

Analisa Perbandingan Algoritma DT C.45 dan Naïve Bayes Dalam Prediksi Penerimaan Kredit Motor

Fadly Ariadi

STMIK ERESHA

Fadly_ariadi@yahoo.com

ABSTRACT

Along with the rapid development and growth of business, the problem of credit is very interesting to be discussed and studied. Several studies have been conducted with the aim of reducing credit risk that can cause losses to the company. In this research comparison algorithm C4.5 and naïve bayes which tested to data creditors who have got credit financing motor, whether that problem in installment payment or not. From the test results by measuring the performance of both algorithms using Cross Validation, Confusion Matrix and ROC Curve, it is known that C4.5 algorithm has accuracy value 91.43%, while naïve bayes method has accuracy value 89.52%. The AUC value for C4.5 method also shows the highest value, that is 0.959 while the AUC naïve bayes value is 0.949. Looking at the AUC values of both methods, the two belong to a very good classification group because their AUC values are between 0.90-1.00.

Article History

Received 2020-09-25
Revised 2020-10-07
Accepted 2020-11-04

Key words

Credit
C4.5
Data Mining
Motorcycle
Naïve Bayes

ABSTRAK

Seiring dengan pesatnya perkembangan dan pertumbuhan bisnis, persoalan perkreditan sangat menarik untuk dibahas dan dipelajari. Beberapa penelitian telah dilakukan dengan tujuan untuk mengurangi resiko kredit yang dapat menyebabkan kerugian pada perusahaan. Dalam penelitian ini dilakukan komparasi algoritma C4.5 dan naïve bayes yang diuji terhadap data kreditur yang telah mendapat pembiayaan kredit motor, baik itu yang bermasalah dalam pembayaran angsurannya maupun tidak. Dari hasil pengujian dengan mengukur kinerja kedua algoritma tersebut menggunakan metode pengujian Cross Validation, Confusion Matrix dan Kurva ROC, diketahui bahwa algoritma C4.5 memiliki nilai accuracy 91.43%, sedangkan metode naïve bayes memiliki nilai accuracy 89.52%. Nilai AUC untuk metode C4.5 juga menunjukkan nilai tertinggi, yaitu 0.959 sedangkan nilai AUC naïve bayes, yaitu 0.949. Melihat nilai AUC dari kedua metode tersebut maka keduanya termasuk kelompok klasifikasi sangat baik karena nilai AUC-nya antara 0.90-1.00.

PENDAHULUAN

Seiring dengan pesatnya perkembangan dan pertumbuhan bisnis, persoalan pengkreditan sangat menarik untuk dibahas dan dipelajari karena pada dasarnya dunia bisnis selalu bergerak maju, dan berkembang dari waktu ke waktu. Sementara itu, kebutuhan akan dana atau pinjaman meningkat dengan pesatnya, namun ketersediaan sumber dana yang terbatas dapat menjadi penghambat dalam memajukan usaha.

Lembaga keuangan adalah suatu badan usaha yang kekayaannya berbentuk aset keuangan (*financial asset*) atau tagihan (*claims*) dibandingkan dengan aset non keuangan (*non financial asset*). Lembaga keuangan terutama memberikan kredit dan menanamkan dananya dalam surat-surat berharga [1]. Di samping itu, lembaga keuangan juga menawarkan secara luas berbagai jenis jasa keuangan antara lain: simpanan, kredit, proteksi asuransi, program pensiun, penyediaan mekanisme pembayaran, dan mekanisme transfer dana. Dalam kegiatannya, lembaga keuangan banyak diterpa masalah dan tidak habis diperbincangkan dan diulas dalam berbagai kesempatan. Dari sisi perkreditan, masalah yang dihadapi dan penyebab munculnya permasalahan, dari waktu ke waktu relatif hampir sama. Penyebab munculnya permasalahan bukan karena sistem dan perangkat peraturan yang disiapkan oleh Bank Indonesia yang kurang memadai, tetapi lebih

banyak dipengaruhi oleh kualitas SDM yang mengelola perkreditan pada lembaga keuangan tersebut.

Menurut Surat Keputusan Bersama Menteri Keuangan, Perindustrian dan Perdagangan No.1169/KMK.01/1991 tanggal 21 November 1991 tentang kegiatan Sewa Guna Usaha, *Leasing* adalah setiap kegiatan pembiayaan perusahaan dalam bentuk penyediaan barang-barang modal untuk digunakan oleh suatu perusahaan untuk jangka waktu tertentu, berdasarkan pembayaran-pembayaran berkala disertai dengan hak pilih (opsi) bagi perusahaan tersebut untuk membeli barang-barang modal yang bersangkutan atau memperpanjang jangka waktu *leasing* berdasarkan nilai sisa yang telah disepakati. Sama halnya dengan lembaga keuangan lainnya, *leasing* juga tidak lepas dari permasalahan-permasalahan kredit. Masalah yang sering terjadi banyak disebabkan karena ulah konsumen, seperti konsumen yang dianggap layak ternyata menunggak angsuran, ada juga yang sampai akhirnya motor harus ditarik oleh perusahaan pemberi kredit, bahkan ada konsumen yang melarikan diri bersama dengan motor kreditannya. Munculnya permasalahan itu salah satunya diakibatkan karena ketidaktajaman analisis *account officer* (analisis kredit). *Account officer* tidak melakukan monitor secara berkala serta lemahnya pengawasan kredit setelah konsumen mendapatkan fasilitas, baik pengawasan langsung maupun administratif.

Agar kredit yang diberikan mencapai sasaran, yaitu aman, maka analisis kredit dilakukan. Analisis kredit adalah kajian yang dilakukan untuk mengetahui kelayakan dari suatu permasalahan kredit. Melalui hasil analisis kreditnya, dapat diketahui apakah usaha nasabah layak (*feasible*) dan *marketable* (hasil usaha dapat dipasarkan), dan *profitable* (menguntungkan), serta dapat dilunasi tepat waktu [2]. Mengingat resiko tidak kembalinya kredit pada sebuah *leasing* selalu ada, maka analisis kredit dengan cermat perlu dilakukan. Penilaian kuantitatif dan kualitatif dalam menganalisa kredit akan memberikan kejelasan bagi pembuat keputusan. Untuk mewujudkan hal tersebut, perlu dilakukan persiapan kredit, yaitu dengan mengumpulkan informasi dan data untuk bahan analisis. Kualitas hasil analisis tergantung pada kualitas SDM, data yang diperoleh, dan teknik analisis.

Beberapa penelitian terdahulu yang terkait dengan tema Pemberian Kredit diantaranya: Penerapan data mining dengan metode naive bayes untuk memprediksi kelayakan pengajuan kredit pada koperasi rukun artha santosa juwana pati (Heru Purwanto & Khafiizh Hastuti, 2011) [3]. Penelitian ini bertujuan untuk menemukan suatu perhitungan yang akurat untuk menghindari kredit macet pada koperasi rukun artha santosa. Hasilnya algoritma naive bayes menghasilkan tingkat akurasi 71%, dimana dalam pengujian model data, data set dibagi menjadi dua bagian yaitu data training dan data testing dengan persentase 70% untuk data training dan sisanya 30% untuk data testing. Model pohon keputusan untuk klasifikasi persetujuan kredit menggunakan algoritma c4.5 (Ivandari, 2015) [4]. Penelitian ini bertujuan menemukan sebuah perhitungan yang dapat memprediksi kredit macet pada nasabah. Hasilnya algoritma C4.5 dapat memodelkan pohon keputusan untuk klasifikasi persetujuan kredit. Model pohon keputusan yang terbentuk dapat dengan mudah merepresentasikan aturan dari klasifikasi persetujuan kredit. Tingkat akurasi klasifikasi persetujuan kredit dengan menggunakan algoritma C4.5 sebesar 94,516% dan tergolong dalam *best classification*.

Dari hasil kedua penelitian di atas, persentase tingkat akurasi yang didapatkan algoritma C4.5 lebih besar daripada algoritma naive bayes, maka dengan penelitian ini penulis mengkomparasikan kedua algoritma tersebut ke dalam sebuah penelitian, Apakah algoritma C4.5 memiliki hasil tingkat akurasi lebih baik dari algoritma naive bayes?

TINJAUAN PUSTAKA

Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data [5]. Klasifikasi merupakan bagian dari algoritma data mining, klasifikasi ini adalah algoritma yang menggunakan data dengan target (class/label) yang berupa nilai kategorikal/nominal. Proses klasifikasi didasarkan pada empat komponen mendasar, yaitu Kelas (*Class*), Prediktor (*Predictor*), Pelatihan dataset (*Training*

dataset), Dataset Pengujian (*Testing Dataset*) [6]. *Knowledge Discovery in Database* (KDD) merupakan sebuah proses dengan beberapa tingkatan, tidak sepele, interaktif dan berulang untuk identifikasi pola yang dipahami, sah, baru dan secara potensial berguna mulai dari sekumpulan data yang sangat besar [7]. KDD dikarakteristikan sebagai sebuah proses yang terdiri dari beberapa tahap operasional : *Preprocessing*, *Data Mining* dan *Post Processing*.

Algoritma C4.5

Algoritma C4.5 (sering disebut dengan pohon keputusan) yang merupakan pengembangan dari algoritma ID3. Algoritma ini memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan) [7]. Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [8] yaitu:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

S = himpunan kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

3. Kemudian hitung nilai gain menggunakan rumus:

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

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

| Si | = proporsi Si terhadap S

| S | = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua record terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua record dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam record yang dipartisi lagi.
 - c. Tidak ada record di dalam cabang yang kosong.

Algoritma Naïve Bayes

Klasifikasi Bayes [8] adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Klasifikasi Bayes juga dikenal dengan Naïve Bayes, memiliki kemampuan sebanding dengan dengan pohon keputusan dan *neural network*. Naïve Bayes, yang juga disebut *idiot's Bayes*, *simple Bayes*, dan *independence Bayes*, adalah

metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk data set berukuran besar. Mudah diinterpretasikan sehingga pengguna yang tidak punya keahlian dalam bidang teknologi klasifikasi pun bisa mengerti. Klasifikasi Bayes didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes (1702-1761), yaitu:

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)}$$

keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$ = probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$ = probabilitas hipotesis x (*prior probability*)

$P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

Naïve bayes adalah penyederhanaan metode bayes. Teorema bayes disederhanakan menjadi:

$$P(x|y) = P(y|x) P(x)$$

Penggunaan teorema Bayes pada algoritma Naïve Bayes yaitu dengan mengkombinasikan prior probability dan probabilitas bersyarat dalam sebuah rumus yang bisa digunakan untuk menghitung probabilitas tiap klasifikasi yang mungkin.

Adapun alur dari metode Naive Bayes adalah sebagai berikut :

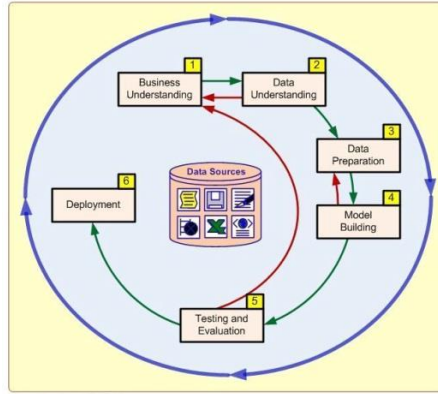
1. Input data training
2. Baca data training.
3. Hitung jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
4. Mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas.

METODE

Metode penelitian yang digunakan dalam eksperimen ini menggunakan model *Cross-Standard Industry for Data Mining* (CRISP-DM). yang terdiri dari 6 fase, yaitu:

1. *Bussines/Research Understanding Phase*
2. *Data Understanding Phase (Fase Pemahaman Data)*
3. *Data Preparation Phase (Fase Pengolahan Data)*
4. *Modeling Phase (Fase Permodelan)*
5. *Evaluation Phase (Fase Evaluasi)*
6. *Deplooyment Phase (Fase Penyebaran)*

Keenam tahap tersebut disajikan pada Gambar di bawah ini:



Source: Adapted from CRISP-DM.org

Gambar 1. Tahap CRISP-DM (Cross Industry standard process for data mining)

Dalam penelitian ini, data testing yang di uji sebanyak 210 record. Data dapat dilihat pada gambar di bawah ini:

no	Status	Jumlah tanggungan	Pendidikan	usia	Pekerjaan	Penghasilan perbulan	Remark	no	Status	Jumlah tanggungan	Pendidik an	usia	Pekerjaan	Penghasilan perbulan	Remark
1	Menikah	1	>S1	<21 / >60	Dokter	>3 Slip	GOOD	71	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
2	Menikah	1	>S1	21-55	Dokter	>2 Slip	GOOD	72	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
3	Menikah	1	SL	21-55	PNS	>3 Slip	GOOD	73	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
4	Menikah	1	SL	21-55	TNI/POLRI	>3 Slip	GOOD	74	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
5	Menikah	1	SL	21-55	TNI/POLRI	>3 Slip	GOOD	75	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD
6	Menikah	1	SL	21-55	Karyawan	>2 Slip	GOOD	76	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD
7	Menikah	1	SL	21-55	Karyawan	>2 Slip	GOOD	77	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD
8	Menikah	1	SL	21-55	Karyawan	>1 Slip	GOOD	78	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD
9	Menikah	1	SL	21-55	Karyawan	>1 Slip	GOOD	79	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
10	Menikah	1	SL	21-55	Karyawan	>1 Slip	GOOD	80	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
11	Menikah	1	SL	21-55	Karyawan	>1 Slip	GOOD	81	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
12	Menikah	1	SL	21-55	Karyawan	>3 Slip	GOOD	82	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
13	Menikah	1	SL	21-55	Karyawan	>3 Slip	GOOD	83	Menikah	2-3	SLTA	<21 / >60	Guru	>1 Slip	GOOD
14	Menikah	1	SL	21-55	Karyawan	>3 Slip	GOOD	84	Menikah	2-3	SLTA	<21 / >60	Guru	>2 Slip	GOOD
15	Menikah	1	SL	21-55	Karyawan	>3 Slip	GOOD	85	Menikah	2-3	SLTA	<21 / >60	Guru	>2 Slip	GOOD
16	Menikah	1	SL	21-55	Karyawan	>2 Slip	GOOD	86	Menikah	2-3	SLTA	<21 / >60	Guru	>3 Slip	GOOD
17	Menikah	1	SL	21-55	Karyawan	>1 Slip	GOOD	87	Menikah	2-3	SLTA	<21 / >60	Guru	>3 Slip	GOOD
18	Menikah	1	SL	21-55	Karyawan	>1 Slip	GOOD	88	Menikah	2-3	SLTA	<21 / >60	Dosen	>3 Slip	GOOD
19	Menikah	1	SL	21-55	Karyawan	>3 Slip	GOOD	89	Menikah	2-3	SLTA	<21 / >60	Dosen	>3 Slip	GOOD
20	Menikah	1	Diploma	21-55	Karyawan	>3 Slip	GOOD	90	Menikah	1	SLTA	55-60	Guru	>3 Slip	GOOD
21	Menikah	1	Diploma	21-55	Karyawan	>3 Slip	GOOD	91	Menikah	1	Diploma	21-55	Guru	>2 Slip	GOOD
22	Menikah	1	Diploma	21-55	Karyawan	>1 Slip	GOOD	92	Menikah	1	SL	21-55	Guru	>2 Slip	GOOD
23	Menikah	1	Diploma	21-55	Karyawan	>2 Slip	GOOD	93	Menikah	1	SLTA	<21 / >60	Guru	>2 Slip	GOOD
24	Menikah	1	Diploma	21-55	Karyawan	>2 Slip	GOOD	94	Menikah	2-3	SLTA	21-55	Guru	>3 Slip	GOOD
25	Menikah	1	Diploma	21-55	Karyawan	>3 Slip	GOOD	95	Menikah	>3	SL	21-55	Guru	>3 Slip	GOOD
26	Menikah	1	Diploma	21-55	Karyawan	>3 Slip	GOOD	96	Menikah	>3	Diploma	21-55	Guru	>3 Slip	GOOD
27	Menikah	1	Diploma	21-55	Karyawan	>3 Slip	GOOD	97	Menikah	Tidak ada	Diploma	21-55	Guru	>3 Slip	GOOD
28	Menikah	1	Diploma	21-55	Karyawan	>3 Slip	GOOD	98	Menikah	Tidak ada	SL	21-55	Guru	>3 Slip	GOOD
29	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD	99	Menikah	2-3	SLTA	21-55	Dosen	>3 Slip	GOOD
30	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	100	Menikah	2-3	SLTA	21-55	Dosen	>3 Slip	GOOD
31	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	101	Menikah	2-3	SLTA	21-55	Dosen	>3 Slip	GOOD
32	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD	102	Menikah	2-3	SLTA	21-55	Dosen	>3 Slip	GOOD
33	Menikah	1	SLTA	21-55	Wiraswasta	>1 Tanpa	GOOD	103	Menikah	2-3	SLTA	21-55	Dosen	>3 Slip	GOOD
34	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD	104	Menikah	2-3	SLTA	21-55	Dosen	>3 Slip	GOOD
35	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD	105	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
36	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD	106	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
37	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	107	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
38	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	108	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
39	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	109	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
40	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	110	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
41	Menikah	1	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD	111	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
42	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD	112	Menikah	2-3	SLTA	21-55	Lawyer	>3 Slip	GOOD
43	Menikah	1	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD	113	Menikah	2-3	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
44	Menikah	1	SLTA	21-55	Karyawan	>1 Slip	GOOD	114	Menikah	2-3	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
45	Menikah	1	SLTA	21-55	Karyawan	>1 Slip	GOOD	115	Menikah	2-3	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
46	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	116	Menikah	2-3	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
47	Menikah	1	SLTA	21-55	Karyawan	>2 Slip	GOOD	117	Menikah	2-3	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
48	Menikah	1	SLTA	21-55	Karyawan	>1 Slip	GOOD	118	Menikah	2-3	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
49	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	119	Menikah	2-3	SLTP	21-55	Karyawan	>2 Slip	BAD
50	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	120	Menikah	2-3	SLTP	55-60	Wiraswasta	>2 Tanpa	GOOD
51	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	121	Menikah	2-3	SD	55-60	Wiraswasta	>2 Tanpa	BAD
52	Menikah	1	SLTA	21-55	Karyawan	>2 Slip	GOOD	122	Menikah	2-3	SD	55-60	Wiraswasta	>3 Tanpa	GOOD
53	Menikah	1	SLTA	21-55	TNI/POLRI	>2 Slip	GOOD	123	Menikah	2-3	SD	55-60	Wiraswasta	>3 Tanpa	GOOD
54	Menikah	1	SLTA	21-55	TNI/POLRI	>2 Slip	GOOD	124	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>2 Slip	BAD
55	Menikah	1	SLTA	21-55	TNI/POLRI	>3 Slip	GOOD	125	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>2 Slip	BAD
56	Menikah	1	SLTA	21-55	TNI/POLRI	>3 Slip	GOOD	126	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>2 Slip	BAD
57	Menikah	1	SLTA	21-55	TNI/POLRI	>3 Slip	GOOD	127	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>2 Slip	BAD
58	Menikah	1	SLTA	21-55	PNS	>3 Slip	GOOD	128	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>1 Slip	BAD
59	Menikah	1	SLTA	21-55	PNS	>2 Slip	GOOD	129	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>1 Slip	BAD
60	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	130	Belum Menikah	Tidak ada	SLTA	21-55	PNS	>2 Slip	BAD
61	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	131	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
62	Menikah	1	SLTA	21-55	Karyawan	>2 Slip	GOOD	132	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
63	Menikah	1	SLTA	21-55	Karyawan	>2 Slip	GOOD	133	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
64	Menikah	1	SLTA	21-55	Karyawan	>1 Slip	GOOD	134	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
65	Menikah	1	SLTA	21-55	Karyawan	>2 Slip	GOOD	135	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
66	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	136	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>2 Tanpa	GOOD
67	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	137	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
68	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	138	Belum Menikah	Tidak ada	SLTA	21-55	Wiraswasta	>3 Tanpa	GOOD
69	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	139	Belum Menikah	Tidak ada	SLTA	21-55	Karyawan	>1 Slip	BAD
70	Menikah	1	SLTA	21-55	Karyawan	>3 Slip	GOOD	140	Belum Menikah	Tidak ada	SLTA	21-55	Karyawan	>2 Slip	BAD

Tabel 1. Model *Confusion matrix* untuk Metode C4.5

Accuracy: 91.43% +/- 6.32% (mikro: 91.43%)			
Precision: 88.15% +/- 9.04% (mikro: 87.65%)			
Recall: 90.00% +/- 9.35% (mikro: 89.87%)			
	True Good	True Bad	Class Precision
Prediksi Good	121	8	93.80%
Prediksi Bad	10	71	87.65%
Class Recall	92.37%	89.87%	

Selanjutnya dengan metode Naïve Bayes, berdasarkan data *training*, pada Tabel 2 diketahui dari 210 data, 122 diklasifikasikan *Good* sesuai prediksi, lalu 13 data diprediksi *Good* tetapi ternyata *Bad*, 66 data *Bad* sesuai prediksi, dan 9 data diprediksi *Bad* ternyata *Good*.

Tabel 2. Model *Confusion matrix* untuk Metode Naïve Bayes

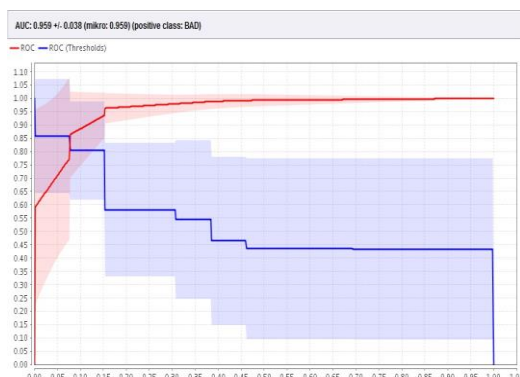
Accuracy: 89.52% +/- 5.13% (mikro: 89.52%)			
Precision: 88.71% +/- 8.22% (mikro: 88.00%)			
Recall: 83.75% +/- 9.76% (mikro: 83.54%)			
	True Good	True Bad	Class Precision
Prediksi Good	122	13	90.37%
Prediksi Bad	9	66	88.00%
Class Recall	93.13%	83.54%	

Dari kedua tabel *Confusion matrix*, selanjutnya dilakukan perhitungan nilai *Accuracy*, *Precision*, dan *Recall*. Perbandingan nilai *Accuracy*, *Precision* dan *Recall* yang telah dihitung untuk metode C4.5, dan Naïve Bayes dapat dilihat pada table 3

Tabel 3. Perbandingan perhitungan *Accuracy*, *Precision*, dan *Recall*

Perbandingan	C4.5	Naïve Bayes
Accuracy	91.43%	89.52%
Precision	88.15%	88.71%
Recall	90.00%	83.75%

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua metode komparasi bisa dilihat pada gambar di bawah ini:



Gambar 2. Kurva ROC C4.5



Gambar 3. Kurva ROC Naïve Bayes

Pada gambar 2 di atas menunjukkan hasil dari perhitungan algoritma C4.5 dalam gambar menunjukkan nilai *false positive rate* yang ditunjukkan dengan garis *vertical*, memiliki rentan nilai 0 hingga 1,05 dan nilai *true positive rate* yang ditunjukkan dengan nilai *horizontal*, memiliki rentan nilai 0 hingga 1,10. Berdasarkan hasil kinerja pada gambar diatas, menunjukkan kurva berwarna merah lebih bagus dibanding kurva berwarna biru, karena garis kurva berwarna merah lebih jauh dari garis bujur (0,0). Untuk membandingkan nilai kinerja kurva merah dan biru dalam bentuk

angka maka dapat dilakukan dengan membandingkan luas area dibawah kurva atau *Area Under Curve* (AUC). Dari gambar dapat disimpulkan luas area kurva berwarna merah lebih besar dibanding luas kurva berwarna biru, dengan nilai luas AUC sebesar 0,959.

Sedangkan pada gambar 3 di atas menunjukkan hasil dari perhitungan algoritma Naïve Bayes. Dalam gambar menunjukkan nilai *false positive rate* yang ditunjukkan dengan garis *vertical*, memiliki rentan nilai 0 hingga 1,05 dan nilai *true positive rate* yang ditunjukkan dengan nilai *horizontal*, memiliki rentan nilai 0 hingga 1,1. Berdasarkan hasil kinerja pada gambar diatas, menunjukkan kurva berwarna merah dan warna biru memiliki kinerja yang hampir sama, karena garis kurva berwarna merah dan warna biru sama-sama menjauh dari garis bujur (0,0). Untuk membandingkan nilai kinerja kurva merah dan biru dalam bentuk angka maka dapat dilakukan dengan membandingkan luas area dibawah kurva atau *Area Under Curve* (AUC). Dari gambar dapat disimpulkan luas area kurva berwarna merah lebih besar dibanding luas kurva berwarna biru, dengan nilai luas AUC sebesar 0,949

Tabel 4. Perbandingan nilai accuracy dan AUC

Perbandingan	C4.5	Naïve Bayes
Accuracy	91.43%	89.52%
AUC	0,959	0,949

KESIMPULAN

Berdasarkan hasil penelitian dari komparasi kedua metode dalam identifikasi pemberian kredit, dapat diambil kesimpulan bahwa:

- Berdasarkan komparasi metode klasifikasi data mining yaitu C4.5 dan Naïve Bayes menunjukkan bahwa metode C4.5 lebih akurat dari pada metode Naïve Bayes. ini dilihat dari nilai accuracy dimana metode C4.5 memiliki nilai accuracy sebesar 91.43%, lebih tinggi dibanding metode Naïve Bayes dengan nilai accuracy 89.52%.
- Berdasarkan metode C4.5 atribut status perkawinan merupakan atribut yang paling berpengaruh terhadap pemberian kredit, ini dibuktikan dengan atribut status perkawinan sebagai root node.
- Berdasarkan nilai AUC kedua metode ini masuk dalam kategori klasifikasi yang sangat baik.

DAFTAR PUSTAKA

- [1] Arthesa, Ade dan Edia Handiman, 2006, Bank & Lembaga Keuangan Bukan Bank, Jakarta: Indeks Kelompok Gramedia.
- [2] Rivai, Veithzal dan Veithzal, Andria Permata, 2006. *Credit Management Handbook*. Jakarta: Raja Grafindo Persada.
- [3] Purwanto, Heru dan Khafiizh Hastuti. (2011). "Penerapan Data Mining Dengan Metode Naive Bayes Untuk Memprediksi Kelayakan Pengajuan Kredit Pada Koperasi Rukun Artha Santosa Juwana Pati". Semarang: Universitas Dian Nuswantoro.
- [4] Ivandari. (2015). "Model Pohon Keputusan Untuk Klasifikasi Persetujuan Kredit Menggunakan Algoritma C4.5" Pekalongan: STMIK WidyaPratama.
- [5] Shraya Taruna R, Saroj Hiranwal, 2013. Enhanced Naïve Bayes Algorithm for Intrusion Detection in Data Mining, International Journal of Computer Science and information Technologies, Vol. 4, 2013.
- [6] Bustami, "Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi" TECHSI, vol. III, pp. 11-14, Oktober 2014.
- [7] Silvia Rissino, Germano Lambert-Torres, 2009. Rough Set Theory-Fundamental Concepts, Principals, Data extraction, and Applications, Data Mining and Knowledge Discovery in Real Life Application, February 2009 I-Tech, Vienna, Austria.
- [8] Kusrini, & Luthfi, E. T. (2009). Algoritma Data Mining. Yogyakarta: Andi Publishing