

KLASIFIKASI DATA MINING UNTUK MENENTUKAN POTENSI KREDIT MACET PADA KOPERASI SIMPAN PINJAM PRIMKOVERI WALERI MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

Awaludin Rizky Wicaksono¹, Yuniarsi Rahayu²

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
Jl. Nakula I No. 5-11 Semarang Telp. (024) 3517261
E-mail : awaludinrizky14@gmail.com¹, yuniarsi.rahayu@dsn.dinus.ac.id²

Abstrak

Analisis klasifikasi prediksi nasabah yang berpotensi kredit macet yang ada di Koperasi Primkoveri masih dilakukan secara sederhana dikarenakan keterbatasan koperasi dalam melibatkan tenaga analis kredit yang membutuhkan biaya cukup mahal dan seringkali memberatkan nasabah, kesederhanaan analisis inilah yang memicu terjadinya kredit macet . Antisipasi yang biasa dilakukan hanyalah dengan melakukan pendekatan-pendekatan personal pada semua nasabah kredit Padahal seharusnya suatu hasil analisis yang baik adalah yang mampu menghasilkan data yang akurat dan tepat dalam proses pembuatannya . Dengan adanya alasan yang telah disebutkan, penulis memilih untuk melakukan penghitungan analisis klasifikasi prediksi nasabah yang berpotensi kredit macet dengan cara lebih efektif yang dapat digunakan untuk menggantikan cara sederhana yaitu dengan menggunakan data mining. Parameter yang digunakan untuk menentukan potensi kredit macet adalah Umur, Status, Pekerjaan, Penghasilan, Maks Kredit, Jangka Waktu Kredit dan Angsuran. Metode yang digunakan dalam klasifikasi adalah metode pengklasifikasian dengan menggunakan desicion tree (pohon keputusan) dengan salah satu algoritma desicion tree yang dipakai adalah C4.5. Hasil dari penelitian dari permasalahan yang dikembangkan dapat disimpulkan bahwa pengimplementasian metode desicion tree terhadap data nasabah kredit Koperasi Primkoveri Waleri tahun 2014 memiliki tingkat akurasi yang baik dalam menyelesaikan solusi klasifikasi data mining dengan algoritma C4.5 yaitu sebesar 86%.

Kata Kunci: Koperasi Primkoveri, Potensi Kredit Macet , Klasifikasi, Pohon Keputusan, Algoritma C4.5

Abstract

Analysis of the classification of the customer predictions that potentially bad credit in Primkoveri Cooperation is still use manual way due to the limitations of the cooperation in the credit analyst power involves a quite expensive cost and often complicating the customer , the simplicity of the analysis here's what triggered the bad credit. Usually, Anticipation are by doing personal approaches to all credit customer.A good analysis should an analysis that capable to generating an accurate and precise data in the manufacturing process. With the reasons already mentioned, the author chose to do the calculation analysis of the classification customer prediction that potentially had a bad debts more effectively which can be used to replace the manual way is to use data mining. The parameters used to determine the potential bad debts are the Age, Status, employment, income, Credit, Max length of time credit and Installment. The methods used in the classification is a method of classification by using desicion tree with one of the algorithms used is tree desicion C 4.5. The results of the research of the problems that developed is can be concluded that the implementation method of the customer data against tree desicion Credit Cooperative Primkoveri Waleri in year 2014 has a good degree of accuracy in completing the solution of classification data mining algorithm with C4.5 is up to 86%.

Keywords: Primkoveri Cooperation, Potentially Bad Credit, Classification, Decision Tree, C4.5 Algorithm

I. PENDAHULUAN

Lembaga finansial adalah suatu lembaga yang bergerak dalam bidang keuangan dimana para nasabahnya terdiri dari berbagai lapisan dan bermacam-macam perilaku. Masyarakat Indonesia dalam melakukan transaksi peminjaman dana dari bank masih banyak menggunakan analisa kebutuhan secara pribadi walaupun dalam pengajuannya menggunakan analisa perusahaan, hal ini dikarenakan tuntutan yang harus dipenuhi karena persyaratan perbankan. Masih banyak lembaga finansial baik itu perbankan maupun koperasi simpan pinjam yang masih mempunyai jumlah kredit macet yang cukup besar, hal ini akan sangat mengganggu kelangsungan koperasi maupun perbankan [1].

Koperasi Primkoveri adalah salah satu koperasi yang bergerak dibidang simpan pinjam dan menawarkan permohonan kredit. Permohonan kredit pada koperasi simpan pinjam biasanya dilakukan secara sederhana tidak seperti di bank, namun masih tetap memenuhi persyaratan prinsip dari sebuah kredit. Penilaian yang terkait dengan jumlah pinjaman, waktu pengembalian, jenis usaha/pekerjaan, jumlah, status perkawinan, angsuran dan ketentuan-ketentuan lain. Penyusunan perencanaan kredit secara umum harus memenuhi kriteria dapat dilaksanakan (*feasible*), dapat memberikan arah yang kita kehendaki (*suitable*), dapat diterima (*acceptable*), mempunyai nilai yang berarti (*valuable*), mudah dicapai (*achievable*), dan hasilnya dapat dievaluasi/diukur (*measurable*) [2].

Analisa yang dilakukan oleh para manajer koperasi simpan pinjam dilakukan secara sederhana karena keterbatasan koperasi dalam melibatkan

tenaga analis kredit yang cukup mahal dan memberatkan nasabah, karena kesederhanaan analisis inilah yang memicu terjadinya kredit macet. Antisipasi yang biasa dilakukan hanyalah dengan melakukan pendekatan-pendekatan personal pada semua nasabah kredit, ini memang tidak efektif manakala jumlah nasabah kredit menunggak sangat banyak [1].

Mengacu dari masalah tersebut maka penulis tertarik untuk mengklasifikasi atau mengelompokkan suatu data mining yang dapat menyajikan informasi secara cepat dan akurat serta mudah digunakan yaitu mengklasifikasi data mining untuk menentukan potensi kredit macet pada Koperasi Simpan Pinjam Primkoveri Waleri menggunakan algoritma *Decision Tree* C4.5 sehingga dapat menjadi acuan pengambilan kredit selanjutnya.

II. TEORI PENUNJANG

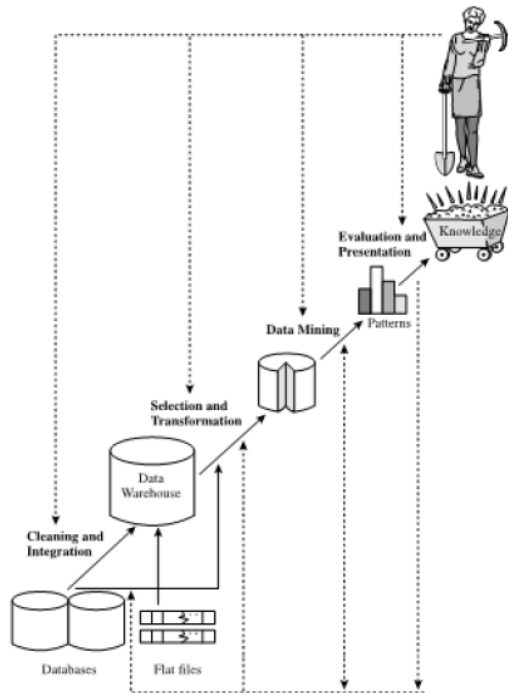
2.1 Data Mining

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar [3]. Hal penting yang terkait dengan data mining adalah sebagai berikut :

1. Data mining merupakan proses otomatis terhadap data yang dikumpulkan di masa lalu.
2. Data yang akan digunakan dalam proses data mining berupa data yang sangat besar.
3. Tujuan dari data mining adalah menemukan hubungan-hubungan atau pola-pola yang mungkin

memberikan indikasi yang bermanfaat.

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan pada Gambar 2.1 di bawah ini.



Gambar 2.1 Tahap-tahap Data Mining

Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantara knowledge base. Tahapan-tahapan tersebut, diantaranya:

1. Pembersihan data (untuk membuang data yang tidak konsisten dan noise).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Seleksi dan Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining).
4. Aplikasi teknik data mining.
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai).
6. Presentasi pola yang ditemukan untuk menghasilkan aksi.

2.2 Teknik Klasifikasi

Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi (taksonomi) merupakan proses penempatan objek atau konsep tertentu ke dalam satu set kategori berdasarkan objek yang digunakan. Salah satu teknik klasifikasi yang populer digunakan adalah decision tree [4]. Klasifikasi sendiri terbagi menjadi dua tahap, yaitu pengklasifikasian dan pembelajaran. Pada tahap pembelajaran, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis training data. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan $y=f(x)$ di mana y adalah kelas hasil prediksi dan X adalah tuple yang ingin diprediksi kelasnya.

Pada tahap pengklasifikasian, model yang telah dihasilkan akan digunakan untuk melakukan klasifikasi terhadap data-data yang belum diketahui. Akan tetapi, sebuah model hanya boleh digunakan untuk klasifikasi jika model tersebut mempunyai tingkat akurasi yang cukup tinggi..

2.3 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang telah secara luas digunakan, khususnya di area machine learning yang memiliki beberapa perbaikan dari algoritma sebelumnya yaitu ID3. Algoritma C4.5 dan ID3 model yang tak terpisahkan, karena membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5 Diakhir tahun 1980- an, J. Ross Quinlan seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3. Walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian

Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis supervised learning [7].

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 [4], yaitu:

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

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

Keterangan :

S = Himpunan Kasus

n = Jumlah Partisi S

pi = proporsi S_i terhadap S

Kemudian setelah nilai entropy pada masing-masing atribut sudah diperoleh maka hitung nilai gain dengan 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

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = Jumlah Kasus dalam S

2.4 Pohon Keputusan

Algoritma decision tree merupakan

salah satu algoritma klasifikasi di dalam data mining yang bekerja berdasarkan teori informasi (information theory). Decision tree menggunakan representasi struktur pohon (tree) dimana setiap node mempresentasikan atribut, cabangnya mempresentasikan nilai dari atribut, dan daun mempresentasikan kelas. Decision tree juga memiliki beberapa keunggulan, yaitu mudah dalam pengembangan sebuah model, mudah dipahami oleh penggunanya, dan mampu menangani noisy data dan unknown data.

Banyak algoritma yang digunakan dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5 [6]. Data dalam pohon keputusan biasanya berbentuk tabel dengan atribut. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon.

2.5 Confusion Matrix

Confusion matrix merupakan sebuah table yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi. [5].

Tabel 2.1 Tabel Confusion Matrik 2 Kelas

Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a (true positive-TP)	b (false negative-FN)
Class = No	c (false positive-FP)	d (true negative-TN)

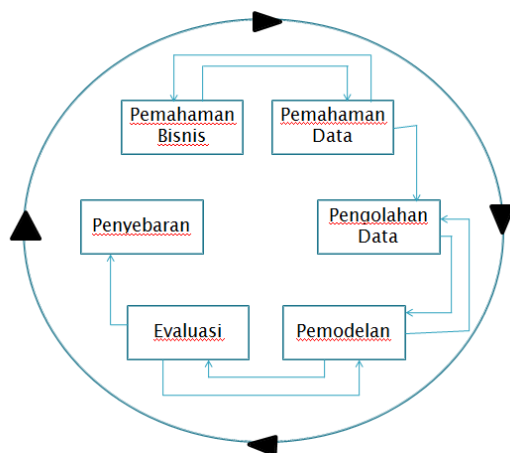
Rumus untuk menghitung tingkat akurasi pada matrik adalah :

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} = \frac{A+D}{A+B+C+D}$$

III. METODE PENELITIAN

3.1 Desain Penelitian

CRISP-DM menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase (Gambar.3.1). Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antarfase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase pemodelan. Berdasar pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase pengolahan data untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase evaluasi [3].



Gambar 3.1 Proses Data Mining menurut CRISP-DM

a. Pemahaman Bisnis

Permohonan kredit pada koperasi

simpan pinjam biasanya dilakukan secara sederhana tidak seperti di bank, namun masih tetap memenuhi persyaratan prinsip dari sebuah kredit. Penilaian yang terkait dengan jumlah pinjaman, waktu pengembalian, jenis usaha/pekerjaan, jumlah, status perkawinan, angsuran dan ketentuan-ketentuan lain. Permohonan ini biasanya diisi oleh calon nasabah dalam blanko yang telah tersedia untuk berkas pengajuan kredit, selebihnya ditentukan oleh hasil analisis team yang telah ditentukan koperasi simpan pinjam.

b. Pemahaman Data

Data yang digunakan dalam penelitian ini adalah sumber data primer. Data diperoleh dari data nasabah yang mengajukan kredit di Koperasi Simpan Pinjam Primkoveri Weleri sebanyak 300 sebagai bahan uji.

Atribut yang dijadikan sebagai acuan adalah Umur, Status, Pekerjaan, Penghasilan, Maks Kredit, Jangka Waktu Kredit Angsuran dan atribut Hasil yaitu Kredit Lancar atau Macet sebagai keterangan bahwa calon nasabah tersebut berpotensi kredit macet atau lancar.

c. Pengolahan Data

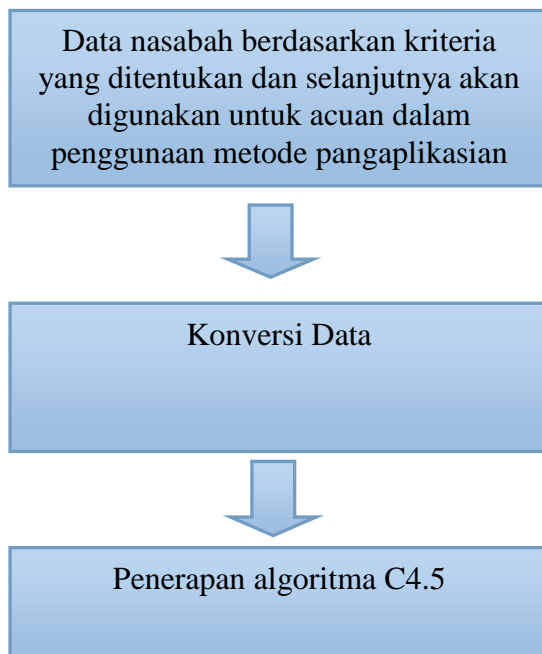
Data kemudian dilakukan pemilihan atribut dan sebagian dari data dalam atribut yang ada akan dikonversikan untuk memudahkan proses data mining, karena data akan diproses dengan tools bantu data mining.

Tabel 3.1 Detail Atribut Data Penelitian

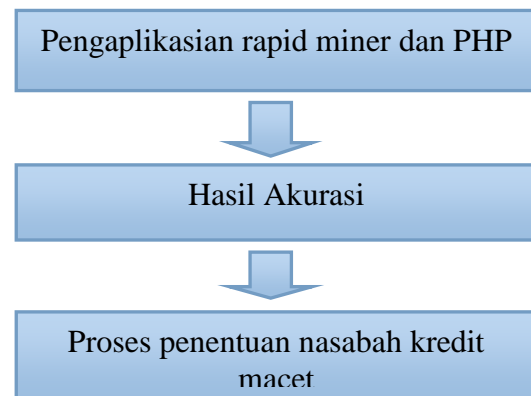
Atribut	Detail Penggunaan	
Nomor Kredit	√	Nilai Model
Nama Nasabah	×	No
Alamat	×	No
Jenis Kelamin	×	No
Umur	√	Nilai Model
Agama	×	No
Status	√	Nilai Model
Pekerjaan	√	Nilai Model
Penghasilan	√	Nilai Model
Maksimal Kredit	√	Nilai Model
Jangka Waktu	√	Nilai Model
Angsuran	√	Nilai Model
Hasil	√	Label Target

d. Permodelan

Metode yang akan digunakan dalam penelitian ini adalah Algoritma C4.5. Lemudian pengukuran akurasi dalam penelitian ini akan menggunakan perhitungan manual dan *Rapid Miner* sebagai pembanding.



Lanjutan



Gambar 3.2. Alur permodelan yang diusulkan

e. Validasi dan Evaluasi

Dalam tahapan ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang terdapat dalam framework *Rapid Miner* Confusion Matrix untuk pengukuran tingkat akurasi model, dan pengujian manual.

f. Penyebaran

Hasil dari penelitian ini berupa analisa yang mengarah ke decision Suport System (DSS), yang diharapkan dapat digunakan oleh Koperasi Simpan Pinjam Weleri dalam menyeleksi nasabah yang ingin mengajukan kredit dan juga dapat digunakan sebagai bahan rujukan untuk penelitian selanjutnya.

IV. HASIL DAN PEMBAHASAN

Table 4.1: Data Nasabah Pemohon Kredit Tahun 2014

No	No Kredit	Umur	Status	Pekerjaan	Penghasilan	Maks Kredit	JW Kredit (Bulan)	Angsuran	Hasil
1	1301010161936	37	KAWIN	PNS	1610400	149000000	120	402600	Kredit lancar
2	1301010175676	41	KAWIN	PNS	1613600	450000000	180	403400	Kredit lancar
3	1301010177644	29	KAWIN	PNS	1643200	483000000	240	410800	Kredit lancar
4	1301020916119	35	KAWIN	PNS	1728400	200000000	180	432100	Kredit lancar
5	1301020902649	28	BELUM KAWIN	WIRASWASTA	1728800	220000000	180	432200	Kredit macet
6	1301010155529	50	KAWIN	PNS	1822000	960000000	180	455500	Kredit macet
7	1301010154638	33	KAWIN	SWASTA	1822000	1000000000	180	455500	Kredit macet
8	1301020915163	36	KAWIN	PNS	1859200	1580000000	180	464800	Kredit macet
9	1301020913569	34	KAWIN	WIRASWASTA	1859200	2040000000	180	464800	Kredit lancar
10	1301030099339	44	KAWIN	WIRASWASTA	1878400	856000000	120	469600	Kredit lancar
...
296	1301020902982	51	KAWIN	PNS	2091300	2640000000	180	697100	Kredit lancar
297	1301020910820	33	KAWIN	SWASTA	2119200	860000000	180	706400	Kredit macet
298	1301020901481	30	KAWIN	PNS	2196000	1120000000	180	732000	Kredit macet
299	1301020902916	29	BELUM KAWIN	PNS	2682900	4400000000	180	894300	Kredit macet
300	1301010168629	37	KAWIN	PNS	2707800	5500000000	240	902600	Kredit macet

a. Konversi Data

Proses konversi data asli pelamar menjadi sebuah bilangan bulat untuk mempermudah perhitungan manual dan *Rapid Miner* untuk melakukan proses klasifikasi.

Tabel 4.2 : Data pelamar setelah dikonversi

No	No Kredit	Umur	Status	Pekerjaan	Penghasilan	Maks Kredit	JW Kredit (Bulan)	Angsuran	Hasil
1	1301010161936	37	KAWIN	PNS	1610400	149000000	120	402600	Kredit lancar
2	1301010175676	41	KAWIN	PNS	1613600	450000000	180	403400	Kredit lancar
3	1301010177644	29	KAWIN	PNS	1643200	483000000	240	410800	Kredit lancar
4	1301020916119	35	KAWIN	PNS	1728400	200000000	180	432100	Kredit lancar
5	1301020902649	28	BELUM KAWIN	WIRASWASTA	1728800	220000000	180	432200	Kredit macet
6	1301010155529	50	KAWIN	PNS	1822000	960000000	180	455500	Kredit macet
7	1301010154638	33	KAWIN	SWASTA	1822000	1000000000	180	455500	Kredit macet
8	1301020915163	36	KAWIN	PNS	1859200	1580000000	180	464800	Kredit macet
9	1301020913569	34	KAWIN	WIRASWASTA	1859200	2040000000	180	464800	Kredit lancar
10	1301030099339	44	KAWIN	WIRASWASTA	1878400	856000000	120	469600	Kredit lancar
...
296	1301020902982	51	KAWIN	PNS	2091300	2640000000	180	697100	Kredit lancar
297	1301020910820	33	KAWIN	SWASTA	2119200	860000000	180	706400	Kredit macet
298	1301020901481	30	KAWIN	PNS	2196000	1120000000	180	732000	Kredit macet
299	1301020902916	29	BELUM KAWIN	PNS	2682900	4400000000	180	894300	Kredit macet
300	1301010168629	37	KAWIN	PNS	2707800	5500000000	240	902600	Kredit macet

Konversi data ini merupakan proses penting dalam perhitungan dalam sistem yang dibangun agar

memudahkan pengkodean dalam pembuatan sistem.

Dari Tabel 4.2 di atas diberikan penjelasan keterangan dari table tersebut diatas yaitu diambil dari tujuh variabel terpenting sebagai atribut untuk menentukan label / keluaran yaitu sebagai berikut:

Tabel 4.3 Range atau Interval

UMUR	
>35	
<=35	
STATUS	
Kawin	
Belum Kawin	
PEKERJAAN	
PNS	
Wiraswasta	
Swasta	
PENGHASILAN	
>3341350	
<=3341350	
MAKS KREDIT	
>92500000	
<=92500000	
JWAKTU KREDIT	
>138	
<=138	
ANGSURAN	
>4672000	
<=4672000	

b. Pengaplikasian PHP

Tampilan pada gambar 4.1 dan gambar 4.2 merupakan tampilan yang menampilkan menu import dari *Ms.Excel* dan tampilan hasil yang dapat diakses oleh user. Desain sistem yang digunakan yaitu menggunakan PHP.



Gambar 4.1. Tampilan import dari Ms.Excel



Gambar 4.2. Hasil dari import Ms.Excel

C. Hasil Akurasi

Untuk menentukan tingkat akurasi ini menggunakan data testing bisa dilihat di gambar 4.5 dengan acuan table 4.2 dengan data berjumlah 200 kemudian bisa dihitung menggunakan tabel confusion matrix dibawah ini:

Tabel 4.4. Confussion matrix

Correct Classification	Classified as	
	+	-
+	TP= 14	FN=6
-	FP= 1	TN=29

Tingkat akurasi dari seluruh klasifikasi ditentukan dengan jumlah klasifikasi yang benar dibagi dengan total jumlah record klasifikasi.

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{14+29}{14+1+6+29} \\
 &= 0.86
 \end{aligned}$$

Untuk menghitung prosentasi akurasi, maka tingkat sukses dikalikan 100%. Ini berarti prosentase error dapat dicari dengan cara 100% dikurangi dengan prosentase sukses.

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{14+29}{14+1+6+29} * 100 \\
 &= 0.86\%
 \end{aligned}$$

IV. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian pada klasifikasi prediksi nasabah yang berpotensi kredit macet dapat diambil beberapa kesimpulan sebagai berikut:

1. Klasifikasi data mining untuk menentukan penyebab kredit macet pada Koperasi Simpan Pinjam Primkoveri Waleri dapat mengklasifikasikan nasabah tersebut masuk dalam tahapan kredit lancar atau tidak.
2. Dari 50 data nasabah yang digunakan menunjukkan tingkat akurasi dengan algoritma C4.5 sebesar 86%.

5.2 Saran

Saran yang diajukan dalam penelitian ini yaitu untuk penelitian selanjutnya dengan permasalahan yang sama dengan metode yang sama sehingga dapat ditingkatkan salah satunya dengan melakukan pruning terhadap algoritma C4.5 jika pohon yang terbentuk terlalu besar, dengan melakukan pruning maka pohon yang terbentuk akan otomatis terjadi pemangkasan walupun dengan jumlah data yang sangat besar. Ini dilakukan untuk mempermudah kinerja

dari algoritma C4.5 tanpa mengurangi tingkat akurasi yang nantinya akan diperoleh.

Selain itu saran dari penulis adalah penerapan rules dari algoritma C4.5 selanjutnya dibuat sistem aplikasi yang lebih baik lagi sehingga dapat diperuntukkan untuk pihak Koperasi Simpan Pinjam Primkoveri Cabang Weleri karena ini masih sebatas prototype sehingga dapat digunakan dalam klasifikasi hasil proses seleksi yang berupa lancar atau tidaknya nasabah pemohon kredit untuk bahan pertimbangan bagi pemohon kredit yang berikutnya.

DAFTAR PUSTAKA

- [1] Mudrajat Kuncoro, Suhardjono, "Manajemen Perbankan" Teori dan Aplikasi, Edisi Kedua, Fakultas Ekonomi dan Bisnis UGM, 2012.
- [2] Dr. (Cand) Taswan, SE, M.Si, "Manajemen Perbankan" Konsep, Teknik dan Aplikasi, Edisi Kedua UPP STIM YKPN, Yogyakarta, 2010.
- [3] Larose, Daniel T, "Discovering Knowledge in Data: An Introduction to Data Mining," John Willey & Sons, Inc, 2005.
- [4] Anik Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," 2012.
- [5] Liliana Swastina, Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," 2013.
- [6] Swastina Liliana, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," Gema Aktualita, Juni 2013.
- [7] Fatayat and Joko Risanto, "Proses Data Mining dalam Meningkatkan Sistem Pembelajaran pada Pendidikan Sekolah Menengah Pertama," 2013.