dilakukan dengan melakukan pengamatan langsung dan mengumpulkan data dari buku-buku dan dari jurnal-jurnal yang menunjang dan relevan. Hasil yang didapatkan berupa prosedur pengambilan data penelitian dan spesifikasi kebutuhan sistem.

2. *RAD Design Workshop*

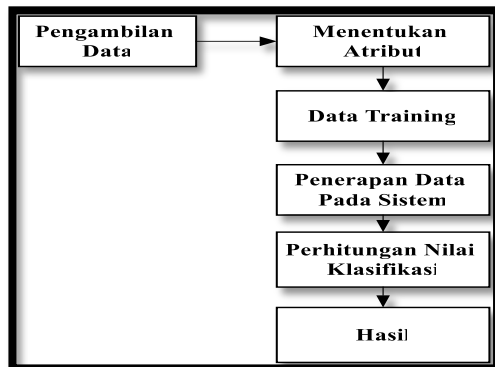
Untuk merancang semua kegiatan dalam arsitektur sistem secara keseluruhan dengan melibatkan identifikasi dan deskripsi abstraksi sistem perangkat lunak yang mendasar dan hubungan-hubungannya.

3. *Implementation* (Penerapan)

Untuk mengimplementasikan metode, program sesuai dengan kebutuhan sistem. Aktivitas yang dilakukan yaitu menentukan implementasi Basis Data, Pemrograman, antarmuka, dan Pengujian.

untuk mengetahui nasabah memiliki nilai lancar.

3. Metode Klasifikasi Algoritma C4.5



Gambar 2. Alur Analisis Data

Berdasarkan alur analisis data di atas maka tahapan yang akan digunakan dalam penelitian ini adalah sebagai berikut:

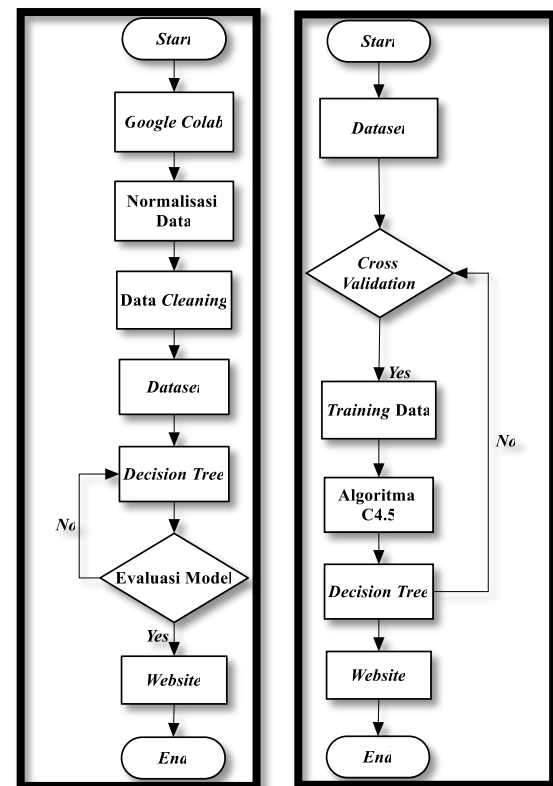
- Melakukan pengambilan data dari data nasabah kredit KUR subsidi bunga 0% pemerintah kota Pagar Alam.
- Menentukan atribut yang digunakan dalam penelitian untuk dilakukan perhitungan.

Tabel 1. Atribut Penelitian

Atribut	Sub Atribut	Nilai
Ideb (Status Kredit)	Pernah	5
	Tidak Pernah	2
Analisis Usaha	Sangat Lancar	5
	Lancar	4
	Cukup Lancar	3
	Tidak Lancar	2
Jaminan	Ada	5
	Tidak Ada	2
Jumlah Pinjaman	30-50 Juta	2
	15-30 Juta	3
	< 15 Juta	5
Tempat Tinggal	Milik Sendiri	5
	Sewa	2

- Data yang digunakan untuk perhitungan yaitu berupa sampel sebanyak 834 data.
- Hasil akurasi dari *data training* dengan hasil yang dari sistem yang dipakai

4. Flowchart Alur Kerja Pengolahan Data Python dan RapidMiner



Gambar 3. (a) Flowchart Python, (b) Flowchart RapidMiner

Alur kerja pada *flowchart Python* yaitu:

- Membuka web browser *Google Colab*.
- Melakukan normalisasi data mentah yang di impor dengan format csv.
- Data dibersihkan dari data kotor atau data yang kurang lengkap.
- Data yang telah dibersihkan akan menjadi *dataset* yang akan diolah.
- Membuat model *data mining* memakai fitur *machine learning* dengan memilih metode *decision tree*.
- Setelah itu melakukan evaluasi model yang telah dibuat untuk mengecek akurasi, presisi, sensitivitas, dan spesifitas dari model.

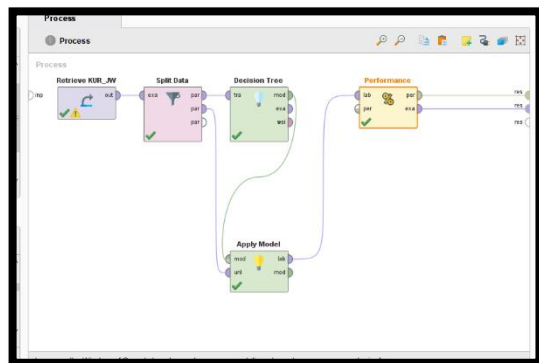
7. Kemudian tahap terakhir adalah merepresentasikan hasil model ke dalam *website*.

Sedangkan alur kerja pada *flowchart RapidMiner* yaitu:

1. Siapkan *dataset*.
2. Mengecek kebenaran *dataset*.
3. Jika data sudah benar maka akan dibentuk training data yang terdiri dari atribut dan labelnya.
4. Setelah itu pada algoritma C4.5 dihitung nilai *entropy* dan *gain ratio*.
5. Maka akan terbentuk sebuah pohon keputusan. Jika pohon keputusan *error* maka harus dicek lagi ke *cross validation*,
6. Tahap terakhir jika pohon keputusannya benar maka dapat kita implementasikan ke *website*.

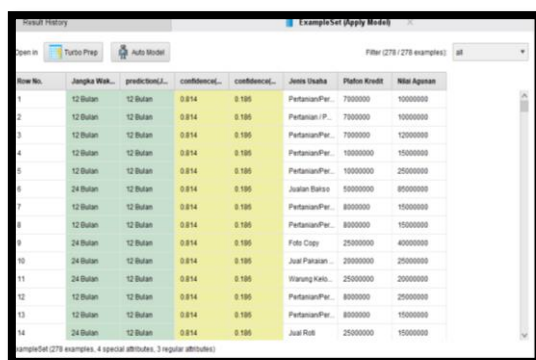
III. HASIL DAN PEMBAHASAN

1. Hasil dari *RapidMiner*



Gambar 4. Desain Data *RapidMiner*

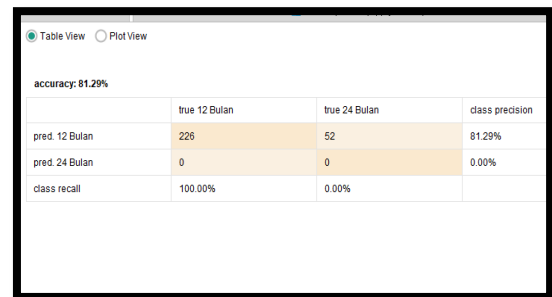
Gambar 4 menunjukkan hasil desain data yang dilakukan pada aplikasi *RapidMiner*.



Row No.	Jangka Waktu	predictions	confidence	confidence	Jenis Usaha	Papan Kredit	Nilai Agunan
1	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	7000000	10000000
2	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	7000000	10000000
3	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	7000000	12000000
4	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	10000000	15000000
5	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	10000000	25000000
6	24 Bulan	12 Bulan	0.814	0.190	Jualan Batas	10000000	80000000
7	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	8000000	10000000
8	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	8000000	10000000
9	24 Bulan	12 Bulan	0.814	0.190	Fito Cggy	25000000	40000000
10	24 Bulan	12 Bulan	0.814	0.190	Jual Peranan	20000000	25000000
11	24 Bulan	12 Bulan	0.814	0.190	Warung Kaki	25000000	20000000
12	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	8000000	25000000
13	12 Bulan	12 Bulan	0.814	0.190	PertanianPer	8000000	15000000
14	24 Bulan	12 Bulan	0.814	0.190	Jual Roll	25000000	15000000

Gambar 5. Hasil *RapidMiner*

Dari gambar 5 menunjukkan kesimpulan dari total 1391 data yang dimasukkan atau data yang diolah didapatkan hasil pengelompokan berdasarkan jangka waktu dan hasil prediksi jangka waktu berdasarkan nilai *confidence* jangka waktu 12 bulan dan jangka waktu 24 bulan.



accuracy: 81.29%			
	true 12 Bulan	true 24 Bulan	class precision
pred. 12 Bulan	226	52	81.29%
pred. 24 Bulan	0	0	0.00%
class recall	100.00%	0.00%	

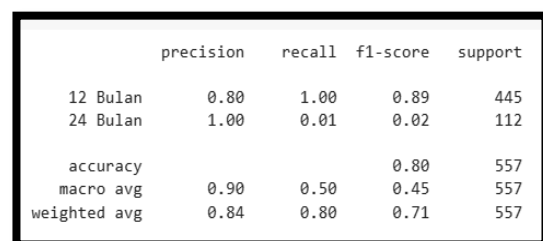
Gambar 6. *Performance*

Gambar 6 menunjukkan performa *vector* yang didapatkan dari hasil klasifikasi dari performa *vector* di dapatkan hasil akurasi 81,29% dan didapatkan *confusion matrik* sebagai berikut;

Tabel 2. *Confusion matrik*

True	12 Bulan	24 Bulan
12 Bulan	226	55
24 Bulan	0	0

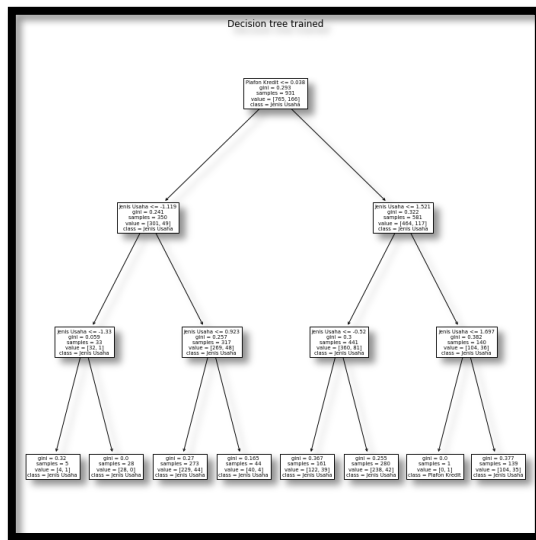
2. Hasil *Python*



	precision	recall	f1-score	support
12 Bulan	0.80	1.00	0.89	445
24 Bulan	1.00	0.01	0.02	112
accuracy			0.80	557
macro avg	0.90	0.50	0.45	557
weighted avg	0.84	0.80	0.71	557

Gambar 7. Laporan Klasifikasi

Pada gambar 7 dapat dilihat bahwa untuk jangka waktu 12 bulan mendapatkan *precision* sebanyak 80%, *recall* 100% dan *f1-score* 89% sedangkan untuk jangka waktu 24 bulan mendapatkan *precision* sebanyak 100%, *recall* 1% dan *f1-score* 2% serta didapatkan akurasi sebesar 80%.



Gambar 8. Decision tree trained

IV. KESIMPULAN

Berdasarkan hasil dari penelitian yang dilakukan maka didapatkan kesimpulan sebagai berikut;

1. Dari total keseluruhan data sebanyak 1391 data nasabah dapat disimpulkan bahwa jumlah nasabah terbanyak berada di plafon kredit > Rp.9.000.000 (diatas sembilan juta) dengan jangka waktu 12 bulan terdapat sebanyak 834 sampel data nasabah.
2. Berdasarkan hasil pengelolaan *python* didapatkan data jangka waktu 12 bulan mendapatkan *precision* sebanyak 80%, *recall* 100% dan *f1-score* 89% sedangkan untuk jangka waktu 24 bulan mendapatkan *precision* sebanyak 100%, *recall* 1% dan *f1-score* 2% serta didapatkan akurasi sebesar 80%.

V. SARAN

Saran peneliti untuk peneliti lain adalah sebagai berikut;

1. Data akurasi yang di dapatkan sudah cukup baik namun diharapkan bisa lebih baik lagi jika nilai akurasi dapat lebih di tingkatan

2. Dapat menggunakan algoritma yang lain atau mengkomparasi antara beberapa algoritma
3. Pengelolaan data yang lebih besar mungkin dapat mencakup wilayah yang lebih besar, dan data yang digunakan data yang lebih banyak mungkin 2 atau 3 tahun.

VI. DAFTAR PUSTAKA

- Adela Ainun D, Zulfi A., Andy S. (2022). *Penerapan Data Mining Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4.5 Pada Pegadaian Tanjung Balai*. *Jurnal Teknologi dan Sistem Informasi*. Vol 2 No. 1. 1-6.
- Anggraini, Y. . P. D. . & D. D. (2020). *Sistem Informasi Penjualan Sepeda Berbasis Web Menggunakan Framework Codeigniter*. *Jurnal Teknologi Dan Sistem Informasi*. 1(2), 64–70.
- Asmira, A. (2019). *Penerapan Data Mining Untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4. 5 Pada Bank Bri Unit Anduonohu Kendari*. Router Research. <https://Www.Ejournal.Stipwunarah.A c.Id/Index.Php/Router/Article/View/316>
- David, D. (2022). *Analisis Kelayakan Menerima Pinjaman Kredit Dengan Algoritma C4. 5 Pada PT BPR Buana Arta Mulia*. Repository.Upbatam.Ac.Id. [Http://Repository.Upbatam.Ac.Id/1480/](http://Repository.Upbatam.Ac.Id/1480/)
- Ibnu Ubaedi, Yan Mitha J. (2022). *Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection dan Stratified Sampling untuk Prediksi Kelayakan Kredit*. *Jurnal Sistem Informasi*, Vol 9.17-26.
- Jiawei, H., & Micheline, K. (2006). *Data Mining: Concepts And Techniques Second Edition*. Simon Fraser University. Morgan.
- Maulana, D., & Putra, R. R. B. (2019).



- Penerapan Data Mining Untuk Memprediksi Pola Nasabah Menggunakan Algoritma C4. 5.* Jurnal Sigma.
<https://Jurnal.Pelitabangsa.Ac.Id/Index.Php/Sigma/Article/View/587>
- Nawary, A. P., & Kurniati, K. (2021). *Penerapan Data Mining Dalam Memprediksi Kelancaran Kredit Nasabah Menggunakan Algoritma C4. 5 (Studi Kasus Pada Pt. Astra Bina Darma Conference On* <https://Conference.Binadarma.Ac.Id/Index.Php/Bdccc/Article/View/2881>
- Parida, M. (2018). *Penerapan Data Mining Klasifikasi Nasabah Menggunakan Algoritma C4. 5 Pada Pt. Bpr Syariah Kotabumi.* Prosiding Seminar Nasional Darmajaya.
<https://Jurnal.Darmajaya.Ac.Id/Index.Php/psnd/Article/View/1264>
- Ridho, A., & Niani, C. R. (2022). *Implementasi Enkripsi Dengan Vigenere Cipher Dan Reverse Cipher Menggunakan Bahasa Pemrograman Python.* 9–15.
- Rosa A.S, M. . (2018). *Rekayasa Perangkat Lunak.* Bandung: Informatika.
- Sibero. (2013). *Web Programing Power Pack.* Yogyakarta: MediaKom.
- Tahir, M. A. (2019). *Perancangan Aplikasi Data Mining Menggunakan Metode Decision Tree Untuk Analisa Pemberian Kredit Pada Bri Unit Lalabata Rilau.* Jurnal Ilmiah Sistem Informasi Dan Teknik
<https://Journal.Jisti.Unipol.Ac.Id/Index.Php/Jisti/Article/View/23>