



Algoritma Decision Tree Untuk Klasifikasi Calon Debitur LPD Desa Adat Anggungan

I Gusti Ayu Mirah Purnama Dewi¹, Wayan Gede Suka Parwita², I Made Dedy Setiawan³
^{1,2,3}Program Studi Teknik Informatika, STMIK STIKOM Indonesia, Denpasar - Bali, Indonesia
E-mail: ¹mirahgek201@gmail.com, ²gede.suka@stiki-indonesia.ac.id, ³dedy.setiawan86@yahoo.com

Received on 13 September 2021	Revised on 23 September 2021	Accepted on 25 September 2021
----------------------------------	---------------------------------	----------------------------------

Abstract

LPD Desa Adat Anggungan is a business entity engaged in financial services. The Anggungan Traditional Village LPD serves credit for customers. In providing credit services, of course there are many big risks faced, one of which is bad credit. To be able to make the right decisions in providing credit lending, classification is required based on criteria according to standards. There fore this research was conducted to help solve this problem, by designing a data mining application to classify credit customers. This classification aims to make it easier for the LPD to extend credit or not. In this study, the data mining technique applied is classification, while the method used is the Decision Tree, C4.5 Algorithm. The result of this study is an application that helps the credit department at the LPD to reduce errors in making decisions when extending credit.

Keywords: *Classification, Decision Tree, C4.5 Algorithm*

Abstrak

LPD Desa Adat Anggungan merupakan salah satu badan usaha yang bergerak dibidang jasa keuangan. LPD Desa Adat Anggungan melayani perkreditan untuk nasabah. Dalam memberikan layanan kredit tentunya ada banyak resiko besar yang dihadapi, salah satunya kredit macet. Untuk dapat mengambil keputusan yang tepat dalam memberikan peminjaman kredit diperlukan pengklasifikasian berdasarkan kriteria yang sesuai standar. Oleh karena itu penelitian ini dilakukan untuk membantu menyelesaikan masalah tersebut, dengan merancang suatu aplikasi data mining untuk mengklasifikasikan nasabah kredit. Klasifikasi ini bertujuan untuk memudahkan pihak LPD dalam memberikan perpanjangan kredit atau tidak. Dalam penelitian ini teknik data mining yang diterapkan adalah klasifikasi sedangkan metode yang digunakan adalah Decision Tree, Algoritma C4.5. Hasil dari penelitian ini adalah sebuah aplikasi yang membantu bagian kredit pada LPD untuk lebih mengurangi kesalahan dalam mengambil keputusan saat perpanjangan kredit.

Kata Kunci: *Klasifikasi, Pohon Keputusan, Algoritma C4.*

PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi, mengakibatkan banyaknya terjadi perubahan di segala bidang. Lembaga Perkreditan Desa sebagai lembaga keuangan yang berfungsi menyimpan dana dari masyarakat dan menyalurkannya dalam bentuk pinjaman, LPD berusaha memberikan kemudahan dalam setiap layanan yang diberikannya, misalnya saja dalam layanan memberikan kredit bagi seorang debitur.

LPD Desa Adat Anggungan merupakan salah satu LPD yang berada di Desa Adat Anggungan, Lukluk, Kabupaten Badung yang memberikan layanan kredit/pinjaman kepada masyarakat. Dalam memberikan layanan kredit tentunya ada banyak resiko besar yang dihadapi, maka dari itu banyak hal yang dipertimbangkan dengan memperhatikan asas-asas perkreditan. Pihak LPD dalam memberikan kredit saat ini belum memiliki pertimbangan yang tepat, jika ada yang mengajukan diterima dengan memenuhi persyaratan. Selain itu, saat mengevaluasi permintaan kredit harus di analisis dengan teliti dan melakukan penilaian melalui suatu prosedur terhadap kondisi calon debitur yang diperkirakan dapat mempengaruhi kemampuan mereka dalam memenuhi kewajibannya dan untuk mengurangi resiko terjadinya “kredit macet” dalam pemberian kredit.

Untuk mengurangi risiko tersebut, jaminan pemberian kredit dalam arti keyakinan atas kemampuan dan kesanggupan debitur untuk melunasi hutangnya sesuai dengan yang diperjanjikan merupakan faktor yang sangat penting yang harus diperhatikan. Oleh karena itu untuk dapat mengambil keputusan yang tepat dalam memberikan peminjaman kredit diperlukan pengklasifikasian berdasarkan kriteria-kriteria yang sesuai standar. Salah satu yang dapat digunakan untuk mengklasifikasi kriteria nasabah yaitu analisis *Data Mining*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat. Analisis ini dapat membantu dalam mengambil keputusan dan untuk menganalisis target pasar sehingga dapat memunculkan peluang-peluang baru yang nantinya dapat meningkatkan keuntungan dan mengetahui nasabah-nasabah yang berpotensi melakukan kredit. Sehingga pihak LPD mudah untuk menetapkan kredit yang ditolak atau kredit yang diterima aplikasinya berdasar data-data yang sudah diolah tersebut.

Adapun teknik *data mining* yang diterapkan pada aplikasi yang dibangun adalah klasifikasi sedangkan metode klasifikasi yang digunakan adalah *Decision Tree* (pohon keputusan). Algoritma yang dipakai sebagai algoritma pembentuk pohon keputusannya adalah Algoritma *C4.5 Decision tree*. Algoritma C4.5 dipilih untuk membangun pohon keputusan karena secara rekursif mengunjungi setiap simpul keputusan, memilih cabang optimal, sampai tidak ada cabang lagi yang mungkin dihasilkan menurut [1]. Oleh karena itu, penelitian ini fokus pada pemanfaatan *data mining* untuk memprediksi kriteria nasabah kredit, sehingga dapat diketahui apakah nasabah yang bersangkutan merupakan nasabah yang berpotensi menjadi nasabah kredit yang produktif atau tidak.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah Teknik Data Mining Klasifikasi metode *Decision Tree* dengan Algoritma C4.5.

KLASIFIKASI

Menurut [2], Klasifikasi adalah proses menemukan model (atau fungsi) yang akan mengelompokkan kelas data sehingga dapat memprediksi kelas obyek yang tidak diketahui. Model yang diperoleh bias diwakili diberbagai bentuk, seperti klasifikasi berbentuk aturan IF-THEN, pohon keputusan, rumus matematika, atau jaringan saraf tiruan. *Rule based classification* adalah model berbentuk seperangkat aturan berbentuk IF-THEN. Aturan berbentuk IF-THEN adalah cara yang baik untuk merepresentasikan pengetahuan. Aturan dapat diekstraksi dari pohon keputusan.

ALGORITMA C4.5

Menurut [2], Algoritma C4.5. merupakan kelompok algoritma pohon keputusan (*decision tree*). Algoritma ini mempunyai input berupa training *samples* dan *samples*. Training *samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data. Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan training data yang telah disediakan.

Secara garis besar langkah-langkah yang dilakukan oleh Algoritma C.45 dalam membentuk pohon keputusan adalah sebagai berikut:

Input: sample training, label training, atribut

1. Membuat simpul akar untuk pohon yang dibuat
2. Jika semua sampel positif, berhenti dengan suatu pohon dengan satu simpul akar, beri tanda (+)
3. Jika semua sampel negatif, berhenti dengan suatu pohon dengan satu simpul akar, beri tanda (-)
4. Jika atribut kosong, berhenti dengan suatu pohon dengan satu simpul akar, dengan label sesuai nilai yang terbanyak yang ada pada label training

Output: Decision Tree.

Mengubah tree yang dihasilkan dalam beberapa rule. Jumlah rule sama dengan jumlah path yang mungkin dapat dibangun dari root sampai leaf node. Tree Praining dilakukan untuk menyederhanakan tree sehingga akurasi dapat bertambah.

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

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

DECISION TREE

Menurut [3] *Decision Tree* mengklasifikasikan sampel secara *top down*, mulai dari simpul akar dengan menjaga jarak dengan hasil dari tes node internal, sampai simpul daun yang dicapai oleh kelas label yang

ditugaskan. *Decision tree* dengan algoritma C45 merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Pada *decision tree* terdapat 3 jenis node, yaitu:

1. *Root Node*, merupakan node paling atas, pada node ini tidak ada *input* dan bias tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal Node*, merupakan node percabangan, pada node ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf node* atau terminal node, merupakan node akhir, pada node ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Menurut [4] Node adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *gain* suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut:

$$Gain(S, A) = Entropy(s) - \sum \frac{|S_i|}{|S|} + Entropy(S_i) \quad (1)$$

Keterangan:

S = Himpunan Kasus A=Atribut n=jumlah partisi atribut A |Si|= Proporsi Si terhadap S

|S| = jumlah kasus dalam S

Sementara itu, menurut [4] untuk menghitung nilai *Entropy* dapat dilihat pada persamaan berikut ini:

$$Entropy(S) = \sum -p_i * \log_2 p_i \quad n \ i=1 \quad (2)$$

Keterangan: S=Himpunan Kasus n=jumlah partisi S Pi = proporsi Si terhadap S

CONFUSION MATRIX

Menurut [9] *Confusion matrix* memberikan keputusan yang diperoleh dalam *training* dan *testing*, *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah. *Confusion matrix* berisi informasi aktual (*actual*) dan informasi terprediksi (*predicted*) pada sistem klasifikasi. *Confusion Matrix* merupakan sebuah metode untuk evaluasi yang menggunakan tabel *matrix* seperti pada tabel 1. Pada Tabel tersebut dapat kita lihat bahwa jika *dataset* terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif. Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*.

Tabel 1: Tabel Confusion Matrix

Correct Classification	Classified as	
	+	-
+	True_positive	False_negative
-	False_positive	True_negative





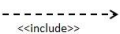
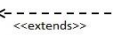
UNIFIED MODELING LANGUAGE (UML)

Menurut [10] *Unified Modeling Language* (UML) adalah bahasa spesifikasi standar yang dipergunakan untuk mendokumentasikan, menspesifikasikan dan membangun perangkat lunak.

USE CASE DIAGRAM

Menurut [11] *Use Case Diagram* merupakan pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat. *Use case* digunakan untuk mengetahui fungsi apa saja yang ada di dalam sistem informasi dan siapa saja yang berhak menggunakan fungsi-fungsi tersebut. Simbol-simbol yang digunakan dalam *Use Case Diagram* dapat dilihat pada tabel 2.








Tabel 2: Use Case Diagram

Gambar	Keterangan
	<i>Use Case</i> menggambarkan fungsionalitas yang disediakan sistem sebagai unit-unit yang bertukar pesan antar unit dengan aktor, yang dinyatakan dengan menggunakan kata kerja
	<i>Actor</i> atau Aktor adalah <i>Abstraction</i> dari orang atau sistem yang lain yang mengaktifkan fungsi dari target sistem. Untuk mengidentifikasi aktor, harus ditentukan pembagian tenaga kerja dan tugas-tugas yang berkaitan dengan peran pada konteks target sistem.
	Asosiasi antara aktor dan <i>use case</i> , digambarkan dengan garis tanpa panah yang mengindikasikan siapa atau apa yang meminta interaksi secara langsung dan bukannya mengindikasikan data.
	Asosiasi antara aktor dan <i>use case</i> yang Menggunakan panah terbuka untuk mengindikasikan bila aktor berinteraksi secara pasif
	<i>Include</i> , merupakan di dalam <i>use case</i> lain (<i>required</i>) atau pemanggilan <i>use case</i> oleh <i>use case</i> lain, contohnya adalah pemanggilan sebuah fungsi program
	<i>Extend</i> , merupakan perluasan dari <i>use case</i> lain jika kondisi atau syarat terpenuhi

Activity Diagram

Menurut [11] *Activity Diagram* menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis. Simbol-simbol yang digunakan dalam *Activity Diagram* dapat dilihat pada table 3.

Tabel 3: Activity Diagram

Gambar	Keterangan
	<i>Start Point</i> , diletakkan pada pojok kiri atas dan merupakan awal aktivitas
	<i>End Point</i> , akhir aktivitas
	<i>Activities</i> , menggambarkan suatu proses/kegiatan bisnis
	<i>Fork</i> /percabangan, digunakan untuk menunjukkan kegiatan yang dilakukan secara paralel atau untuk menggabungkan dua kegiatan paralel menjadi satu
	<i>Join</i> (penggabungan) atau <i>rake</i> , digunakan untuk menunjukkan adanya dekomposisi
	<i>Decision Points</i> , menggambarkan pilihan untuk pengambilan keputusan, <i>true</i> atau <i>false</i>
	<i>Swimlane</i> , pembagian <i>activity diagram</i> untuk menunjukkan siapa melakukan apa

Sequence Diagram

Diagram Urutan (*Sequence Diagram*), *Sequence Diagram* menggambarkan kelakuan objek pada *use case* dengan mendeskripsikan waktu hidup objek dan pesan yang dikirimkan dan diterima antar objek.

Class Diagram

Merupakan hubungan antar kelas dan penjelasan detail tiap-tiap kelas di dalam model desain dari suatu sistem, juga memperlihatkan aturanaturan dan tanggung jawab entitas yang menentukan perilaku sistem. *Class Diagram* juga menunjukkan atribut-atribut dan operasioperasi dari sebuah kelas dan *constraint* yang berhubungan dengan objek yang dikoneksikan. *Class Diagram* secara khas meliputi: Kelas (*Class*), Relasi *Assosiations*, *Generalitiation* dan *Aggregation*, atribut (*Attributes*), operasi (*operation/method*) dan *visibility*, tingkat akses objek eksternal kepada suatu operasi atau atribut. Hubungan antar kelas mempunyai keterangan yang disebut dengan *Multiplicity* atau *Cardinality*.

HASIL DAN PEMBAHASAN

ANALISIS DATA MINING

Dalam proses klasifikasi terdapat beberapa langkah pemrosesan terhadap data yaitu:

1. *Data Cleaning*, data *cleaning* merupakan suatu pemrosesan terhadap data untuk menghilangkan *noise* dan penanganan terhadap *missing value* pada suatu *record*.
2. Analisis Relevansi, pada tahap ini, dilakukan penghapusan terhadap atribut – atribut yang *redundant* ataupun kurang berkaitan dengan proses klasifikasi yang akan dilakukan. Analisis relevansi dapat meningkatkan efisiensi klasifikasi karena waktu yang diperlukan untuk pembelajaran lebih sedikit daripada proses pembelajaran terhadap data – data dengan atribut yang masih lengkap (masih terdapat redundansi).
3. Transformasi Data, pada data dapat dilakukan generalisasi menjadi data dengan level yang lebih tinggi. Misalnya dengan melakukan diskretisasi terhadap atribut dengan nilai kontinyu. Pembelajaran terhadap data hasil generalisasi dapat mengurangi kompleksitas pembelajaran yang harus dilakukan karena ukuran data yang harus diproses lebih kecil. Proses pembentukan *Decision Tree* sebagai berikut:
 - a. Mengubah bentuk data (tabel) menjadi *tree*
 - b. Mengubah model *tree* menjadi *rule*
 - c. Menyederhanakan *rule*
 - d. Apabila individu-individu suatu populasi dapat diklasifikasikan dalam dua variabel (kategori), tiap-tiap kategori dapat terdiri dari beberapa alternative.
 - e. Menentukan hasil akhir

PENENTUAN DAN PERHITUNGAN ATRIBUT

Prosedur pencarian nasabah yang dilakukan oleh Bagian Kredit LPD Desa Adat Anggunan. Berdasarkan data yang dikumpulkan maka dilakukan proses seleksi, *cleaning* dan transformasi data. Pada penerapan aslinya akan menggunakan tujuh atribut, diantaranya:

1. Kolektibilitas
2. Jenis agunan

3. Pekerjaan
4. Jumlah angsuran
5. Umur

Transformasi data dilakukan dengan mengubah tipe data numerik menjadi interval dan menginisialisasi nilai atau isian karakter terlalu panjang pada beberapa atribut. Hasil transformasi data yang dilakukan dapat dilihat pada tabel 4.

Tabel 4: Transformasi Data

Atribut	Kategori	Transformasi data
Kolektibilitas	Lancar	LA
	Kurang Lancar	KL
Jenis Agunan	Sertifikat Tanah	JA1
	BPKB	JA2
Pekerjaan	PNS	PK1
	Pegawai swasta	PK2
	Wiraswasta	PK3
Pinjaman	Kecil	PJ1
	Sedang	PJ2
	Besar	PJ3
Umur	20-35 thn	U1
	36-50 thn	U2
	51-65 thn	U3

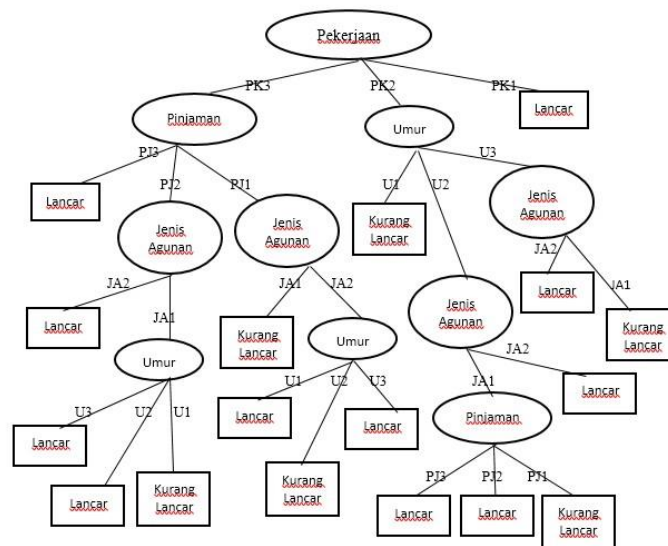
Setelah proses transformasi selesai maka dilanjutkan ke proses *data mining*, sehingga menghasilkan pohon keputusan dan aturan (*rule*). Data-data yang dikumpulkan dikelompokkan berdasarkan atribut – atribut yang ada. Data-data yang telah dikelompokkan dianalisa dengan persamaan matematik agar dapat menentukan *root* (akar). Untuk itu dilakukan proses perhitungan dan mencari nilai *Gain* dari masing - masing atribut yang ada. Jika sudah didapat nilai *Gain* dari masing-masing atribut maka diambil nilai *Gain* tertinggi sebagai *root* (akar). Hasil dari perhitungan Entropy dan Gain dapat dilihat pada tabel 5.

Tabel 5: Tabel Hasil Pehitungan Entropy dan Gain

Atribut	Jumlah	Lancar	Kurang Lancar	Entropy	Gain
Jumlah kasus	38	20	18	0,998000884	
Pekerjaan					
Wiraswata	20	10	10	1	0,109199897
Pns	3	0	3	0	
Peg swasta	15	10	5	0,918295834	
Jenis agunan					
Sert. Tanah	19	8	11	0,981940787	0,032304483
BPKB	19	12	7	0,949452015	
Pinjaman					
Kecil	22	8	14	0,945660305	0,08815951
Sedang	11	8	3	0,845350937	
Besar	5	4	1	0,721928095	
Umur					
20-35 thn	3	0	3	0	0,09976537
36-50 thn	21	13	8	0,958711883	
51-65 thn	14	7	7	1	

Dari tabel 6 Hasil Pehitungan *Entropy* dan *Gain* di atas dapat diketahui bahwa atribut yang memiliki gain tertinggi adalah atribut pekerjaan yaitu dengan nilai 0,109199897. Ada tiga nilai dari atribut pekerjaan yaitu

wiraswasta, pegawai swasta dan PNS. Maka atribut pekerjaan menjadi akar untuk membuat pohon keputusan. Dari hasil perhitungan tersebut dapat digambarkan pohon keputusan sementara yang tampak dapat dilihat pada gambar 1.



Gambar 1. Pohon Keputusan

Dari pohon keputusan akan terbentuk aturan – aturan dari penurunan data. Dari sampel calon nasabah yang akan melakukan perpanjangan kredit maka terbentuklah aturan atau rule.

Daftar aturan yang terbentuk adalah:

- a. Jika pekerjaan wiraswasta, pinjaman besar maka lancar.
- b. Jika pekerjaan wiraswasta, pinjaman sedang, jenis agunan BPKB maka lancar.
- c. Jika pekerjaan wiraswasta, pinjaman sedang, jenis agunan sertifikat tanah, umur 20-35 maka tidak lancar.
- d. Jika pekerjaan wiraswasta, pinjaman sedang, jenis agunan sertifikat tanah, umur 36-50 maka lancar.
- e. Jika pekerjaan wiraswasta, pinjaman sedang, jenis agunan sertifikat tanah, umur 51-65 maka lancar.
- f. Jika pekerjaan wiraswasta, pinjaman kecil, jenis agunan sertifikat tanah maka kurang lancar.
- g. Jika pekerjaan wiraswasta, pinjaman kecil, jenis agunan BPKB, umur 20-35 maka lancar.
- h. Jika pekerjaan wiraswasta, pinjaman kecil, jenis agunan BPKB, umur 36-50 maka kurang lancar.
- i. Jika pekerjaan wiraswasta, pinjaman kecil, jenis agunan BPKB, umur 51-65 maka lancar.
- j. Jika pekerjaan pegawai swasta, umur 20-35 maka lancar.
- k. Jika pekerjaan pegawai swasta, umur 36-50, jenis agunan BPKB maka lancar.
- l. Jika pekerjaan pegawai swasta, umur 36-50, jenis agunan sertifikat tanah, pinjaman kecil maka kurang lancar.
- m. Jika pekerjaan pegawai swasta, umur 36-50, jenis agunan sertifikat tanah, pinjaman sedang maka lancar.
- n. Jika pekerjaan pegawai swasta, umur 36-50, jenis agunan sertifikat tanah, pinjaman besar maka lancar

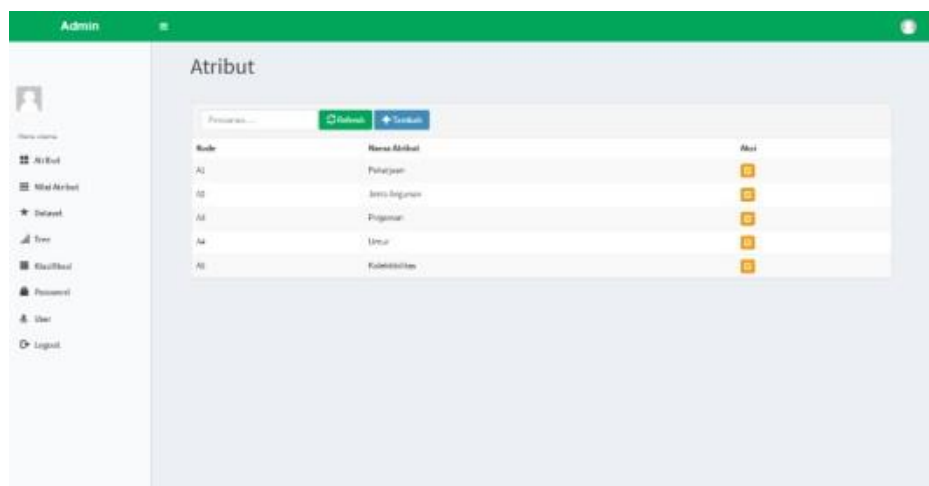
- o. Jika pekerjaan pegawai swasta, umur 51-65, jenis agunan BPKB maka lancar.
- p. Jika pekerjaan pegawai swasta, umur 51-65, jenis agunan sertifikat tanah maka kurang lancar.
- q. Jika pekerjaan PNS maka kurang lancar.

PEMBAHASAN

Implementasi Sistem pada Klasifikasi Nasabah dengan Metode *Decision Tree* sebagai dasar pemberian kredit pada LPD akan dibatasi sesuai dengan ketentuan yang telah dibahas pada bab-bab sebelumnya.

Halaman Atribut

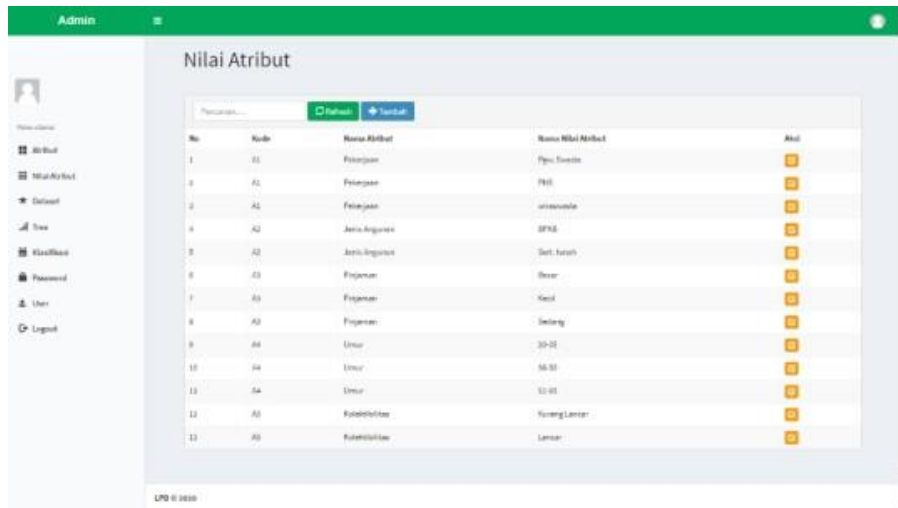
Pada gambar 2 merupakan halaman atribut yang digunakan oleh pengguna untuk memasukan data atribut yang digunakan. Halaman ini juga berisi aksi untuk mengubah data, tombol untuk merefresh data dan *button* tambah untuk menambahkan data baru.



Gambar 2. Halaman Atribut

Halaman Nilai Atribut

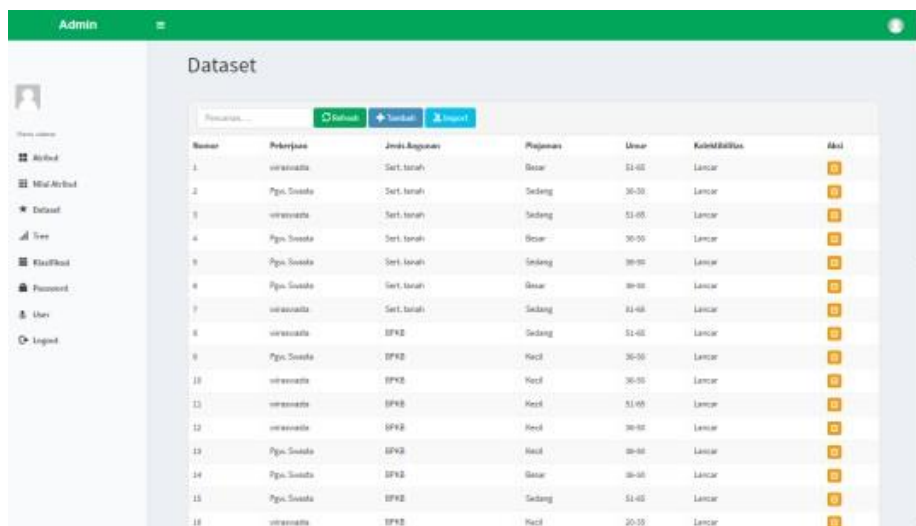
Pada gambar 3 merupakan halaman nilai atribut. Pada halaman nilai atribut ini berisi data nilai atribut yang sudah dimasukan sebelumnya yaitu kode, nama atribut dan nilai atribut. Halaman ini juga berisi aksi untuk mengubah data, *button* untuk merefresh data dan *button* tambah untuk menambahkan data baru.



Gambar 3 Halaman Nilai Atribut

Halaman Dataset

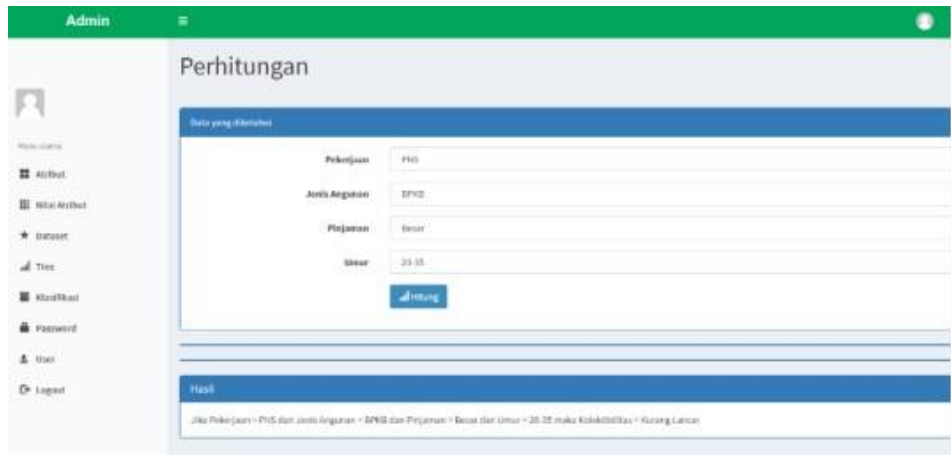
Pada gambar 4 merupakan halaman dataset untuk mengimport dan menambah data nasabah. Halaman dataset berisi kolom tempat dataset yaitu pekerjaan, jenis agunan, pinjaman, umur, kolektibilitas.



Gambar 4. Halaman Dataset

Halaman Klasifikasi

Pada gambar 5 merupakan halaman klasifikasi untuk menghitung dan mengetahui apakah nasabah yang bersangkutan kolektibilitasnya lancar atau tidak. Halaman Klasifikasi berisi pilihan yang harus dipilih seperti pekerjaan, jenis agunan, pinjaman dan umur. Setelah memilih pengguna dapat menekan *button* hitung untuk mengetahui hasil klasifikasinya. Maka hasil perhitungannya akan muncul dibawahnya.



Gambar 5. Halaman Klasifikasi

AKURASI KLASIFIKASI DECISION TREE

Pada penelitian ini untuk mengetahui akurasi datanya menggunakan metode *Confussion matrix*. Metode *Confussion matrix* melakukan pengujian untuk memperkirakan obyek yang benar dan salah [9]. Urutan pengujian ditabulasikan dalam *confussion matrix* dimana kelas yang diprediksi ditampilkan di bagian atas matriks dan kelas yang diamati di bagian kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi.

Jumlah data yang digunakan dalam penelitian ini sebanyak 38 *record*. Untuk data yang digunakan dalam akurasi ini sebanyak 20 *record* Hasil dari akurasi *confussion matrix* klasifikasi nasabah dengan *decision tree* algoritma C4.5 dapat dilihat pada tabel 6.

Tabel 6: Tabel Confusion Matrix

Confusion Matrix				
Klasifikasi	TP	FP	TN	FN
Lancar	14	2	4	0
Kurang Lancar	4	0	14	2

Keterangan:

- True Positive (TP): kasus dimana nasabah diprediksi (Positif) Lancar, memang benar(True) Lancar. Dalam penelitian ini nilai TP(Lancar) adalah 14 dan
- TP(Kuranglancar) adalah 4.
- True Negative (TN): kasus dimana nasabah diprediksi (Negatif) kurang lancar dan sebenarnya nasabah tersebut memang (True) kurang lancar. Dalam penelitian ini nilai TN(Lancar) adalah 4 dan TN(Kuranglancar) adalah 14.
- False Positve (FP): kasus dimana nasabah yang diprediksi positif lancar, ternyata kurang lancar. Prediksinya salah (False).
- Dalam penelitian ini nilai FP(Lancar) adalah 2 dan FP(Kuranglancar) adalah 0.

- f. False Negatif (FN): kasus dimana nasabah yang diprediksi kurang lancar (Negatif), tetapi ternyata sebenarnya (TRUE) lancar. Dalam penelitian ini nilai FN(Lancar) adalah 0 dan FN(Kuranglancar) adalah 2.

Accuracy merupakan proporsi jumlah prediksi benar.

$$Accuracy = (TP + TN) / (TP+TN+FP+FN) \quad (3)$$

Precision adalah proporsi prediksi kasus positif yang benar.

$$Precision = (TP) / (TP+FP) \quad (4)$$

Recall atau tingkat positif benar (TP) adalah proporsi kasus positif yang diidentifikasi dengan benar.

$$Recall = (TP) / (TP + FN) \quad (5)$$

Specificity

$$Specificity = TN/(TN+FP) \quad (6)$$

Nilai AUC (*Area Under Kurve*)

$$\begin{aligned} AUC &= (Recall + Specificity) / 2 \quad (7) \\ &= (0.834+0.834)/2 \\ &= 0.834 \end{aligned}$$

Untuk klasifikasi data mining menurut Gorunescu (2011) [11], bahwa nilai AUC (*Area Under Curve*) dapat dibagi menjadi beberapa kelompok yaitu:

- a. 0,90 – 1.00 = Excellent Classification
- b. 0,80 – 0,90 = Good Classification
- c. 0,70 – 0,80 = Fair Classification
- d. 0,60 – 0,70 = Poor Classification
- e. 0,50 – 0,60 = Failure

Pada penelitian ini nilai AUC yaitu 0.834 dimana termasuk pada interval nilai 0,80-0,90. Jadi metode yang digunakan dalam penelitian ini termasuk kedalam *Good Clasification*.

KESIMPULAN

Berdasarkan hasil dari klasifikasi nasabah kredit pada LPD Desa Adat Anggungan yang telah dilakukan dapat disimpulkan:

1. Analisis klasifikasi nasabah kredit ini dimulai dari pengumpulan data dengan metode wawancara, observasi, dokumentasi dan kepustakaan untuk mengumpulkan data primer dan sekunder. Setelah itu menganalisis masalah dan kendala yang sedang terjadi. Kemudian membuat suatu rancangan usulan sistem dengan menyusun rancangan tersebut menggunakan UML, struktur tabel dan *user interface*.
2. Untuk tahap implementasi, sistem ini dibangun dengan menggunakan bahasa pemrograman php dan mysql sebagai *database*. Pada tahap pengujian dilakukan dengan menggunakan *black box testing*, dimana sistem diuji prosesnya untuk menunjukkan bahwa sistem dapat berjalan sesuai dengan rancangan yang telah dibuat.

3. Secara fungsional pembangunan sistem sudah sesuai diantaranya, fungsi login, fungsi menambah data seperti data atribut dan nilai atribut, fungsi ubah seperti ubah data atribut dan nilai atribut, fungsi klasifikasi, fungsi *import* dataset nasabah dan fungsi pohon keputusan(*tree*).
4. Penelitian ini menggunakan *confusion matrix* untuk mengetahui akurasi klasifikasi dengan *decision tree* algoritma c4.5. Hasil *Accuracy* adalah 0.9, *Precision* adalah 0.938, *Recall* adalah 0.834 dan nilai AUC dari penelitian ini yaitu 0.834 termasuk kedalam *Good Clasification*.

DAFTAR PUSTAKA

- [1] H. Marcos, I. Hidayah, J., 2014 Teknik, E. Dan, T. Informatika, and U. G. Mada, "Implementasi Data Mining Untuk Klasifikasi Nasabah Kredit Bank " X " Menggunakan Classification Rule," *Semin. Nas. Teknol. Inf. dan Multimed.*, no. Februari, pp. 1–7.
- [2] E. P. Cynthia and E. Ismanto, 2018, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 3.
- [3] I. Menarianti, 2015, "Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, pp. 1–10, [Online]. Available: <http://ejournal.upgrismg.ac.id/index.php/JITEK/article/view/836>.
- [4] M. A. Sembiring, 2016, "Penerapan Metode Decision Tree Algoritma C45 Untuk Memprediksi Hasil Belajar Mahasiswa Berdasarkan Riwayat Akademik," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 60–65.
- [5] Wiryaningtyas, 2016, "Pengaruh Keputusan Nasabah Dalam Pengambilan Kredit Pada Bank Kredit Desa Kabupaten Jember," pp. 49–57.
- [6] Maharunti, Mukid, and Rusgiyono, 2016, "Analisis diskriminan fisher populasi ganda untuk klasifikasi nasabah kredit," vol. 5, no. 1936, pp. 575–581.
- [7] A. Saleh, 2015, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217.
- [8] Y. Mardi, 2016, "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".
- [9] M. Yusa *et al.*, 2015, "Evaluasi Model Decision Tree C4 . 5 Guna PrediksI," pp. 147–152.
- [10] Suendri, 2018, "Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan)," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 1, pp. 1–9, [Online]. Available: <http://jurnal.uinsu.ac.id/index.php/algoritma/article/download/3148/1871>.
- [11] A. Hendini, 2016, "Pemodelan Uml Sistem Informasi Monitoring Penjualan Dan Stok Barang," *J. Khatulistiwa Inform.*, vol. 2, no. 9, pp. 107–116.

