

## Implementasi Algoritma Decision Tree Untuk Mengetahui Faktor Kredit Macet Dan Lancar di Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur

Mahpuz<sup>1\*</sup>, Yahya<sup>2</sup>, Muhammad Wasil<sup>3</sup>

<sup>1,3</sup>Program Studi Teknik Informatika, Universitas Hamzanwadi

<sup>2</sup>Program Studi Sistem Informasi, Universitas Hamzanwadi

\*mahfuzuma@gmail.com

### Abstrak

Kredit merupakan produk utama dari koperasi simpan pinjam dalam upaya meningkatkan profitabilitasnya. Semakin besar kredit yang dikeluarkan maka semakin besar pula laba diperoleh oleh suatu koperasi tersebut, setiap koperasi akan mengemas produk kredit sedemikian rupa agar dapat menarik perhatian setiap nasabahnya. Akan tetapi suatu koperasi dapat menemukan masalah dalam menyalurkan kreditnya, seperti yang ada pada koperasi Serba Usaha Daruzzakah yang terletak di Desa Rensing Kecamatan Sakra Barat Kabupaten Lombok Timur, dalam menentukan kelayakan pemberian kredit mereka selalu menggunakan data statistik, akan tetapi data tersebut tidak menjamin apakah selama pembayaran kredit lancar atau macet. Algoritma Decision Tree akan memberikan beberapa opsi untuk melihat faktor apa yang menjadi kredit macet atau lancar, hal ini dilakukan untuk mengklasifikasikan kredit mana yang sekiranya lancar dan macet sehingga kedepannya bisa dilakukan antisipasi lebih awal

Kata Kunci : Kredit, Koperasi, *Decision Tree*

### Abstract

Credit is the main product of savings and loan cooperatives in an effort to increase profitability. The greater the credit issued, the greater the profit earned by a cooperative, each cooperative will package credit products in such a way as to be able to attract the attention of each customer. However, cooperative suati can find problems in channeling credit, such as those in the Daruzzakah Multipurpose Cooperative located in Rensing Village, West Sakra Subdistrict, East Lombok Regency, however, the cooperative some can find problems in channeling credit, such as those in the Daruzzakah Multipurpose Cooperative which is located in Rensing Village, West Sakra District, East Lombok Regency, in determining the creditworthiness of credit they always use statistical data, but the data does not guarantee whether during credit or non-performing loans. The Decision Tree algorithm will provide several options to see what factors are bad or current loans, this is done to classify which loans are smooth and bad so that in the future it can be anticipated early.

Keywords: Credit, Cooperatives, Decision Tree

### 1. Pendahuluan

Perkembangan ilmu pengetahuan dan teknologi menyebabkan terjadinya perubahan besar dalam semua aspek terutama aspek perekonomian. Meningkatnya kebutuhan hidup membuat pengeluaran yang harus dikeluarkan semakin banyak. Persaingan pasar bebas yang semakin tinggi menimbulkan sinergi positif dimana

keterlibatan berbagai macam pihak dalam upaya perbaikan perekonomian.

Koperasi sebagai gerakan ekonomi yang tumbuh dari masyarakat merupakan organisasi swadaya masyarakat yang lahir atas kehendak, kekuatan dan partisipasi masyarakat dalam menentukan tujuan, sasaran kegiatan serta pelaksanaannya. Keberadaan koperasi sebagai wadah untuk

mewujudkan kesejahteraan bersama bagi seluruh Rakyat Indonesia.

Koperasi Daruzzakah Rensing merupakan koperasi simpan pinjam yang kegiatannya usahanya menghimpun dan menyalurkan dana kepada para anggotanya dengan bunga yang rendah. Koperasi ini disebut juga sebagai koperasi kredit dimana pengelolaannya dilakukan secara mandiri dan demokratis, serta anggotanya bergabung secara sukarela, ada pula yang menyebut bahwa koperasi simpan pinjam merupakan lembaga keuangan non-bank yang memiliki kegiatan usaha dari anggotanya dan memberikan pinjaman kepada para anggota dengan bunga rendah. sehingga untuk mengatasi masalah ini penulis akan menerapkan data mining dengan algoritma decision tree dalam menganalisa pola yang akan membantu untuk mencari tau kenapa kredit dari nasabah itu dikatakan macet atau lancar, yang dimana algoritma decision tree merupakan algoritma klasifikasi yang memberikan *output* berupa sebuah pohon keputusan yang menampilkan setiap detail dari masalah yang diteliti.

## 2. Landasan Teori

### 2.1. Penelitian Terkait

Beberapa penelitian terkait untuk penggunaan *Data Mining* metode klasifikasi dengan menggunakan algoritma *decision tree* dalam mencari pola untuk menentukan yang terbaik atau untuk memprediksi yang akan terjadi kedepannya, diantaranya adalah :

- Penelitian yang dilakukan oleh Yahya, dan Reni Zuliana pada jurnal yang berjudul “Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decision Tree (C4.5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur”(2018) didapat kesimpulan jika Dari hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa penggunaan BBM dapat diprediksi dan dievaluasi dengan memanfaatkan teknik *Data Mining* menggunakan algoritma *decision tree* (C4.5), dari data penggunaan BBM dengan data training yang telah diperoleh. Metode algoritma *Decision Tree* (C4.5) yang digunakan pada prediksi ini cukup baik karena menghasilkan nilai akurasi 67.33%. Dari tiga kali percobaan yang telah dilakukan dapat diketahui hasil nilai *K-Fold Validation* yaitu, semakin besar nilai *K-Fold Validation* yang digunakan maka semakin tinggi nilai akurasi yang akan diperoleh. Metode algoritma *Decision Tree* (C4.5) yang digunakan dapat diketahui pula tingkat penggunaan yang tinggi dan yang baik, sehingga dalam menganalisa data metode ini cukup akurat untuk tetap mengoptimalkan pemakaian BBM [1].
- Penelitian yang dilakukan oleh Yupi Kuspani Putra dan Yeni Kurniatul Mashun pada jurnal yang berjudul “Prediksi Jumlah Retribusi Sampah Perbulan menggunakan algoritma *decision tree* (C4.5) pada Kantor Dinas Lingkungan Hidup dan Kebersihan

Kecamatan Selong Kabupaten Lombok Timur” didapatkan kesimpulan Dari pengujian yang telah dilakukan sebelumnya dengan algoritma Decition Tree dengan *number of validation* yang berbeda, hasil yang diperoleh menunjukkan bahwa algoritma *Decition Tree* (C4.5) memperoleh hasil yang unggul. Dari tiga kali uji coba, ketiganya menunjukkan hasil akurasi dan nilai AUC yang tinggi. Hasil yang paling unggul yaitu dengan menggunakan *number of validation* 3, dengan nilai akurasi yang diperoleh sebesar 96.51% dan nilai AUC sebesar 0.976% dengan tingkat diagnosa nilai akurasi excellent classification. Dengan demikian dapat disimpulkan bahwa algoritma *Decition Tree* (C4.5) terbukti akurat dalam prediksi retribusi persampahan<sup>[2]</sup>.

- Penelitian yang dilakukan oleh afnia Sartika Hutasoid, Pilipus Tarigan, Dan Robinson Siagian dalam jurnal yang berjudul “Implementasi Data Mining Klasifikasi Gizi Balita Pada Posyandu Medan Timur Dengan Menggunakan Metode C4.5” didapatkan kesimpulan bahwa Algoritma C45 (Decision tree) dapat digunakan untuk memudahkan dalam pengambilan keputusan dengan memproyeksikan data-data yang ada ke dalam bentuk pohon keputusan, berdasarkan nilai entropy dan gain yang dimiliki masing-masing atribut data dan algoritma C4.5 dapat dipakai untuk menghasilkan keputusan yang akurat untuk memberikan hasil lebih baik dalam mengklasifikasikan setatus gizi karena

data status gizi balita berupa data kelompok yang cocok dengan sifat klasifikasi algoritma C4.5<sup>[3]</sup>.

## 2.2. Landasan Teori

### 2.2.1. Pengertian koperasi

Undang-undang Nomor 25 tahun 1992 menyatakan bahwa koperasi merupakan badan usaha yang beranggotakan orang atau badan hukum koperasi yang melandaskan kegiatannya berdasarkan atau asas kekeluargaan. Sementara pernyataan Standar Akutansi Nomor 27 menyatakan bahwa koperasi adalah badan usaha yang mengorganisasi pemanfaatan dan pendayagunaan sumber daya ekonomi para anggotanya atas dasar prinsip-prinsip koperasi dan kaidah usaha ekonomi untuk meningkatkan taraf hidup anggota dan masyarakat pada umumnya<sup>[1]</sup>.

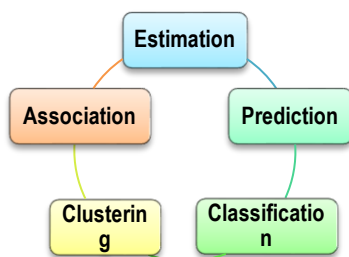
### 2.2.2. Datamining

Data *mining* adalah proses untuk menggali (*mining*) pengetahuan dan informasi baru dari data yang berjumlah banyak, dengan menggunakan kecerdasan buatan (*artificial Intelegence*), statistik dan matematika. Data mining merupakan teknologi yang diharapkan dapat menjembatani komunikasi antara data dan pemakainya.

*Data mining* merupakan disiplin ilmu yang bertujuan untuk menemukan, menggali pengetahuan dari data atau informasi yang kita miliki. Data mining pencarian pola yang diinginkan pada database untuk membantu

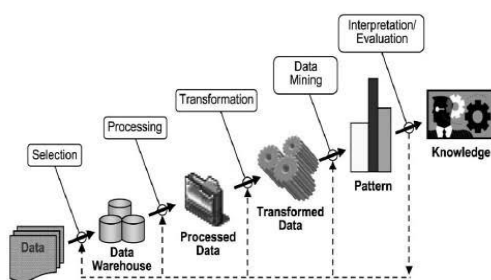
pengambilan keputusan di waktu yang akan datang. Text mining mengekstrak pola yang berguna dari sumber data teks melalui eksplorasi pola tertentu. Pola-pola ini dapat diketahui oleh perangkat tertentu yang memberikan suatu analisa data yang berguna kemudian dapat dipelajari lebih teliti, yang memungkinkan penggunaan perangkat pendukung lainnya<sup>[2]</sup>.

Dalam prosesnya datamining dibagi menjadi lima proses tergantung jenis data yang akan diolah, yaitu



Gambar 1. Proses dalam data mining

Tahapan yang dilakukan pada proses datamining diawali dari seleksi data dari data sumber ke data target, tahap processing untuk memperbaiki kualitas data, transformasi data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik secara detail dijelaskan sebagai berikut :



Gambar 2. Tahapan pengolahan data mining

### 1. Pembersihan data (*Data Cleaning*)

Merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau tidak relevan. Pada umumnya, data yang diperoleh baik dari database suatu perusahaan maupun hasil eksperimen memiliki isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang, pembersihan data juga akan mempengaruhi performansi dari teknik data mining, karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

### 2. Integrasi data (*Data Integration*)

Merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. 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.

### 3. Seleksi data (*Data Selection*)

Data yang ada pada database seringkali tidak semuanya dipakai. Oleh karena itu, hanya

data yang sesuai untuk dianalisis yang akan diambil dari database.

#### 4. 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 di aplikasikan.

#### 5. Proses mining

Proses mining merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

#### 6. Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi di evaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

#### 2.2.3. Dataset

Dataset adalah objek yang mempresentasikan data dan relasinya di memori, strukturnya mirip dengan data yang ada di database<sup>[3]</sup>. *Dataset*

berisi koleksi dari data tabel dan data, jenis dataset ada dua macam yaitu :

##### 1. Private dataset

*Private dataset* adalah data yang di dapatkan atau diambil dari Organisasi yang dijadikan tempat atau sebagai objek penelitian.

##### 2. Public Dataset

*Public dataset* adalah dataset yang dapat diambil dari *repository public* (penyimpanan umum) yang telah disepakati oleh para peneliti untuk dipublikasikan kepada umum

##### 3. Algoritma Decision Tree

*Decision tree* merupakan suatu metode klasifikasi yang menggunakan struktur pohon, dimana setiap node merepresentasikan atribut dan cabangnya merepresentasikan nilai dari atribut, sedangkan daunnya digunakan untuk merepresentasikan kelas. *Node* teratas dari *decision tree* ini disebut dengan *root*.

Dalam beberapa aplikasi, akurasi dari sebuah klasifikasi atau prediksi adalah satu-satunya hal yang ditonjolkan dalam metode ini, contohnya perusahaan surat langsung membuat model yang akurat untuk memprediksi anggota mana yang berpotensi merespon permintaan terlepas dari bagaimana atau mengapa model itu berfungsi<sup>[4]</sup>.

Berikut beberapa jenis dari node yang ada pada algoritma *decision tree*

##### 1. Akar

Akar merupakan *node* yang berada paling awal, tidak ada input dan outputnya bisa saja

tidak ada atau bahkan bisa lebih dari satu output

2. Internal node

Internal *node* merupakan *node* percabangan, disini bisa terjadi dua percabangan.

3. Daun

Ini merupakan kesimpulan dari internal *node*, yang dimana hanya akan terdapat satu *output* saja atau satu kesimpulan.

Memilih atribut berdasarkan nilai “*gain*” tertinggi dari atribut-atribut yang ada, rumusnya sebagai berikut [5]

1. Perhitungan Gain

*Gain* adalah yaitu ukuran efektifitas suatu atribut dalam mengklasifikasi data.

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

Keterangan:

S : himpunan

A : atribut

n : jumlah partisi atribut A

Si : jumlah kasus pada partisi ke-i

S : jumlah kasus dalam S

2. Menghitung Nilai Entropy

*Entropy* adalah besaran sebuah nilai ketidakpastian dalam suatu atribut dari kumpulan objek

$$Entropy(S) = \sum_{i=0}^n - pi * log_2 pi$$

dimana :

S : himpunan kasus

A : fitur

n : jumlah partisi S

pi : proporsi dari Si terhadap S

3. Metode Penelitian

3.1. Metode yang diusulkan

Dalam melakukan eksperimen, penulis melakukan beberapa tahap agar mendapatkan hasil yang maksimal, diantaranya :

- Tahap Pemahaman Bsisnis

Sesuai dengan laporan data kredit Koperasi Serba Usaha Daruzzakah Rensing Tahun 2018, terdapat beberapa masalah kredit yang sangat berpengaruh terhadap kestabilan koperasi. Untuk itu, maka dikembangkan model algoritma C 4.5 (*Decision tree*) untuk menentukan kelayakan faktor kredit lancar dan macet nasabah, agar untuk kedepannya bisa diantisipasi.

- Tahap Pemahaman Data

Data kredit yang penulis gunakan ini diambil dari data Koperasi Serba Usaha Daruzzakah Rensing Tahun 2018. Di dalam data ini terdapat 254 nasabah yang di antaranya 88 orang dianggap lancar, 3 orang kurang lancar, dan 163 orang dinyatakan macet. Dalam data tersebut ada beberapa data nasabah yang digunakan sebagai atribut, yaitu no rekening nasabah, realisasi, jangka waktu, jumlah pinjaman, sisa pokok, sisa basil, tunggakan, dan kolektibilitasnya.

- Tahap Persiapan (*preparation*) data

Pada tahapan ini, data sebanyak 254 dan terdiri dari 8 atribut akan dilakukan beberapa penyeleksian untuk menghasilkan data yang dibutuhkan, tahapannya yaitu :

1. *Data Cleaning* (pembersihan data) untuk membersihkan nilai kosong atau tupel yang kosong.
2. *Data integration* berfungsi menyatukan tempat penyimpanan (arsip) yang berbeda kedalam satu data. Dalam hal ini selama setahun data dipisahkan perbulan, maka dengan tahap ini data perbulan akan disatukan dalam satu tahun.
3. *Data reduction* (pengurangan), jumlah atribut yang digunakan hanya beberapa saja, dan atribut yang tidak diperlukan akan dihapus. Objek dalam dataset mungkin terjadi duplikasi atau terjadi atribut yang sama, sehingga untuk memprediksi jumlah atribut, atribut yang sama akan dijadikan dalam satu atribut untuk mewakili atribut tersebut. Berikut adalah dataset yang digunakan :

Tabel 1. Dataset nasabah Daruzzakah Rensing

No. Rekening	Realisasi	JKW	Jumlah Pinjaman	Sisa Pokok	Sisa Basil	Tunggakan	Kolektibilitas
12.010.000.002	2-Nov-09	10	500000	100000	30	100000	Macet
12.010.000.003	6-Nov-09	102	500000	215000	43	214999	Macet
12.010.000.004	12 Desember 2009	10	500000	100000	10	100000	Macet
12.010.000.007	26-Apr-10	119	1200000	588000	98	587999	Macet
12.010.000.008	21 Mei 2010	14	1000000	72000	8	719999	Macet
12.010.000.010	15 Juni 2010	72	300000	193000	46	193000	Macet
12.010.000.011	06 Agustus 2010	11	750000	150	20	150000	Macet
12.010.000.012	27-Sep-10	98	700000	55	20	55000	Macet
12.010.000.013	28 Oktober 2010	10	1000000	300	30	300	Macet
12.010.000.014	13-Nov-10	15	1000000	31	24	31000	Macet
12.010.000.015	18-Nov-10	5	2200000	720	298	720	Macet
12.010.000.016	23-Nov-10	120	2000000	340	66	340000	Macet
12.010.000.017	26-Nov-10	15	1000000	456	100000	456000	Macet
12.010.000.019	23 Februari 2011	19	1000000	400	120	400000	Macet

#### 4. Hasil dan Pembahasan

Tujuan dari penelitian ini adalah untuk menguji keakuratan analisa prediksi faktor kredit macet

dan lancar pada nasabah koperasi Daruzzakah Rensing menggunakan metode Algoritma *decision tree*. Data yang dianalisa adalah data kredit nasabah koperasi, yaitu semua data tersebut telah disetujui oleh pihak koperasi Serba Usaha Daruzzakah Rensing. Dari uji coba menggunakan *K-Fold Validation 2* sampai 10.

Tabel 2. Hasil pengujian berdasarkan x validation

Item	Uji Coba (Number of validation)									
	2	3	4	5	6	7	8	9	10	
Accuracy	95,63%	96,03%	96,03%	95,64%	96,43%	93,65%	95,68%	96,43%	96,45%	
Sensitiv	98,80%	98,80%	98,20%	98,20%	98,80%	97,01%	97,60%	98,20%	98,80%	
Specificity	89,41%	90,59%	91,76%	90,59%	91,76%	87,06%	91,76%	92,94%	91,76%	
Ppv	94,83%	95,38%	95,91%	95,35%	95,93%	93,64%	95,88%	94,47%	95,93%	
Npv	97,44%	97,47%	96,30%	96,25%	97,50%	93,67%	95,12%	96,34%	97,50%	

Dalam pengujian yang dilakukan maka didapatkan nilai akurasi tertinggi pada *K-Fold Validation 9* dan 10 dengan *sensitivity*, *specificity*, *Ppv* dan *Npv* sebagai berikut:

Tabel 3. Nilai akurasi tertinggi

Item	Uji Coba (Number of validation)	
	9	10
Accuracy	96,43%	96,45%
Sensitiv	98,20%	98,80%
specificity	92,94%	91,76%
Ppv	94,47%	95,93%
Npv	96,34%	97,50%

##### 1. K-Fold Validation 9

Pengujian ini dilakukan dengan membagi 9 bagian pada 252 data perkreditan nasabah yang akan diuji, yang terdiri dari data *training* dan data *testing*. Model *confusion matrix* akan membentuk *matrix* yang terdiri dari *true* positif atau tupel positif dan *true* negatif atau tupel negatif, kemudian masukkan data *testing* yang telah disiapkan kedalam *confusion matrix* sehingga mendapatkan hasil sebagai berikut

	true Macet	true Lancar	class precis
pred. Macet	164	6	96.47%
pred. Lancar	3	79	96.34%
class recall	98.20%	92.94%	

Gambar 3. Hasil akurasi *K-Fold Validation 9*

Gambar di atas merupakan hasil pengujian menggunakan *K-Fold Validation 9* yang menghasilkan nilai *accuracy* 96,43%.

Tabel 4. Hasil akurasi *K-Fold Validation 9*

Predicate Class		
Observed Class	Macet	Lancar
Macet	164	6
Lancar	3	79

Jumlah *true positif* (TP) adalah 164 *record* diprediksikan MACET menerima kredit dan *false negatif* (FN) sebanyak 3 *record* diprediksikan MACET, akan tetapi LANCAR untuk diberikan kredit. Berikutnya 79 *record* untuk *true negatif* (TN) diprediksikan sebagai LANCAR, dan 6 *record false positif* (FP) diprediksikan LANCAR namun ternyata MACET. Nilai *accuracy*, *sensitivity*, *specitivity*, *ppv*, dan *nvp* dapat dihitung dengan menggunakan persamaan di bawah ini :

$$\begin{aligned} \text{Accuracy} &= \frac{tp+tn}{tp+tn+fp+fn} \\ &= \frac{164+79}{164+79+6+3} \times 100 \% = 96,43 \\ &\% \end{aligned}$$

Dengan pengujian *K-Fold Validation 9* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil akurasi 96.43%, hal ini berarti dari 252 data yang di uji, 243 data yang dapat diprediksi. Sehingga tingkat kesalahan yang diperoleh

adalah 4%, dalam hal ini 9 data tidak dapat diprediksi.

$$\begin{aligned} \text{Sensitivity} &= \frac{tp}{tp+fn} \times 100\% \\ &= \frac{164}{164+3} \times 100\% = 98,20\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *true positif* (TP) prediksi MACET sebesar 98,20%, hal ini berarti 247 data yang diprediksi Kreditnya MACET. Sehingga tingkat kesalahan yang diperoleh adalah 2%, dalam hal ini 5 data diprediksi LANCAR.

$$\begin{aligned} \text{Specificity} &= \frac{tn}{tn+fp} \\ &= \frac{79}{79+6} \times 100\% = 92,94\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *true negatif* (TN) prediksi LANCAR sebesar 92,94%, hal ini berarti 234 data yang diprediksi Kreditnya LANCAR. Sehingga tingkat kesalahan yang diperoleh adalah 8%, dalam hal ini 18 data diprediksi MACET

$$\begin{aligned} \text{Ppv} &= \frac{tp}{tp+fp} \\ &= \frac{164}{164+6} \times 100\% = 94,47\% \end{aligned}$$

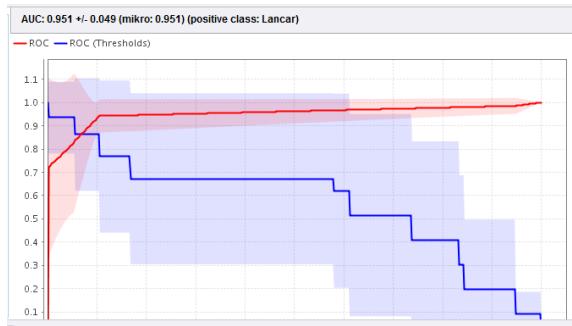
Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *false positif* (FP) prediksi LANCAR namun pada kebenarannya adalah MACET memperoleh hasil 94.47%, hal ini berarti 238 data yang



diprediksi LANCAR namun pada kebenarannya adalah MACET. Sehingga tingkat kesalahan yang diperoleh adalah 6%, dalam hal ini 14 data diprediksi MACET namun pada kebenarannya adalah LANCAR.

$$\begin{aligned} Npv &= \frac{tn}{tn+fn} \\ &= \frac{79}{79+3} \times 100\% = 96,34\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *false negatif* (FN) prediksi MACET namun pada kebenarannya adalah LANCAR memperoleh hasil 96.34%, hal ini berarti 243 data yang diprediksi MACET namun pada kebenarannya adalah LANCAR. Sehingga tingkat kesalahan yang diperoleh adalah 4%, dalam hal ini 9 data diprediksi LANCAR namun pada kebenarannya adalah MACET.



Gambar 4. Hasil AUC dari *K-Fold Validation* 9

Gambar di atas adalah grafik ROC dengan nilai AUC (*Area Under Curve*) dengan hasil *accuracy* dengan *K-Fold Validation* 9 sebesar 0.951 dengan nilai akurasi *Excellent Classification*.

## 2. *K-Fold Validation* 10

Pengujian ini dilakukan dengan membagi 254 data menjadi 10 bagian, yang terdiri dari data

*training* dan data *testing*. Model *confusion matrix* akan membentuk *matrix* yang terdiri dari *true positif* atau *tupel positif* dan *true negatif* atau *tupel negatif*, kemudian masukkan data *testing* yang telah disiapkan kedalam *confusion matrix* sehingga mendapatkan hasil. Nilai *accuracy*, *sensitivity*, *specitivity*, *ppv*, dan *nvp* dapat dihitung dengan menggunakan persamaan di bawah ini :

accuracy: 96.45% +/- 3.67% (mikro: 96.43%)			
	true Macet	true Lancar	class pred
pred. Macet	165	7	95.93%
pred. Lancar	2	78	97.50%
class recall	98.80%	91.76%	

Gambar 5. Hasil akurasi *K-Fold Validation* 10

Gambar di atas merupakan hasil pengujian menggunakan *K-Fold Validation* 10 yang menghasilkan nilai *accuracy* 96,45%.

Tabel 5. Hasil akurasi *K-Fold Validation* 10

Predicate Class		
Observed Class	Macet	Lancar
Macet	165	7
Lancar	2	78

Jumlah *true positif* (TP) adalah 165 *record* diprediksikan kredit MACET dan *false negatif* (FN) sebanyak 2 *record* diprediksikan MACET, akan tetapi LANCAR. Berikutnya 78 *record* untuk *true negatif* (TN) diprediksikan sebagai LANCAR membayar kredit, dan 7 *record false positif* (FP) diprediksikan LANCAR namun ternyata MACET.

$$\begin{aligned} Accuracy &= \frac{tp+tn}{tp+tn+fp+fn} \\ &= \frac{165+78}{165+78+7+2} \times 100\% = 96,45\% \end{aligned}$$

Dengan pengujian *K-Fold Validation* 10 yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh

hasil akurasi 96.45%, hal ini berarti dari 252 data yang di uji, 243 data yang dapat diprediksi. Sehingga tingkat kesalahan yang diperoleh adalah 4%, dalam hal ini 24 data tidak dapat diprediksi.

$$\begin{aligned} \text{Sensitivity} &= \frac{tp}{tp+fn} \\ &= \frac{165}{165+2} \times 100\% = 98,80\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *true positif* (TP) prediksi MACET sebesar 98,80%, hal ini berarti 249 data yang diprediksi Kreditnya MACET. Sehingga tingkat kesalahan yang diperoleh adalah 2%, dalam hal ini 3 data diprediksi LANCAR.

$$\begin{aligned} \text{Specificity} &= \frac{tn}{tn+fp} \\ &= \frac{165}{165+7} \times 100\% = 91,76\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *true negatif* (TN) prediksi LANCAR sebesar 91,76%, hal ini berarti 231 data yang diprediksi kreditnya LANCAR. Sehingga tingkat kesalahan yang diperoleh adalah 9%, dalam hal ini 21 data diprediksi MACET.

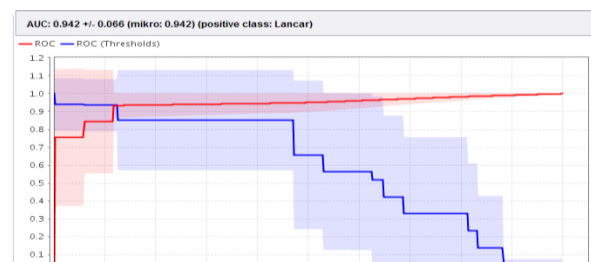
$$\begin{aligned} \text{Ppv} &= \frac{tp}{tp+fp} \\ &= \frac{165}{165+7} \times 100\% = 95,93\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *false positif* (FP) prediksi LANCAR namun

pada kebenarannya adalah MACET memperoleh hasil 95.93%, hal ini berarti 242 data yang diprediksi LANCAR namun pada kebenarannya adalah MACET. Sehingga tingkat kesalahan yang diperoleh adalah 5%, dalam hal ini 10 data diprediksi MACET namun pada kebenarannya adalah LANCAR.

$$\begin{aligned} \text{Npv} &= \frac{tn}{tn+fn} \\ &= \frac{78}{78+2} \times 100\% = 97,50\% \end{aligned}$$

Dengan pengujian algoritma *decision tree* yang terdapat 252 data untuk diprediksi yang terdiri dari data *training* dan data *testing* memperoleh hasil *false negatif* (FN) prediksi MACET namun pada kebenarannya adalah LANCAR memperoleh hasil 97.50%, hal ini berarti 246 data yang diprediksi MACET namun pada kebenarannya adalah LANCAR. Sehingga tingkat kesalahan yang diperoleh adalah 3%, dalam hal ini 6 data diprediksi LANCAR namun pada kebenarannya adalah MACET.



Gambar 6. Hasil AUC dari *K-Fold Validation* 10

Gambar di atas adalah grafik ROC dengan nilai AUC (*Area Under Curve*) dengan hasil *accuracy* dengan *K-Fold Validation* 10 sebesar 0.942% dengan nilai akurasi *Excellent Classification*.

Berdasarkan hasil uji coba maka didapatkan *accuracy* tertinggi pada dua *K-Fold Validation* yaitu pada *K-Fold Validation* 9 dan *K-Fold*

*Validation* 10. Dimana didapatkan hasil *accuracy* sebesar 96,43 % pada *K-Fold Validation* 9 dan *accuaracy* sebesar 96,45% pada *K-Fold Validation* 10. Dan selisih akurasi keduanya adalah 0.02% yang menunjukkan bahwa terjadi peningkatan 0.02% pada rentang *k-fold validation* 9 dan 10. Peningkatan pada setiap uji coba tidaklah selalu sama ini dipengaruhi oleh seberapa banyak pembagian data yang dilakukan

Algoritma *decision tree* mempunyai kemampuan yang baik dalam memecahkan masalah *data mining* walaupun dengan sampel yang terbatas. Eksperimen menggunakan metode *decision tree* yang terbaik menghasilkan nilai akurasi sebesar 96.45%. Hasil tersebut diperoleh dengan metode *K-Fold Validation* 10, dimana data dibagi menjadi 10 bagian untuk *ditraining* dan *ditesting* . serta didapatkan juga nilai AUC (*Area Under Curve*) sebesar 0.942 yang dimana AUC merupakan parameter yang digunakan dalam analisis klasifikasi dalam menentukan model terbaik untuk prediksi suatu kelas atau atribut, AUC sendiri mempunyai rentang nilai dari 0 sampai 1, yang artinya semakin nilai AUC mendekati angka 1 maka prediksi atau diagnosa atributnya semakin bagus. Ini menunjukkan nilai AUC sebesar 0.942 yang didapatkan setelah melakukan pengujian dengan *K-Fold Validation* 10 sangat baik karena hampir mendekati angka 1.

## 5. Kesimpulan

Setelah dilukukannya pengujian sebanyak 9 kali yaitu dari *validation* 2 sampai dengan 10 maka didapatkan dua hasil akurasi terbaik dari *K-Fold Validation* 9 dan *K-Fold Validation* 10 yang masing-masing memiliki nilai *accuracy* sebesar 96,43% pada *K-Fold Validation* 9 dan 96,45% dari *K-Fold Validation* 10. Dan selisih akurasi keduanya adalah 0.02% yang menunjukkan bahwa terjadi peningkatan 0.02% pada rentang *k-fold validation* 9 dan 10. Peningkatan pada setiap uji coba tidaklah selalu sama ini dipengaruhi oleh seberapa banyak pembagian data yang dilakukan

Maka didapatkan nilai *accuracy* terbaik pada *K-Fold Validation* 10 dengan akurasi sebesar 96,45% dimana data dibagi menjadi 10 bagian untuk *ditraining* dan *ditesting* . serta didapatkan juga nilai AUC (*Area Under Curve*) sebesar 0.942 yang dimana AUC merupakan parameter yang digunakan dalam analisis klasifikasi dalam menentukan model terbaik untuk prediksi suatu kelas atau atribut, AUC sendiri mempunyai rentang nilai dari 0 sampai 1, yang artinya semakin nilai AUC mendekati angka 1 maka prediksi atau diagnosa atributnya semakin bagus. Ini menunjukkan nilai AUC sebesar 0.942 yang didapatkan setelah melakukan pengujian dengan *K-Fold Validation* 10 sangat baik karena hampir mendekati angka 1.

## 6. Daftar Pustaka

- [1] R. Z. Yahya, "Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decition Tree (C4.5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur," vol. 1, no. 1, pp. 56–63, 2018.
- [2] Y. K. Putra, Y. K. Mashun, F. Teknik, and U. Hamzanwadi, "Prediksi Jumlah Retribusi Sampah Perbulan menggunakan Algoritma Decition Tree (C4.5) pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur," vol. 1, no. 1, 2018.
- [3] A. C Hutasoit, P. Tarigan, and E. R. Siagian, "IMPLEMENTASI DATA MINING KLASIFIKASI STATUS GIZI BALITA PADA POSYANDU MEDAN TIMUR DENGAN MENGGUNAKAN," vol. 17, pp. 355–361, 2018.
- [4] A Jajang W. Mahri, "Pelayanan dan manfaat koperasi, serta pengaruhnya terhadap partisipasi anggota," 2016.
- [5] E. I. Siburian, G. Ginting, and R. D. Sianturi, "PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELAYAKAN PEMAKAIAN ALAT LABORATORIUM KOMPUTER (STUDI KASUS: SMK HKBP SIDIKALANG)," *Maj. Ilm. INTI*, vol. 13, no. September, pp. 256–262, 2018.
- [6] A. Zulkifli, "Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler," vol. 2, no. 1, p. 41, 2016.
- [7] P. Alkhairi, "Pembelajaran dasarr klasifikasi algoritma decision tree," pp. 1–5, 2018.
- [8] S. K. Andie, "PENERAPAN DECISION TREE UNTUK MENGANALISIS MAHASISWA BARU," vol. 7, no. 1, pp. 8–14, 2016.