

**SKRIPSI**  
**PENERAPAN DATAMINING KLASIFIKASI POLA NASABAH**  
**MENGGUNAKAN ALGORITMA C4.5 PADA BANK BRI**  
**BATANG**

***APPLICATION OF DATA MINING FOR CLASSIFICATION OF***  
***CUSTOMER PATTERN USING C4.5 ALGORITHM ON BANK***  
***BRI BATANG***

Diajukan untuk memenuhi salah satu syarat memperoleh gelar

Sarjana Teknik Informatika



Disusun Oleh :

Nama : Ayu Rizqi Oktaviana

NIM : A11.2011.06042

Program Studi : Teknik Informatika – S1

**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS DIAN NUSWANTORO**  
**SEMARANG**

**2016**

## **PERSETUJUAN SKRIPSI**

Nama : Ayu Rizqi Oktaviana  
NIM : A11.2011.06042  
Program Studi : Teknik Informatika-S1  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Penerapan Data Mining Klasifikasi Pola Nasabah  
Menggunakan Algoritma C4.5 pada Bank BRI Batang

Tugas Akhir ini telah diperiksa dan disetujui,

Semarang, 22 Februari 2016

Menyetujui,  
Pembimbing

**Heribertus Himawan, M.Kom**

Mengetahui,  
Dekan Fakultas Ilmu Komputer

**Dr. Abdul Syukur**

## **PENGESAHAN DEWAN PENGUJI**

Nama : Ayu Rizqi Oktaviana  
NIM : A11.2011.06042  
Program Studi : Teknik Informatika-S1  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Penerapan Data Mining Klasifikasi Pola Nasabah  
Menggunakan Algoritma C4.5 pada Bank BRI Batang

Tugas Akhir ini telah diujikan dan dipertahankan dihadapan dewan penguji pada sidang tugas akhir tanggal 22 Februari 2016. Menurut pandangan kami, tugas akhir ini memadai dari segi kualitas maupun kuantitas untuk tujuan penganugrahan gelar Sarjana Komputer (S. Kom)

Semarang, 22 Februari 2016

Dewan Penguji,

Noor Ageng Setiyanto, M.Kom.

Anggota 1

Khafiizh Hastuti, M.Kom.

Anggota 2

Slamet Sudaryanto N.,ST, M.Kom

Ketua Penguji

## **PERNYATAAN KEASLIAN SKRIPSI**

Sebagai mahasiswa Universitas Dian Nuswantoro, yang bertanda tangan dibawah ini, saya :

Nama : Ayu Rizqi Oktaviana

NIM : A11.2011.06042

Menyatakan bahwa karya ilmiah saya yang berjudul :

### **Penerapan Data Mining Klasifikasi Pola Nasabah Menggunakan Algoritma C4.5 pada Bank BRI Batang**

Merupakan karya asli saya (kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya dan perangkat pendukung seperti laptop dll). Apabila dikemudian hari, karya saya disinyalir bukan karya asli saya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar saya beserta hak dan kewajiban yang melekat pada gelar tersebut. Demikian Surat pernyataan ini saya buat dengan sebenarnya.

Semarang, 22 Februari 2016

Yang membuat pernyataan,

Ayu Rizqi Oktaviana

## **PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai mahasiswa Universitas Dian Nuswantoro, yang bertanda tangan dibawah ini, saya :

Nama : Ayu Rizqi Oktaviana

NIM : A11.2011.06042

Demi mengembangkan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Dian Nuswantoro Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

### **Penerapan Data Mining Klasifikasi Pola Nasabah Menggunakan Algoritma C4.5 pada Bank BRI Batang**

Beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini, Universitas Dian Nuswantoro berhak untuk menyimpan data, mengcopy ulang, mempergunakan dan mengelola dalam bentuk *database*, serta mendistribusikan dan menampilkan/mempublikasikan karya ilmiah ini di internet atau media lain untuk kepentingan akademis tanpa izin dari saya selama tetap mencantumkan saya sebagai penulis dan pemilik hak cipta. Segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta karya ilmiah ini menjadi tanggungjawab saya pribadi.

Semarang, 22 Februari 2016

Yang membuat pernyataan,

Ayu Rizqi Oktaviana

## **UCAPAN TERIMA KASIH**

Dalam kesempatan yang baik ini, perkenankan penulis menyampaikan ucapan terima kasih kepada beliau-beliau yang berperan penting dalam keseharian penulis selama ini sehingga laporan tugas akhir dengan judul “PENERAPAN DATAMINING KLASIFIKASI POLA NASABAH MENGGUNAKAN ALGORITMA C4.5 PADA BANK BRI BATANG” dapat penulis selesaikan sesuai dengan rencana. Oleh karena itu, penulis menyampaikan terima kasih kepada :

1. Allah SWT atas limpahan rahmat serta karunia-Nya.
2. Bapak Dr. Ir. Edi Noersasongko, M.Kom, selaku Rektor Universitas Dian Nuswantoro Semarang.
3. Bapak Dr. Abdul Syukur, Drs, MM, selaku Dekan Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang.
4. Bapak Dr. Heru Agus Santoso, Ph.D, selaku Ka.Progdi Teknik Informatika Universitas Dian Nuswantoro Semarang.
5. Bapak Heribertus Himawan, M.Kom, selaku pembimbing tugas akhir yang memberikan bimbingan yang berkaitan dengan penelitian penulis.
6. Doses-dosen pengampu di Fakultas Ilmu Komputer Teknik Informatika Universitas Dian Nuswantoro Semarang yang telah memberikan ilmu dan pengetahuannya masing-masing.
7. Pimpinan dan Staf Bank BRI Batang yang telah membantu dan mengizinkan penulis melakukan wawancara dan pengambilan data untuk proses penulisan laporan Tugas Akhir.
8. Kedua orangtua dan keluarga tercinta yang telah memberikan dorongan, nasehat, kasih sayang, doa, dukungan material, dan spiritual yang tak terhingga kepada penulis dalam pembuatan laporan Tugas Akhir ini.
9. Teman-teman seperjuangan Teknik Informatika Universitas Dian Nuswantoro Semarang yang selalu membantu dan memberikan semangat dalam setia kesempatan.

10. Semua pihak yang namanya tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa mungkin masih terdapat kekurangan dalam Laporan Tugas Akhir ini. Oleh karena itu, kritik dan saran dari pembaca sangat bermanfaat bagi penulis. Semoga laporan ini dapat bermanfaat bagi semua pihak yang membacanya.

Semarang, 22 Februari 2016

Penulis

## ABSTRAK

Data Mining merupakan proses analisa data dari sudut yang berbeda dan mengolahnya menjadi informasi-informasi penting yang bisa digunakan untuk meningkatkan keuntungan. Secara teknis, data mining dapat disebut juga sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field*. Pada data mining dapat digunakan metode *decision tree* untuk melakukan klasifikasi. Kredit tidak lagi menjadi hal yang asing bagi masyarakat luas, banyaknya kebutuhan membuat masyarakat mengambil kredit untuk memenuhi kebutuhan konsumtif mereka. Seringkali pembayaran kredit yang macet membuat bank kerepotan, karena semakin banyaknya nasabah yang membayar secara macet dapat berdampak buruk untuk kesehatan bank. Oleh karena itu, data nasabah dari Bank BRI Batang menjadi bahan acuan untuk menganalisa pola nasabah pemohon kredit. Pemohon kredit termasuk dalam kategori lancar, atau macet. Untuk mengatasi masalah tersebut, dapat digunakan metode *decision tree*. Sehingga penelitian ini bisa dijadikan acuan pihak Bank untuk menilai nasabah dengan record data yang ada untuk pengambilan kredit selanjutnya. Informasi yang ditampilkan berupa tingkat akurasi data nasabah lancar dan macet. Dalam penelitian ini, *decision tree* menggunakan bahasa pemrograman java. Kemudian hasil akurasi dari aplikasi yang diimplementasikan akan dibandingkan dengan hasil menggunakan software rapidminer. Sehingga diperoleh akurasi dengan *decision tree* sebesar 89,5%.

Kata Kunci : *data mining*, klasifikasi, algoritma C4.5, kredit, *decision tree*



## ABSTRACT

*Data mining* is the process of analyzing data from different angles and process them into important information that can be used to increase profits. Technically, data mining can also be called as a process of finding correlation or patterns of hundreds or thousand of field. In data mining decision tree method can be used for the classification. Credit is no longer a strange thing for the general public, many need to make people take credit to meet their consumer needs. Often jammed credit payments made bank hassles, because more and more customers who pay jams can be bad for the health of banks. Thereofer, customer data from Bank BRI become a reference for analyzing pattern of customer credit applicants. Credit applicants included in the current category, or jammed. To overcome these problems, can use decision tree method. So this research could be used as a reference for Bank to assess clients with exsisting data record for further credit decision. The information is displayed in the form of the accuracy of customer data smoothly and jammed. In this research, a decision tree using the java programming language. Then the accuracy of the results of application that is implemented will be compared with results using software Rapidminer. This obtained with less accuracy of decision tree by 89,5%.

Keyword : *data mining, classification, algorithmC4.5, credit, decision tree*

## DAFTAR ISI

SKRIPSI.....	i
PERSETUJUAN SKRIPSI .....	ii
PENGESAHAN DEWAN PENGUJI.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI .....	v
UCAPAN TERIMA KASIH.....	vi
ABSTRAK .....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR .....	xiv
BAB I .....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah .....	4
1.3    Batasan Masalah.....	4
1.4    Tujuan Penelitian.....	5
1.5    Manfaat Penelitian.....	5
1.5.1    Manfaat Bagi Akademik .....	5
1.5.2    Manfaat Bagi Penulis .....	5
1.5.3    Manfaat Bagi Pembaca .....	6
1.5.4    Manfaat Bagi Bank BRI.....	6
BAB II.....	7
2.1    Penelitian Terkait .....	7

2.2	Landasan Teori .....	12
2.2.1	Kredit.....	12
2.2.2	Syarat Pengambilan Kredit Bank BRI Batang .....	15
2.2.3	Kriteria Keterlambatan Kredit Bank BRI .....	15
2.2.4	Data Mining .....	16
2.2.5	Decision Tree .....	23
2.2.6	Algoritma C4.5.....	24
2.2.7	Klasifikasi .....	26
2.2.8	Java.....	28
2.2.9	MySQL.....	30
2.2.10	Kerangka Pemikiran.....	31
BAB III	.....	32
3.1	Instrumen Penelitian .....	32
3.1.1	Bahan.....	32
3.1.2	Peralatan .....	32
3.2	Metode Pengumpulan Data .....	33
3.3	Teknik Analisa Data .....	34
3.4	Desain Sistem .....	36
3.5	Pemodelan .....	38
3.6	Validasi dan Evaluasi .....	38
3.7	Analisis Sistem .....	39
3.7.1	Use Case Diagram.....	40
3.8	Pengujian Sistem .....	43
BAB IV	.....	46
4.1	Pengolahan Data.....	46

4.2	Pengujian Data .....	47
4.3	Pemodelan Datamining Menggunakan Algoritma C4.5 .....	48
4.4	Hasil Pengujian Sistem.....	58
4.5	Tampilan Program .....	59
4.6	Hasil Pengujian.....	63
BAB V.....		65
KESIMPULAN DAN SARAN.....		65
5.1	Kesimpulan.....	65
5.2	Saran .....	65
DAFTAR PUSTAKA .....		lxvi
LAMPIRAN.....		lxviii

## DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian.....	9
Tabel 3.1 Penentuan Atribut Nasabah Kredit .....	34
Tabel 3.2 Skenario Use Case Login .....	41
Tabel 3.3 Skenario Use Case Pengolahan Data .....	42
Tabel 3.4 Skenario Use Case Penentu Pola Nasabah.....	42
Tabel 3.5 Skenario Use Case Pengolahan Data .....	43
Tabel 3.6 Rencana Pengujian Black-box Sistem .....	44
Tabel 4.1 Keterangan Atribut Data Nasabah Kredit .....	46
Tabel 4.2 Skenario Detail Atribut Data Nasabah Kredit.....	47
Tabel 4.3 Data Testing .....	48
Tabel 4.4 Sample Jumlah Kasus Tiap Atribut .....	49
Tabel 4.5 Hasil Perhitungan Nilai Entropy dan Gain .....	56
Tabel 4.6 Hasil Uji Black-box .....	58

## DAFTAR GAMBAR

Gambar 2.1 Proses Data Mining .....	22
Gambar 2.2 Konsep Data Mining .....	23
Gambar 2.3 Kerangka Pemikiran .....	31
Gambar 3.1 Alur yang Diusulkan .....	34
Gambar 3.2 Analisa Data .....	35
Gambar 3.3 Desain Sistem Login .....	36
Gambar 3.4 Desain Halaman Utama.....	36
Gambar 3.5 Desain Input Data.....	37
Gambar 3.6 Desain Cetak Data Nasabah .....	37
Gambar 3.7 Use Case Diagram .....	40
Gambar 4.1 Pohon Keputusan.....	57
Gambar 4.2 Tampilan Halaman Login Sistem.....	60
Gambar 4.3 Tampilan Halaman Utama .....	60
Gambar 4.4 Tampilan Input Data Nasabah.....	61
Gambar 4.5 Tampilan Proses Print Data.....	62
Gambar 4.6 Tampilan Cetak Proses Data .....	62

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Tujuan negara yang dirumuskan dalam tujuan pembangunan nasional sebagaimana telah digariskan pada Tap MPR RI No. 4/MPR/1999 tentang GBHN tahun 1999-2004 yang pada hakekatnya adalah untuk mewujudkan masyarakat yang adil dan makmur berdasarkan Pancasila dan Undang-Undang Dasar 1945. Upaya mewujudkan tujuan negara diatas, salah satunya adalah dengan melaksanakan pembangunan.[1]

Pembangunan nasional yang dilaksanakan mencakup upaya peningkatan disegala bidang kehidupan yaitu pembangunan dibidang ekonomi, politik, sosial, budaya, dan hankam. Pelaksanaan pembangunan membutuhkan dana yang cukup besar, dana tersebut berasal dari APBN dan bantuan luar negeri. Selain itu juga dari sekitar lembaga keuangan. Menurut kepemilikannya, lembaga keuangan terdiri dari lembaga keuangan milik pemerintah (BUMN) dan lembaga keuangan milik swasta, sedangkan jika dilihat dari jenisnya, sektor lembaga keuangan ini terdiri dari lembaga keuangan bank dan lembaga keuangan bukan bank.[1]

Bank merupakan salah satu lembaga keuangan formal. Bank adalah lembaga keuangan yang biasanya didirikan dengan kewenangan yang pada umumnya digunakan untuk menyimpan uang dan meminjamkan uang kepada masyarakat [2]. Sedangkan menurut undang-undang perbankan, bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak.[3]

Kredit bukan merupakan sesuatu yang asing lagi bagi masyarakat luas, masyarakat desa saja mengenal kredit. Banyak masyarakat mengambil kredit untuk memenuhi kebutuhan mereka, seperti, untuk membangun rumah, kebutuhan sehari-hari, modal, biaya anak sekolah, dan lain-lain. Kredit adalah kepercayaan.

Menurut Eric, L. Kohler, “Kredit adalah kemampuan untuk melaksanakan suatu pembelian atau mengadakan suatu pinjaman dengan suatu janji pembayarannya akan dilakukan dan ditanggihkan pada suatu jangka waktu yang disepakati”[4]. Dalam pemberian kredit, tidak hanya memiliki unsur kepercayaan penuh terhadap penerima kredit, tetapi juga terjaganya kepercayaan akan kejujuran dan kemampuan penerima kredit dalam mengembalikan pinjaman tepat pada waktunya. Dengan kata lain seseorang atau perusahaan yang akan menentukan kredit harus mempunyai kredibilitas, atau kelayakan seseorang untuk menerima kredit. Kredit memiliki banyak macam, seperti Kartu Kredit, Kredit Pegawai, Kredit Umum, Kredit Kendaraan Bermotor (KKB), Kredit Usaha Rakyat (KUR), dan Kredit Keperilikan Rumah (KPR).

PT. Bank Rakyat Indonesia (Persero) adalah salah satu bank yang menyediakan layanan Kredit Pegawai kepada para nasabahnya. Pengajuan kredit Pegawai (BRIGuna) bisa dilakukan oleh pemohon yang telah memenuhi syarat dari pihak bank. Pemberian kredit ini masih menjadi masalah yang krusial, karena pihak bank masih sangat sulit untuk menentukan pola debitur atau pemohon yang menyebabkan kredit lancar atau macet. Banyak faktor yang menyebabkan kredit macet, diantaranya jumlah penghasilan tiap bulan dan banyaknya kebutuhan ekonomi. Pembayaran kredit yang macet membuat performa bank menjadi buruk.

Pada Bank BRI Cabang Batang kredit yang banyak diambil adalah kredit pegawai, dimana hanya yang berpenghasilan tetap yang diperbolehkan mengambil kredit tersebut. Syarat untuk mengajukan kredit pada Bank BRI juga sangatlah mudah, tetapi walaupun demikian, masih ada juga nasabah yang membayar kredit secara macet. Apalagi jika nasabah tersebut tiba-tiba menghilang, tidak terlihat lagi dalam instansinya. Sedangkan untuk nasabah yang meninggal sebelum kreditnya lunas, biaya itu ditanggung oleh asuransi.

Pembayaran kredit nasabah yang macet ini dapat berdampak negatif untuk kesehatan Bank, karena apabila nilai *NPL (Non Performing Loan)* pada Bank ini mencapai 2% dapat dikatakan Bank tidak memiliki performa yang bagus atau bahkan bangkrut, karena seharusnya nilai *NPL* pada Bank adalah dibawah 2%. Untuk menekan nilai *NPL*, pihak Bank harus selektif terhadap nasabah kredit



yang membayar secara macet pada kredit. Sehingga penelitian ini bisa dijadikan acuan pihak Bank untuk menilai nasabah dengan record data yang ada untuk pengambilan kredit selanjutnya.

Oleh karena itu, data nasabah dari Bank Rakyat Indonesia (Persero) Kantor Cabang Batang menjadi bahan acuan untuk menganalisa pola nasabah pemohon kredit. Pemohon kredit termasuk dalam kategori lancar, diperlukan perhatian khusus, kurang lancar, atau macet. Dalam kasus ini digunakan teknik data mining *decision tree* untuk mengklasifikasi pola nasabah dengan menggunakan pohon keputusan algoritma C4.5.

Nama data mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis. Data mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, dan belum melewati masa “remaja”, maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya. Maka, Daryl Pregibon menyatakan bahwa “data mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data” yang masih berkembang. Kemunculan data mining berdasarkan pada jumlah data yang tersimpan pada basis data semakin besar. Dalam perusahaan, ada beberapa juta data tersimpan dari setiap kegiatan produksi untuk setiap produk yang dibuat dalam beberapa tahun.[5]

Decision Tree atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Pohon yang dibentuk tidak selalu berupa pohon biner. Apabila semua fitur data set menggunakan dua macam nilai kategorikal maka bentuk pohon yang didapatkan adalah pohon biner. Apabila dalam fitur berisi lebih dari dua macam nilai kategorikal atau menggunakan tipe numerik maka bentuk yang didapatkan biasanya tidak berupa pohon biner. Kefleksibelan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa visualisasi saran (dalam bentuk *decision tree*) yang membuat prosedur prediksinya dapat diamati (Gorunescu, 2011). Decision tree banyak digunakan untuk menyelesaikan kasus penentu

keputusan seperti di bidang kedokteran (diagnosis penyakit pasien), ilmu komputer (struktur data), psikologi (teori pengambilan keputusan) dan sebagainya.

Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3. Dalam ID3, pada decision tree hanya bisa digunakan untuk fitur bertipe kategorikal saja (nominal atau ordinal), sedangkan untuk tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan decision tree, dan penurunan rule set. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi.[5]

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang masalah sebelumnya, maka akan dirumuskan permasalahan bagaimana cara mengklasifikasikan data nasabah kredit BRIGuna pada Bank Rakyat Indonesia (Persero) dengan menggunakan algoritma C4.5.

## **1.3 Batasan Masalah**

Adapun yang menjadi batasan masalah agar tidak menyimpang dari maksud dan tujuan yang sebenarnya, maka penulis membatasi masalah pada penulisan ini, batasan masalahnya sebagai berikut :

- a) Penelitian ini dilakukan untuk mengklasifikasi pola nasabah Kredit BRIGuna pada PT Bank Rakyat Indonesia (Persero) menggunakan algoritma C4.5.
- b) Penerapan datamining menggunakan aplikasi netbean dan mysql.
- c) Data yang digunakan adalah data nasabah Kredit BRIGuna pada Bank BRI Kantor Cabang Batang.

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah yang diuraikan diatas, maka tujuan dari penelitian ini adalah untuk menerapkan data mining pada pengklasifikasian data nasabah kredit pegawai (BRIGuna) pada Bank Rakyat Indonesia (Persero) dengan algoritma C4.5 agar mengetahui pola nasabah pada pembayaran kredit.

#### **1.5 Manfaat Penelitian**

Manfaat yang diperoleh dengan tercapainya tujuan penelitian ini diantaranya yaitu:

##### **1.5.1 Manfaat Bagi Akademik**

Manfaat Penelitian ini bagi Akademik diantaranya yaitu:

1. Menambah literatur di Perpustakaan Universitas Dian Nuswantoro yang dapat digunakan sebagai gambaran dan rujukan dalam membuat laporan tugas akhir.
2. Dapat dijadikan sebagai tolak ukur sejauh mana pemahaman dan penguasaan materi kuliah yang diberikan dalam masa perkuliahan di Universitas Dian Nuswantoro Semarang.

##### **1.5.2 Manfaat Bagi Penulis**

Manfaat Penelitian ini bagi penulis diantaranya yaitu:

1. Sebagai sarana untuk menerapkan ilmu yang telah didapatkan selama perkuliahan.
2. Menambah pemahaman dan pengalaman terutama dalam bidang kajian data mining khususnya algoritma C4.5

### 1.5.3 Manfaat Bagi Pembaca

Dapat digunakan sebagai informasi dan tambahan pengetahuan tentang perancangan sistem untuk klasifikasi pola nasabah dengan menggunakan algoritma C4.5.

### 1.5.4 Manfaat Bagi Bank BRI

Dengan adanya penerapan *datamining* untuk menentukan pola nasabah pada Bank BRI Batang ini dapat membantu menyediakan pengetahuan dan informasi yang bisa dijadikan acuan pihak Bank untuk menilai nasabah dengan record data yang ada untuk pengambilan kredit selanjutnya.

## **BAB II**

### **TINJAUAN PUSTAKA DAN LANDASAN TEORI**

#### **2.1 Penelitian Terkait**

Sebelum memulai penelitian ini, penulis terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Dari studi kepustakaan itu penulis menemukan beberapa penelitian yang mendorong penulis untuk mengangkat tema seperti diatas. Penelitian tersebut yaitu :

1. “Penerapan Algoritma C4.5 pada Program Klasifikasi Mahasiswa Dropout” yang ditulis oleh Anik Andriani [6]. Penelitian ini menggunakan algoritma C4.5 untuk mengklasifikasikan mahasiswa DO (dropout) dan mahasiswa yang masih aktif sebagai acuan dalam membuat kebijakan dan tindakan untuk mengurangi jumlah mahasiswa DO. Kesimpulannya, hasil evaluasi dan validasi dengan confusion matrix menunjukkan tingkat akurasi pada algoritma C4.5 sebesar 97,5%. Sedangkan hasil evaluasi dan validasi dengan ROC/AUC menunjukkan nilai lebih dari 0,9 sehingga dapat dimasukkan ke dalam excellent classification. Penerapan rule dari algoritma C4.5 yang digunakan dalam klasifikasi mahasiswa potensi dropout terhadap data baru diperoleh hasil evaluasi dan validasi dengan confusion matrix menghasilkan tingkat akurasi sebesar 90%.
2. “Penerapan Data Mining untuk Analisa Pola Perilaku Nasabah dalam Pengkreditan Menggunakan Metode C4.5 Studi Kasus pada KSU Insan Kamil Demak” yang ditulis oleh Rina Dewi Indah Sari dan Yuwono Sindunata [7]. Penelitian ini ditulis untuk membantu menganalisis perilaku nasabah dalam pengkreditan. Dalam suatu sistem analisa penentuan nasabah perlu dilakukan pendeteksian kriteria-kriteria nasabah yang digunakan untuk menentukan layak atau tidaknya nasabah dalam pengambilan kredit. Hasil yang didapatkan dari uji coba sebanyak 20

3. data uji dihasilkan 13 data benar dan 7 data salah, sehingga dapat disimpulkan tingkat keakurasian sebesar 65%.
4. “Faktor-Faktor yang Mempengaruhi Penyaluran Kredit Bank Umum di Indonesia” yang ditulis oleh Greydi Normala Sari [8]. Data yang digunakan adalah data sekunder dari Bank umum di Indonesia yang meliputi Dana Pihak Ketiga (DPK), *Capital Adequancy Ratio* (CAR), *Non performing Loan* (NPL), kredit, dan BI Rate pada periode Januari 2008-Februari 2012 (bulanan) data time series. Tujuan penelitian ini adalah untuk mengetahui pengaruh Dana Pihak Ketiga (DPK), *Capital Adequancy Ratio* (CAR), *Non Performing Loan* (NPL), dan BI Rate terhadap penyaluran kredit bank umum di Indonesia. DPK berpengaruh positif terhadap penyaluran kredit. Hal ini ditunjukkan oleh koefisien regresi DPK yaitu sebesar (1.154040). Artinya setiap kenaikan DPK sebesar 1% maka penyaluran kredit akan naik sebesar 1.15%, ceteris paribus. CAR berpengaruh negatif terhadap penyaluran kredit. Hal ini ditunjukkan oleh koefisien regresi CAR yaitu sebesar (-0.013287), artinya setiap kenaikan CAR sebesar 1% maka penyaluran kredit akan turun sebesar 0.013%, ceteris paribus. NPL berpengaruh negatif terhadap penyaluran kredit. Hal ini ditunjukkan oleh koefisien regresi NPL yaitu sebesar (-0.021909), artinya setiap kenaikan NPL sebesar 1% maka penyaluran kredit akan turun sebesar 0.021%, ceteris paribus. BI Rate berpengaruh positif terhadap penyaluran kredit. Hal ini ditunjukkan oleh koefisien regresi BI Rate yaitu sebesar (0.016437), artinya setiap kenaikan BI Rate sebesar 1% maka penyaluran kredit akan naik sebesar 0.016%, ceteris paribus. Kesimpulan dari penelitian ini adalah bahwa secara simultan maupun secara parsial variabel DPK, CAR, NPL, BI Rate berpengaruh terhadap penyaluran kredit Bank Umum di Indonesia.
5. “Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa” yang ditulis oleh David Hartanto Kamagi dan Seng Hansun [9]. Penelitian ini bertujuan untuk mengetahui tingkat

kelulusan mahasiswa pada Universitas Multimedia Nusantara, siapa saja mahasiswa yang lulus tepat waktu, lulus lebih cepat, lulus terlambat atau dropout dengan menggunakan algoritma C4.5. Peneliti mengambil data training dari data alumni mahasiswa program studi Teknik Informatika Universitas Multimedia Nusantara angkatan tahun 2007 dan 2008, sedangkan untuk data testing menggunakan data alumni 2009. Dengan demikian dapat disimpulkan bahwa data mining dengan algoritma C4.5 dapat diterapkan untuk memprediksi tingkat kelulusan mahasiswa dengan empat kategori yaitu lulus cepat, lulus tepat waktu, lulus terlambat, dan dropout. Aplikasi desktop berhasil memprediksi kelulusan mahasiswa dengan presentase 87,5% dari enam puluh data training dan empat puluh data testing. Hasil prediksi kelulusan dari aplikasi ini dapat membantu bagian program studi untuk mengetahui status kelulusan mahasiswa.

Berikut adalah tabel ringkasan penelitian terkait :

Tabel 2.1 Tabel Ringkasan Penelitian

No	Nama Peneliti	Tahun	Judul	Metode	Hasil
1.	Anik Andriani	2012	Penerapan Algoritma C4.5 pada Program Klasifikasi Mahasiswa Dropout	Klasifikasi Mahasiswa Dropout menggunakan algoritma C4.5	hasil evaluasi dan validasi dengan confusion matrix menunjukkan tingkat akurasi pada algoritma C4.5 sebesar 97,5%. Sedangkan hasil evaluasi dan validasi dengan ROC/AUC menunjukkan

					nilai lebih dari 0,9. Hasil evaluasi dan validasi dengan confusion matrix menghasilkan tingkat akurasi sebesar 90%.
2.	Rina Dewi Indah Sari dan Yuwono Sindunata	2014	Penerapan Data Mining untuk Analisa Pola Perilaku Nasabah dalam Pengkreditan Menggunakan Metode C4.5 Studi Kasus pada KSU Insan Kamil Demak	Penerapan Data Mining Menggunakan Metode C.45	Hasil yang didapatkan dari uji coba sebanyak 20 data uji dihasilkan 13 data benar dan 7 data salah, sehingga dapat disimpulkan tingkat keakurasian sebesar 65%.
3.	Greydi Normala Sari	2013	Faktor-Faktor yang Mempengaruhi Penyaluran Kredit Bank Umum di Indonesia”	Menggunakan metode analisis ekonometrika. Model regresi berganda dengan metode kuadrat terkecil OLS (Ordinary Least Square).	Kesimpulan dari penelitian ini adalah bahwa secara simultan maupun secara parsial variabel DPK, CAR, NPL, BI Rate berpengaruh terhadap



					penyaluran kredit Bank Umum di Indonesia.
4.	David Hartanto Kamagi dan Seng Hansun	2014	Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa	Implementasi Data Mining dengan Algoritma C4.5	Aplikasi desktop berhasil memprediksi kelulusan mahasiswa dengan presentase 87,5% dari enam puluh data training dan empat puluh data testing
5.	Ibnu Fatchur Rohman [12]	2015	Penerapan Algoritma C4.5 pada Kepuasan Pelanggan Perum DAMRI	Penerapan Algoritma C4.5	Diketahui nilai precision 86% dengan perbandingan data training dan data testing 40% : 60% adalah 86%, <i>recall</i> 96%, dan F-Measure 90%. Perbandingan 60% : 40% memiliki nilai precision 87%, <i>recall</i> 1%, dan F-Measure

					93%. Sedangkan perbandingan 80% : 20% memiliki nilai precision 1%, <i>recall</i> 89%, dan F- Measure 95%.
--	--	--	--	--	---

## 2.2 Landasan Teori

Dalam Landasan Teori peneliti meninjau dari beberapa sumber buku dan jurnal untuk menjelaskan berbagai hal yang berhubungan dengan topik yang terkait :

### 2.2.1 Kredit

Dalam kehidupan sehari-hari, kredit bukanlah merupakan perkataan yang asing lagi bagi masyarakat luas. Kata-kata kredit tidak hanya dikenal di masyarakat kota besar, bahkan di masyarakat desa pun kata kredit sangat populer.

Istilah kredit berasal dari bahasa Yunani (*credere*) yang berarti kepercayaan (*truth* atau *faith*). Oleh karena itu dasar dari kredit adalah kepercayaan. Seseorang atau suatu badan yang memberikan kredit harus percaya bahwa penerima kredit di masa mendatang akan sanggup memenuhi segala sesuatu yang telah dijanjikan. Adapun yang telah dijanjikan dapat berupa barang, uang, atau jasa. [10]

Menurut Undang-undang Nomor 14 Tahun 1997 tentang Pokok-pokok Perbankan, yang dimaksud dengan kredit adalah “Penyediaan uang atau tagihan-tagihan yang dapat disamakan dengan itu berdasarkan persetujuan pinjam-meminjam antara bank dengan pihak lain dalam hal

mana pihak peminjam berkewajiban melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga yang telah ditetapkan”.

Kredit diberikan oleh seseorang atau suatu lembaga kredit yang didasarkan atas kepercayaan. Ini berarti bahwa seseorang atau suatu lembaga kredit tersebut baru akan memberikan kredit jika ia sudah benar-benar yakin bahwa penerima kredit akan mengembalikannya. Dengan demikian dapat disimpulkan bahwa unsur yang terdapat dalam kredit adalah :

- a. Kepercayaan, yaitu bahwa pemberi kredit harus yakin penuh terhadap penerima kredit bahwa mereka mampu mengembalikan prestasi yang diberikan, baik berupa uang, barang, atau jasa pada jangka waktu yang telah ditentukan.
- b. Waktu, yaitu suatu masa yang memisahkan antara pemberian prestasi dengan kontraprestasi yang akan diterima pada masa yang akan datang.
- c. *Degree of Risk*, yaitu suatu tingkat risiko yang akan dihadapi sebagai akibat dari adanya jangka waktu yang memisahkan antara prestasi dan kontraprestasi yang akan diterima kemudian hari. Semakin lama kredit diberikan, semakin tinggi pula tingkat risikonya. Dengan adanya unsur risiko inilah maka timbullah jaminan dalam pemberian kredit.
- d. Prestasi, atau objek kredit itu sebenarnya tidak hanya berupa uang, bisa berupa barang atau jasa, akan tetapi dalam kehidupan modern seperti sekarang ini, maka uanglah yang sering digunakan dalam pengkreditan.

Dalam kehidupan perekonomian yang moodern seperti sekarang ini, bank memegang peranan penting. Oleh karena itu, organisasi-organisasi bank selalu diikutsertakan dalam menentukan kebijakan di bidang moneter,

pengawasan devisa, pencatatan efek-efek, dan lain-lain. Fungsi kredit perbankan dalam kehidupan perekonomian dan perdagangan antara lain :

- a. Kredit pada hakikatnya dapat meningkatkan daya guna uang  
Biasanya pemilik uang atau modal bersedia meminjamkan uang mereka kepada orang lain atau pengusaha yang membutuhkan modal lebih untuk keperluan peningkatan produksi atau usahanya.
- b. Kredit dapat meningkatkan peredaran dan lalu-lintas uang  
Kredit yang disalurkan melalui rekening giro dapat menciptakan pembayaran baru, sehingga apabila pembayaran-pembayaran dilakukan dengan cek atau giro maka akan meningkatkan peredaran uang giral. Selain itu, kredit perbankan yang ditarik secara tunai dapat pula meningkatkan peredaran uang kartal, sehingga arus lalu-lintas uang akan berkembang pula.
- c. Kredit dapat pula meningkatkan daya guna dan peredaran uang  
Dengan mendapatkan kredit, para pengusaha dapat mengolah bahan baku yang tersedia menjadi barang yang siap dipasarkan, sehingga daya guna barang tersebut meningkat.
- d. Kredit sebagai salah satu alat stabilitas ekonomi  
Dalam keadaan ekonomi yang kurang sehat, kebijakan diarahkan pada usaha-usaha, antara lain :
  1. Pengendalian inflasi,
  2. Peningkatan ekspor, dan
  3. Pemenuhan kebutuhan pokok rakyat.
- e. Kredit dapat meningkatkan kegairahan berusaha  
Ada kalanya, orang yang mempunyai usaha terlahang oleh jumlah modal yang dimiliki, padahal mereka ingin meningkatkan usahanya.
- f. Kredit dapat meningkatkan pemerataan pendapatan  
Setiap pengusaha ingin mengembangkan usaha atau proyek yang dimilikinya, selalu ingin membuka cabang baru untuk memperluas

usaha dan proyeknya, sehingga dapat pula menciptakan lapangan kerja baru bagi masyarakat luas.

- g. Kredit sebagai alat untuk meningkatkan hubungan internasional Perusahaan-perusahaan dalam negeri biasanya secara langsung maupun tidak langsung mendapatkan bantuan dari Bank besar di luar negeri karena mereka memiliki jaringan usaha yang luas.

### 2.2.2 Syarat Pengambilan Kredit Bank BRI Batang

Adapun syarat yang harus dipenuhi calon nasabah kredit pada Bank BRI Batang adalah, sebagai berikut :

1. SK CPNS + PNS
2. SK Terakhir
3. KARPEG (Kartu Pegawai) + Kartu Taspen / ASABRI
4. Rekomendasi Atasan
5. Daftar Rincian Gaji Bulan Terakhir

### 2.2.3 Kriteria Keterlambatan Kredit Bank BRI

Setiap Bank memiliki kriteria keterlambatan pembayaran kredit yang berbeda-beda, berikut adalah kriteria keterlambatan kredit pada Bank BRI :

1. Dalam Pengawasan Khusus

Dalam kasus ini terbagi menjadi 3 kategori, yaitu dalam pengawasan khusus 1, dalam pengawasan khusus 2, dan dalam pengawasan khusus 3.

- a. Dalam Pengawasan Khusus 1

Nasabah terlambat membayar angsurannya selama 1 bulan.

- b. Dalam Pengawasan Khusus 2

Nasabah terlambat membayar angsurannya selama 2 bulan.

- c. Dalam Pengawasan Khusus 3

Nasabah terlambat membayar angsurannya selama 3 bulan.

2. Kurang Lancar

Dalam kasus ini dikatakan kurang lancar apabila nasabah tidak membayar angsurannya selama 4 bulan.

3. Diragukan

Dalam kasus ini dikatakan diragukan apabila nasabah tidak membayar angsuran selama 5-6 bulan.

4. Macet

Dalam kasus ini, nasabah dikatakan macet apabila sampai bulan ke 7 belum membayar angsurannya.

#### 2.2.4 Data Mining

Data Mining adalah proses analisa data dari sudut yang berbeda dan mengolahnya menjadi informasi-informasi penting yang bisa digunakan untuk meningkatkan keuntungan, meminimalisasi biaya pengeluaran, atau bahkan keduanya. Secara teknis, data mining dapat disebut juga sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional basis data yang besar.

Kemampuan data mining dalam mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk [12] :

1. Prediksi dan sifat-sifat bisnis, dimana data mining secara otomatis melakukan proses pencarian informasi untuk memprediksi basis data dalam jumlah besar.
2. Penemuan pola-pola yang tidak diketahui sebelumnya, dimana data mining “menyapu” basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan.

Nama data mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis. Data mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, dan belum melewati masa “remaja”, maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya [5].

#### 2.2.4.1 Pekerjaan dalam Data Mining

Pekerjaan yang berkaitan dengan data mining dapat dibagi menjadi empat kelompok, yaitu: model prediksi, analisis cluster, analisis sosiasi, dan deteksi anomali.

##### a. Model Prediksi

Pekerjaan ini berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat.

##### b. Analisis Cluster

Pekerjaan ini melakukan pengelompokkan data ke dalam kelompok yang ada berdasarkan karakteristik yang sama dari setiap kelompok yang ada.

##### c. Analisis Asosiasi

Analisis asosiasi digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola

yang ditemukan biasanya mempresentasikan bentuk aturan implikasi atau subset fitur.

d. Deteksi Anomali

Pekerjaan deteksi anomali berkaitan dengan penamatan beberapa data dari sisa sejumlah data yang memiliki klasifikasi berbeda.

#### 2.2.4.2 Proses Data Mining

Secara sistematis, ada tiga langkah utama dalam data mining (Gonunescu, 2011):

a. Eksplorasi / pemrosesan awal data

Eksplorasi atau pemrosesan awal data terdiri dari ‘pembersihan’ data, normalisasi data, transformasi data, penanganan data yang salah, reduksi dimensi, pemilihan subset fitur, dan sebagainya.

b. Membangun model dan melakukan validasi terhadapnya

Membangun model dan melakukan validasi terhadapnya berarti melakukan analisis berbagai model dan memilih model dengan kinerja prediksi yang terbaik. Dalam langkah ini digunakan metode-metode seperti kalsifikasi, regresi, analisis cluster, deteksi anomali, analisis asosiasi, analisis pola sekuensial, dan sebagainya.

c. Penerapan

Penerapan berarti menentukan model pada data agar menghasilkan prediksi masalah yang diinvestigasi.

#### 2.2.4.3 Perkembangan Datamining

CRISP-DM (*Cross-Industry Standart Process for Data Mining*) merupakan suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah



ditetapkan sebagai proses standar dalam data mining yang dapat diaplikasikan di berbagai sektor [15].

Adapun siklus hidup pengembangan data mining di antaranya yaitu [15] :

1. *Business Understanding*

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pengetahuan tersebut ke dalam pendefinisian masalah dalam data mining. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

2. *Data Understanding*

Tahap ini akan dimulai dengan pengumpulan data yang kemudian dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah tentang data atau untuk mendeteksi adanya bagian menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.

3. *Data Preparation*

Pada tahap ini meliputi semua kegiatan untuk membangun *dataset* akhir (data yang akan diproses pada tahap pemodelan) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, *record*, dan atribut – atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (*modeling*).

4. *Modeling*

Tahap ini merupakan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal.

Secara khusus ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus, sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya.

#### 5. *Evaluation*

Pada tahap ini model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Kemudian akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal. Kunci dari tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Di akhir dari tahap ini harus ditentukan penggunaan hasil proses data mining.

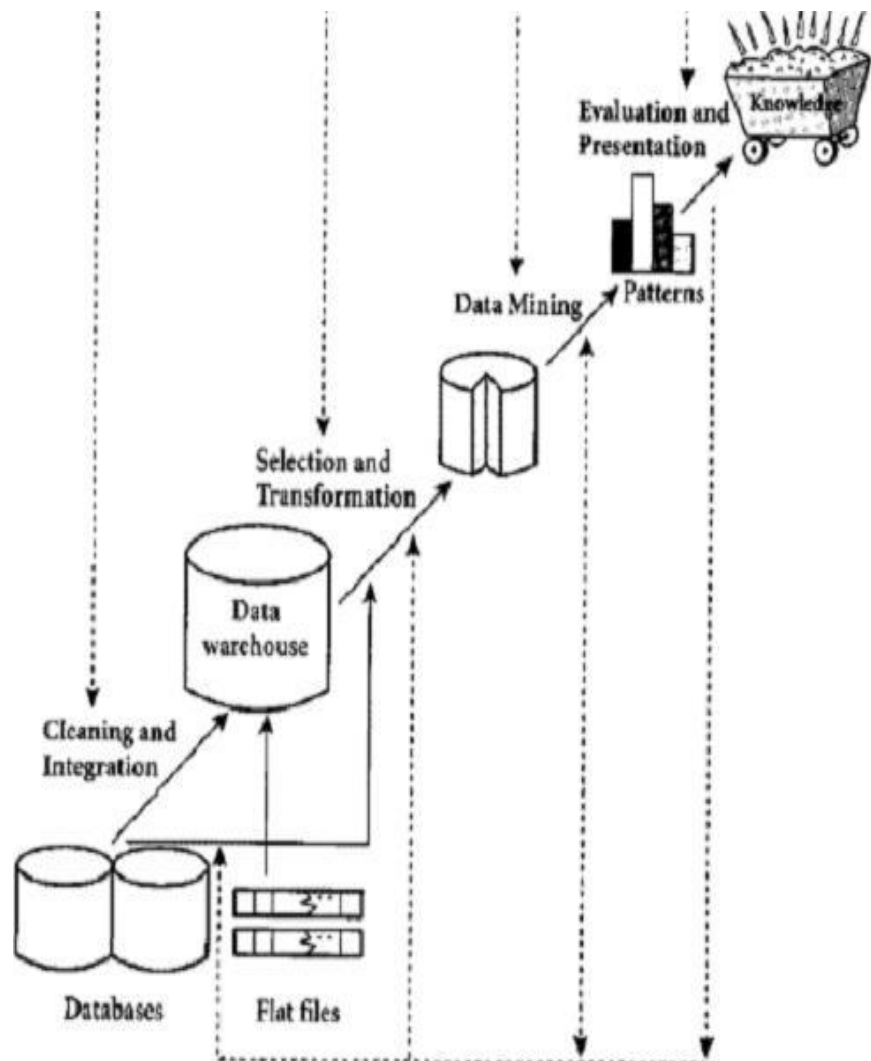
#### 6. *Deployment*

Tahap ini pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap *deployment* dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang dalam perusahaan. Dalam banyak kasus tahap *deployment* melibatkan konsumen, disamping analisis data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.

#### 2.2.4.4 Tahapan- Tahapan Data Mining

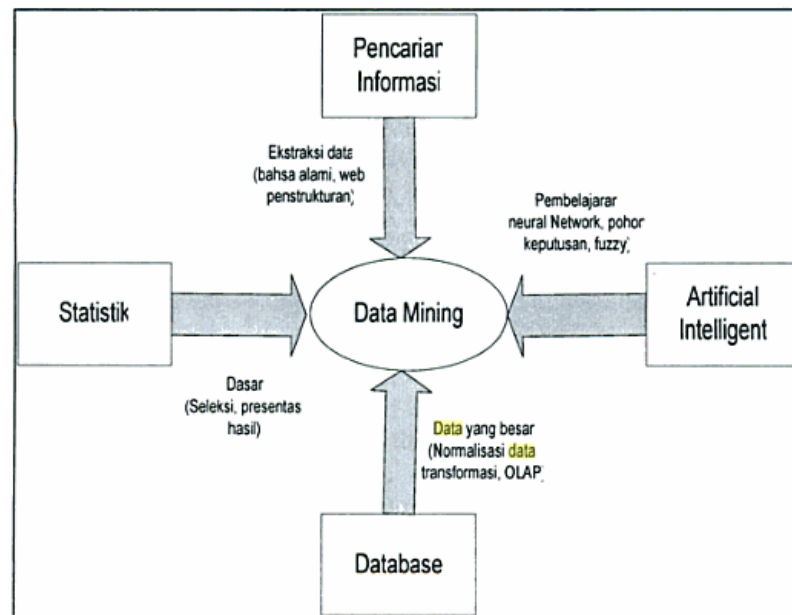
Berikut ini merupakan proses atau langkah-langkah dalam melakukan data mining:

- 1). *Data Cleaning*, merupakan tahap pembersihan data yang tidak konsisten.
- 2). *Data Integration*, merupakan langkah menggabungkan data dari beberapa sumber.
- 3). *Data Selection*, merupakan data yang tidak dikembalikan lagi ke *database* setelah proses *data cleaning*.
- 4). *Data Transformation*, merupakan data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi regresi.
- 5). *Data Mining*, merupakan proses yang digunakan untuk mengolah suatu data dengan menggunakan metode.
- 6). *Evaluation and presentation*, merupakan pengidentifikasian pola berdasarkan tindakan yang digunakan.
- 7). *Knowledge*, yaitu sebuah hasil yang dicapai berupa pengetahuan atau sebuah informasi.



Gambar 2.1 Proses data mining

#### 2.2.4.5 Konsep Data Mining



Gambar 2.2 Konsep data mining

#### 2.2.5 Decision Tree

*Decision tree* atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Pohon yang dibentuk tidak selalu berupa pohon biner. Jika semua fitur dalam data set menggunakan 2 macam nilai kategorikal maka bentuk pohon yang didapatkan berupa pohon biner. Jika dalam fitur berisi lebih dari 2 macam nilai kategorikal atau menggunakan tipe numerik maka bentuk pohon yang didapatkan biasanya tidak berupa pohon biner.

*Decision tree* merupakan metode klasifikasi yang paling sering digunakan. Karena dalam pengerjaannya tidak memerlukan waktu yang lama dan hasilnya pun mudah untuk dipahami dan banyak penelitian dalam kasus ini sering menggunakan decision tree untuk mendapatkan hasil yang maksimal.

Pada decision tree terdapat tiga jenis node, yaitu :

- a. Root Node, merupakan node paling atas, pada node ini tidak ada input dan bisa tidak mempunyai output lebih dari satu.

- b. Internal Node, merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal 2.
- c. Leaf node atau terminal node, merupakan node terakhir, pada node ini hanya terdapat satu input dan tidak mempunyai output.

Kelebihan yang dimiliki decision tree antara lain:

- a. Hasil analisa berupa diagram pohon yang mudah dimengerti.
- b. Mudah untuk dibangun, serta membutuhkan data percobaan yang lebih sedikit dibandingkan algoritma klasifikasi lainnya.
- c. Mampu mengolah data nominal dan kontinyu.
- d. Menggunakan teknik statistik sehingga dapat divalidasi.
- e. Akurasi yang dihasilkan mampu menandingi teknik klasifikasi lainnya.

#### 2.2.6 Algoritma C4.5

Algoritma C4.5 adalah suatu deretan algoritma untuk permasalahan klasifikasi didalam sebuah mesin dan himpunan data. Dengan nilai data yang bervariasi, dimana kejadian diuraikan oleh koleksi atribut dan mempunyai salah satu dari satu set kelas yang eksklusif.

Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi pembenahan dari ID3. Dalam ID3, induksi decision tree hanya dapat digunakan pada fitur bertipe kategorikal (nominal), sedangkan tipe numerik (ratio) tidak bisa digunakan. Perubahan yang membedakan algoritma C4.5 dan ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (*pruning*) decision tree, dan penurunan (*deriving*) rule set. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi.

Algoritma C4.5 adalah satu algoritma untuk klasifikasi penggolongan data. Data yang digunakan adalah data pinjaman kredit kepemilikan rumah pada nasabah yang akan dikelompokkan ke dalam class lancar atau macet.

Terdapat beberapa tahapan untuk menentukan apakah seorang nasabah bermasalah atau tidak menggunakan C4.5.

Hal terpenting dalam induksi decision tree adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting syarat pengujian node :

1. Fitur biner

Fitur yang hanya memiliki 2 nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang.

2. Fitur bertipe kategorikal

Untuk fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda.

3. Fitur bertipe numerik

Untuk fitur bertipe numerik, syarat pengujian dalam node (akar maupun internal) dinyatakan dengan perbandingan pengujian ( $A < v$ ) atau ( $A \geq v$ ) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk  $v_i \leq A < v_{i+1}$  , untuk  $i=1,2,...,k$ .

Dalam algoritma C4.5 sebelum membangun pohon keputusan hal yang paling perlu dilakukan adalah menentukan atribut sebagai akar. Kemudian dibuat cabang untuk setiap nilai dalam akar tersebut. Langkah selanjutnya adalah membagi kasus dalam cabang. Kemudian mengulangi proses pada setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus berikut :

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2.1)$$

Keterangan:

- S : Himpunan kasus  
 A : Atribut  
 N : Jumlah partisi atribut A  
 $|S_i|$  : Jumlah kasus pada partisi ke-i  
 $|S|$  : Jumlah kasus dalam S

Sehingga akan diperoleh nilai gain dari atribut yang paling tertinggi. Gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada tree. Atribut dengan *information gain* tertinggi dipilih sebagai *test attribute* dari suatu *node*. Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2.2)$$

Keterangan:

- S : Himpunan kasus  
 A : Atribut  
 N : Jumlah partisi S  
 $P_i$  : Proporsi dari  $S_i$  terhadap S

### 2.2.7 Klasifikasi

Klasifikasi merupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan tujuan memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan, yaitu:



### 1. Proses Training

Pada proses ini akan digunakan data training set atau data sampel yang telah diketahui label-label atau atribut dari data sampel tersebut untuk membangun model.

### 2. Proses Testing

Pada proses testing ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pada proses training maka dibangun data yang disebut dengan data testing untuk mengklasifikasikan label-labelnya.

Klasifikasi merupakan penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi sekarang ini telah banyak digunakan dalam berbagai aplikasi, sebagai contoh pendeteksian pesan email, spam berdasarkan header dan isi atau mengklasifikasikan galaksi berdasarkan bentuk-bentuknya. Pada proses klasifikasi data yang diinputkan adalah data record atau data sample. Pada setiap record dikenal sebagai instance atau contoh yang ditentukan oleh sebuah tuple  $(x,y)$ . Dimana  $x$  adalah himpunan atribut dan  $y$  adalah atribut tertentu yang menyatakan sebagai label class.

Komponen-komponen utama dari proses klasifikasi antara lain:

- 1). Kelas, merupakan variable tidak bebas yang merupakan label dari hasil klasifikasi. Sebagai contoh adalah kelas loyalitas pelanggan, kelas badai, gempa bumi, dan lain-lain.
- 2). Prediktor, merupakan variable bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasi, misalnya merokok, minuman beralkohol, tekanan darah, status perkawinan, dan sebagainya.
- 3). Set data pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan prediktor untuk dilatih agar dapat mengelompokkan ke dalam kelas yang tepat.
- 4). Set data uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

Sebagian besar istilah-istilah yang ada dalam aktivitas klasifikasi sama dengan yang digunakan dalam aplikasi database. Namun beberapa mungkin tidak begitu dikenal, istilah-istilah tersebut antara lain:

- 1). Set data yang digunakan untuk proses pelatihan dikenal dengan nama-nama yang berbeda, antar lain: records, tuples, vector, instan, objek, dan sample.
- 2). Tiap set data tersebut memiliki suatu atribut.
- 3). Pengklasifikasian, merupakan data-data dengan sifat seperti data pelatihan untuk menguji akurasi dari model yang telah dibuat.

#### 2.2.8 Java

Java adalah bahasa pemrograman dan platform komputasi pertama kali yang dirilis oleh *Sun Microsystem* pada tahun 1995. Java merupakan teknologi yang mendasari kekuatan program untuk *utilitas*, permainan dan aplikasi bisnis. Java berjalan pada lebih dari 850 juta komputer pribadi di seluruh dunia, dan ada miliaran perangkat di seluruh dunia, termasuk ponsel dan perangkat televisi.

Salah satu karakteristik java adapah *portabilitas*, yang artinya bahwa program komputer yang ditulis dalam java harus dijalankan secara sama, pada setiap *hardware / platform* sistem operasi. Hal ini dicapai dengan menyusun kode bahasa java ke sebuah java *bytecode*. Pengguna aplikasi biasanya menggunakan java Runtime Environment (JRE) diinstal pada mesin mereka sendiri untuk menjalankan aplikasi java atau dalam browser web untuk applet java.

Untuk pembuatan dan pengembangan aplikasi berbasis java diperlukan Java Development Kit (JDK), dimana saat ini pemilik lisensi dari JDK adalah Oracle Corporation ang telah resmi mengakuisisi Sun Microsystem pada awal tahun 2010.

#### 2.2.8.1 Java Platform

- a. Java SE (*Standart Edition*), yang khusus digunakan untuk pengembangan aplikasi-aplikasi PC atau workstation.
- b. Java ME (*Micro Edition*), yaitu khusus digunakan untuk pengembangan aplikasi-aplikasi yang ada diperangkat mobile seperti HP, Smartphone, PDA, Tablet dsb.
- c. Java EE (*Enterprise Edition*), yaitu khusus digunakan untuk pengembangan aplikasi skala besar (*enterprise*), dan aplikasi web berbasis java.

Java mempunyai lima prinsip penting yang dijadikan sebagai tujuannya, kelima prinsip ini adalah :

- a. Java harus “sederhana, object oriented dan mudah dimengerti”.
- b. Java harus “kuat dan aman”.
- c. Java harus “netral terhadap arsitektur system (OS, processor) dan bisa jalan di manapun”.
- d. Java harus bisa dijalankan dengan “kinerja yang tinggi”.
- e. Java harus “*interpreted, threaded* dan dinamis”.

Dengan kelima prinsip di atas, aplikasi java mempunyai popularitas yang sangat tinggi terutama di dunia enterprise application. Dimana semua prinsip di atas sangat cocok untuk jenis aplikasi ini. Industri yang mempunyai budget tinggi untuk IT seperti perbankan dan telekomunikasi menggunakan Java secara ekstensif. Banyak aplikasi dengan skala raksasa dibangun menggunakan platform Java.

### 2.2.9 MySQL

*MySQL (My Structured Query Language)* merupakan perangkat lunak sistem basis data, perangkat lunak ini juga sering disebut DBMS (*Database Management System*). Namun berbeda dengan basis data konvensional seperti dbf, dat, dan mdb. *MySQL* sendiri memiliki beberapa kelebihan diantaranya yaitu *multi user*, bersifat *multithread* serta mendukung sistem jaringan [14].

*MySQL* adalah program database yang mampu mengirim dan menerima data dengan sangat cepat dan *multiuser*. *MySQL* memiliki dua bentuk lisensi, yaitu *free software* dan *shareware*. *MySQL* yang *free software* bebas digunakan untuk keperluan pribadi atau usaha tanpa harus membayar atau membeli lisensi, yang berada di bawah lisensi GNU/GPL.

Fitur-fitur dalam *MySQL* adalah sebagai berikut :

a. Multiplatform

*MySQL* tersedia pada beberapa platform seperti Windows, Linux, Unix, dan lain sebagainya.

b. Andal, Cepat, dan Mudah Digunakan

*MySQL* tergolong sebagai server basis data (server yang melayani permintaan terhadap basis data) yang andal, dapat menangani basis data yang besar dengan kecepatan tinggi, mendukung banyak sekali fungsi untuk mengakses basis data, dan mudah digunakan.

c. Dukungan SQL

*MySQL* mendukung perintah SQL (*Structured Query Language*). SQL merupakan standar dalam pengaksesan basis data relasional.

## 2.2.10 Kerangka Pemikiran

Masalah		
Mengklasifikasikan nasabah yang mengambil kredit akan membayar secara lancar atau macet		
Tujuan		
Penerapan dan pengaplikasian dari klasifikasi data nasabah kredit menggunakan algoritma C4.5		
Percobaan		
Tools	Data	Metode
MySql, Netbean, Rapidminer	Kredit Pegawai Bank BRI Batang	Algoritma C4.5
Hasil		
Dapat diketahui tingkat akurasi pembayaran macet dan lancar pada nasabah yang mengambil kredit dan kedepannya pihak bank dapat mempertimbangkan nasabah yang akan mengambil kredit selanjutnya		
Manfaat		
Membantu menyediakan pengetahuan dan informasi yang bisa dijadikan acuan pihak Bank untuk menilai nasabah dengan record data yang ada untuk pengambilan kredit selanjutnya.		

Gambar 2.3 Gambar kerangka Pemikiran

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Instrumen Penelitian**

Penelitian ini dilakukan berdasarkan permasalahan yang telah diuraikan pada bab sebelumnya. Adapun instrumen yang dibutuhkan dalam melakukan penelitian ini yaitu :

Penelitian ini dilakukan berdasarkan permasalahan yang telah diuraikan pada bab sebelumnya. Adapun instrumen yang dibutuhkan dalam melakukan penelitian ini yaitu :

##### **3.1.1 Bahan**

Bahan penelitian ini yang dibutuhkan yaitu berupa data kredit pegawai pada Bank BRI Kantor Cabang Batang, yang akan digunakan untuk menghitung akurasi.

##### **3.1.2 Peralatan**

Tujuan dari deskripsi peralatan adalah untuk mengetahui kebutuhan sistem agar mempermudah perancangan. Peralatan ini meliputi kebutuhan perangkat keras dan perangkat lunak. Kebutuhan-kebutuhan di bawah ini merupakan kebutuhan minimal dari sistem :

##### **A. Kebutuhan Perangkat Lunak**

Perangkat lunak yang dibutuhkan adalah sebagai berikut :

##### **1. Sistem Operasi Windows 7**

Untuk mendukung penelitian ini, maka penulis menggunakan sistem operasi windows 7 dengan versi 64 bit.

##### **2. Tools Netbean dan MySql**

Untuk mengimplementasikan metode yang digunakan, maka penulis menggunakan software Netbean dan MySql.

### 3. Microsoft Office

Untuk menulis laporan ini, penulis menggunakan Microsoft Office Word 2007.

### B. Kebutuhan Perangkat Keras

Selain kebutuhan perangkat lunak, diperlukan pula perangkat keras untuk mendukung penelitian ini. Spesifikasi perangkat keras pada penelitian ini cukup berpengaruh karena penelitian ini membutuhkan prosesor dan memori yang sesuai dengan tools perangkat lunak yang dipakai. Berikut adalah perangkat keras yang digunakan:

1. Laptop dengan prosesor Intel Core i5
2. Komputer PC
3. RAM 2GB
4. Hardisk 500GB
5. Flashdisk 8GB
6. Mouse
7. Printer

## 3.2 Metode Pengumpulan Data

Metode pengumpulan data yang dilakukan penulis untuk mendapatkan data yang valid adalah sebagai berikut :

### 1. Metode Interview atau Wawancara

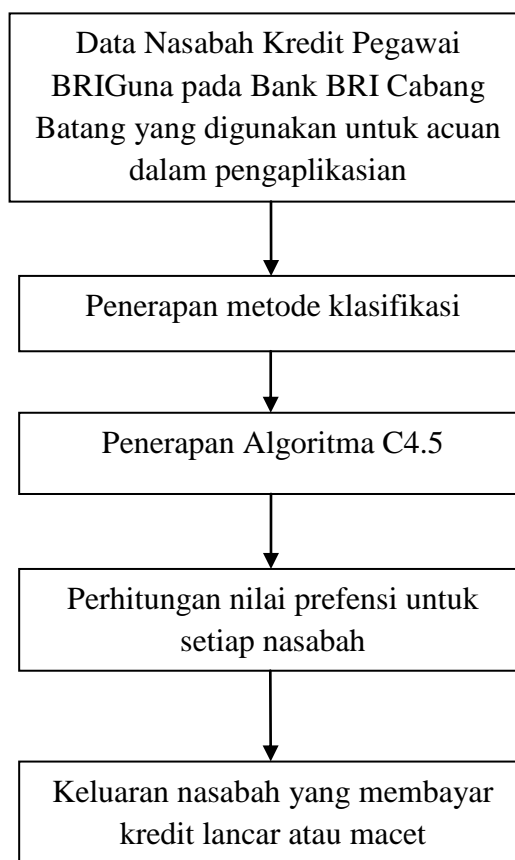
Metode wawancara dilakukan dengan melakukan wawancara terhadap pegawai Bank bagian kredit BRI Batang untuk mendapatkan informasi seputar syarat kredit dan kriteria kredit. Selain itu juga untuk mendapatkan data nasabah kredit untuk dijadikan atribut pendukung pada penelitian ini.

## 2. Metode Studi Literatur

Pada metode studi literatur, penulis mengumpulkan, membaca, mempelajari, dan mencatat literatur dari jurnal maupun buku yang berkaitan dengan algoritma C4.5. Dari metode studi literatur diharapkan dapat memberi gambaran yang dapat diimplementasikan pada aplikasi yang digunakan penulis dalam menyelesaikan penelitian ini.

### 3.3 Teknik Analisa Data

Metode yang diusulkan untuk proses seperti ini menggunakan metode klasifikasi dengan algoritma C4.5. Metode yang diusulkan untuk proses penentuan nasabah kredit adalah seperti gambar dibawah ini :



Gambar 3.1 Alur yang diusulkan

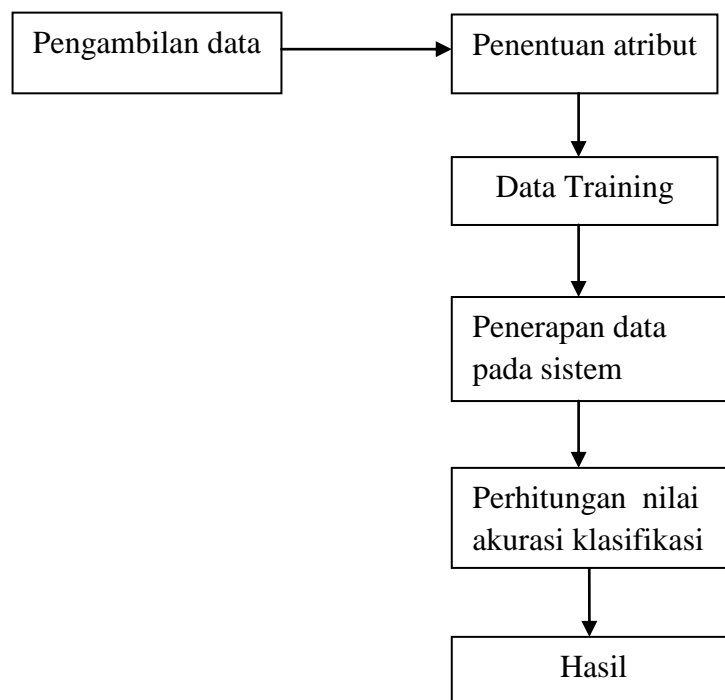


Setelah alur pemodelan ditentukan, maka penerapan algoritma C4.5 berupa perhitungan nilai entropy dan nilai gain dapat dimulai dari atribut yang digunakan dalam penelitian ini, berikut atribut yang digunakan :

Tabel 3.1 Tabel Penentuan Atribut Nasabah Kredit

No	Nama	No. Rek	Umur	Status	Penghasilan	Angsuran	Jangka waktu	hasil
----	------	------------	------	--------	-------------	----------	-----------------	-------

Berikut adalah tahapan analisis data yang digunakan dalam penelitian ini :



Gambar 3.2 Analisis Data

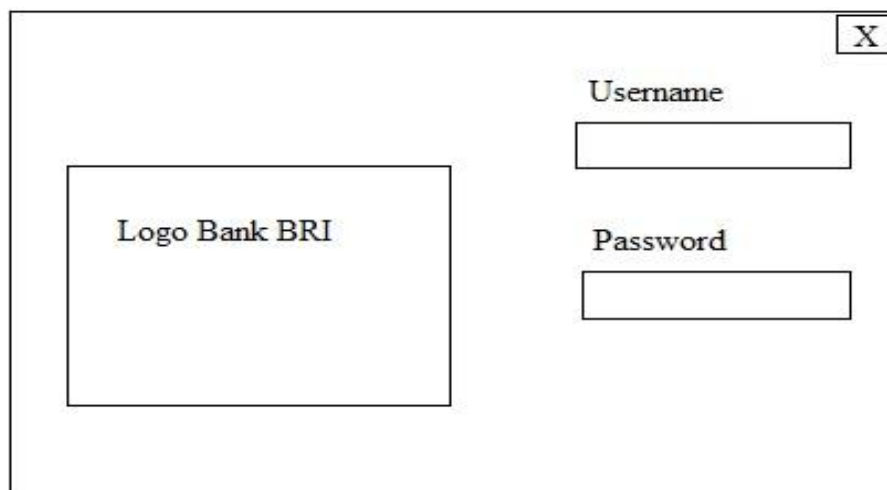
Berdasarkan analisis data diatas :

1. Melakukan pengambilan data dari data nasabah Kredit Pegawai BRIGunapada Bank BRI Cabang Batang.
2. Menentukan atribut yang digunakan dalam penelitian ini untuk dilakukan perhitungan.

3. Data yang digunakan untuk perhitungan berupa sample sebanyak 81 data.
4. Hasil akurasi dari data training dengan hasil dari sistem yang digunakan untuk mengetahui nasabah yang memiliki nilai lancar atau macet.

### 3.4 Desain Sistem

1. Desain yang dibuat untuk halaman login admin atau user yang sudah memiliki username dan password



Gambar 3.3 Gambar desain login

2. Desain halaman utama yang dibuat untuk penerapan datamining



Gambar 3.4 Gambar desain halaman utama

3. Desain untuk input data nasabah, admin atau user memasukkan nama, nomor rekening, usia, status, penghasilan, angsuran, dan jangka waktu

Input Data Nasabah										
Input Data	Cetak Data									
<p><b>Nama</b> <input type="text"/></p> <p><b>No.Rek</b> <input type="text"/></p> <p><b>Usia</b> <input type="text"/></p> <p><b>Status</b> <input type="radio"/> Menik <input type="radio"/> Belum</p> <p><b>Penghasil</b> <input type="text"/></p> <p><b>Angsuran</b> <input type="text"/></p> <p><b>Jangka waktu</b> <input type="text"/></p>	<p><b>Cari Data</b> <input type="text"/></p> <table border="1"> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </table>									

Gambar 3.5 Gambar desain input data

4. Desain cetak data pada sistem penerapan data mining

Input Data Nasabah													
Input Data	Cetak Data												
<table border="1"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table>													<p><input type="text"/></p> <p><b>Output</b></p>

Gambar 3.6 Gambar desain cetak data nasabah

### 3.5 Pemodelan

Pada penelitian ini metode yang akan digunakan adalah Algoritma C4.5 dan dengan menggunakan tools Java Neatbean dan Rapidminer.

Ada beberapa tahap dalam membuat pohon keputusan dengan algoritma C4.5, yaitu sebagai berikut:

1. Menyiapkan data training atau data testing. Data training dan data tessting biasanya diambil dari histori yang pernah terjadi 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 dulu nilai entropy, dengan rumus sebagai berikut:

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i \quad (3.1)$$

3. Kemudian hitung nilai gain, dengan rumus sebagai berikut:

$$\text{Gain(S,A)} = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy (S)} \quad (3.2)$$

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.

### 3.6 Validasi dan Evaluasi

Evaluasi dan validasi hasil dihitung menggunakan rumus akurasi, *precison* dan *recall* berikut ini.

1. Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample *data testing* yang diuji.

$$Akurasi = \frac{\text{jumlah data benar}}{\text{jumlah data testing yang dilakukan}} \quad (3.3)$$

2. Menghitung nilai *precision* dengan cara membagi jumlah data yang benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Positive*).

$$Precision = \frac{TP}{TP+FP} \quad (3.4)$$

3. Sedangkan *recall* dihitung dengan cara membagi data benar yang bernilai positif (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai negative (*False Negative*).

$$Recall = \frac{TP}{TP+FN} \quad (3.5)$$

4. Nilai F-Measure didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall*, kemudian dikalikan dua.

$$F-Measure = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3.6)$$

Keterangan :

TN = True Positif

TN = True Negatif

FP = False Positif

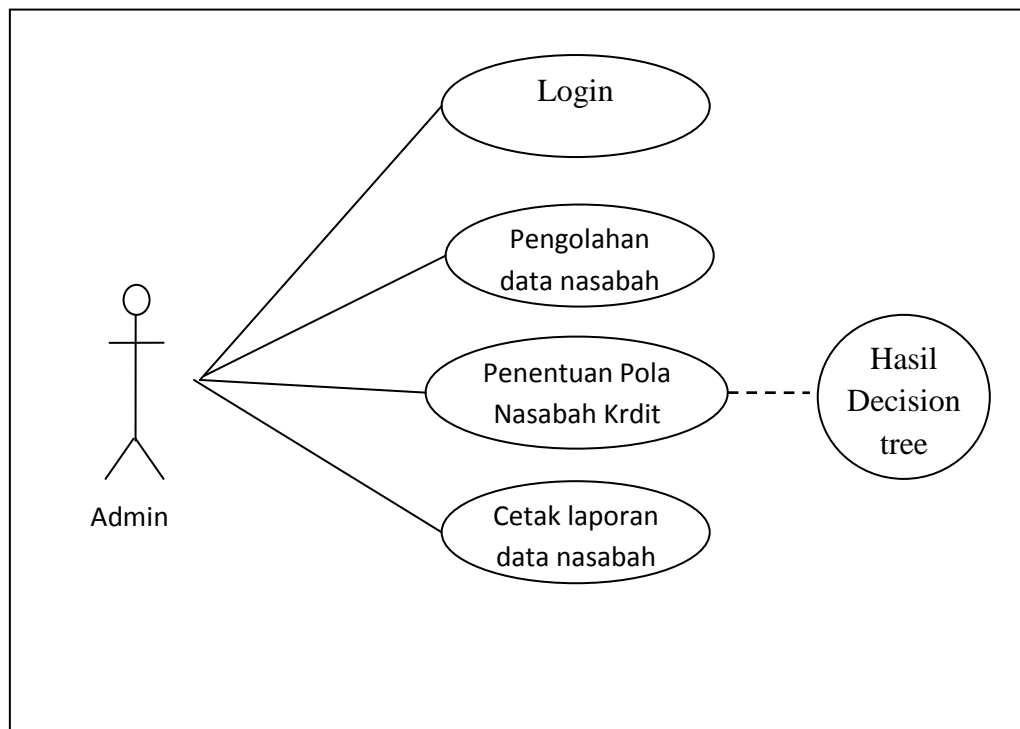
FN = False Negatif

### 3.7 Analisis Sistem

Perancangan sistem yang dibuat untuk klasifikasi calon pendonor darah ini antara lain meliputi *use case diagram*, dan *scenario case*

### 3.7.1 Use Case Diagram

Use case adalah sekelompok yang saling terkait dan membentuk sistem secara teratur yang dilakukan atau diawasi oleh sebuah aktor. Use case digunakan untuk membentuk tingkah laku benda dalam sebuah mode serta direalisasikan oleh sebuah *collaborator*, umumnya use case digambarkan dengan sebuah elips dengan garis yang solid, biasanya mengandung nama. Use case menggambarkan proses system (kebutuhan sistem dari sudut pandang user).



Gambar 3.7 Use Case Diagram

#### 3.7.1.1 Skenario Use Case

##### a. Use Case Login

Nama use case : Login

Aktor : Admin

Tujuan : Menjelaskan inisialisasi admin

Deskripsi:

1. Admin memasukkan username atau password saat sistem penerapan datamining mulai dioperasikan.
2. Sistem tersebut memeriksa kebenaran data yang dimasukkan.

Tabel 3.2 Skenario usecase login

No	Actor	Sistem
1.	Admin memasukkan username atau password saat sistem penerapan datamining mulai dioperasikan	
2.		Sistem tersebut memeriksa kebenaran data yang dimasukkan.

b. Use Case Pengolahan Data Nasabah

Nama use case : Pengolahan data nasabah Bank BRI

Actor : Admin

Tujuan : Menjelaskan proses pengolahan data Nasabah

Deskripsi:

1. Sistem menampilkan form input data nasabah.
2. Admin memasukkan data nasabah sesuai dengan atribut yang dimiliki.
3. Sistem menyimpan data yang dimasukkan.

Tabel 3.3 Skenario Usecase pengolahan data

No	Actor	Sistem
1.		Sistem menampilkan form input data nasabah.
2.	Admin memasukkan data nasabah sesuai dengan atribut yang dimiliki.	
3.		Sistem menyimpan data yang dimasukkan.

## c. Use Case penentu pola nasabah kredit

Nama usecase : Penentu pola nasabah

Aktor : Admin

Tujuan : Menjelaskan proses penentu pola nasabah

Deskripsi:

1. Sistem menampilkan form data penentu pola nasabah kredit.
2. Admin memasukkan data atribut nasabah untuk dilakukan proses perhitungan pola nasabah
3. Sistem memproses data dengan perhitungan algoritma C4.5
4. Sistem menampilkan hasil kategori yang diperoleh sistem.

Tabel 3.4 Skenario use case penentu pola nasabah

No	Actor	Sistem
1.		Sistem menampilkan form data penentu pola nasabah kredit.
2.	Admin memasukkan data atribut nasabah untuk dilakukan proses perhitungan pola nasabah	



3.		Sistem memproses data dengan perhitungan algoritma C4.5
4.		Sistem menampilkan hasil kategori yang diperoleh sistem.

d. Cetak Laporan Data Nasabah

Nama Use Case : Pengolahan laporan data

Tujuan : Memberikan detail dari semua data pola nasabah yang telah dimasukkan.

Deskripsi:

1. Sistem menampilkan laporan data hasil.

Tabel 3.5 Skenario use case pengolahan data

No	Actor	Sistem
1.		Sistem menampilkan laporan data hasil.

### 3.8 Pengujian Sistem

Pengujian yang dilakukan akan menggunakan pengujian *black-box testing*. *Black-box testing* merupakan pengujian suatu perangkat lunak atau sistem dengan menguji secara fungsional berdasarkan pada spesifikasi kebutuhan perangkat lunak. *Black-box testing* melakukan pengujian tanpa mengetahui struktur internal dari sistem atau komponen yang diuji [17]:

Tabel 3.6 Rencana pengujian black-box sistem

No	Skenario Pengujian	Test-case	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
1.	Mengisi semua isian data login dengan sembarangan	Username: bri Password: batang	Sistem akan menolak akses login dan menampilkan username dan password tidak valid	Sesuai/ Tidak Sesuai	Valid/ Tidak Valid
2.	Mengisi semua isian data login dengan benar	Username: riris Password: oktober	Sistem menerima akses login dan kemudian langsung menampilkan halaman utama sistem	Sesuai/ Tidak Sesuai	Valid/ Tidak Valid
3.	Tidak mengisikan salah satu dari form yang tersedia	Nama lengkap: Ayu Rizqi No.Rek: 1234567890 Usia: 23 Penghasilan: 3000000 Angsuran: 800000 Jangka Waktu:-	Saat data di save dan direfresh tidak ada muncul prediksi lancar atau macet.	Sesuai/ Tidak Sesuai	Valid/ Tidak Valid
4	Mengisikan secara lengkap form yang terdapat pada input data nasabah	Nama lengkap: Ayu Rizqi No.Rek: 1234567890 Usia: 23	Ketika data di save dan di refresh, maka akan muncul prediksi	Sesuai/ Tidak Sesuai	Valid/ Tidak Valid

		Penghasilan: 3000000 Angsuran: 800000 Jangka Waktu: 160	lancar atau macet.		
5.	Isikan atau copy kan nomor rekening pada form pencarian untuk proses cetak data	Form Pencarian: 1234567890	Setelah klik proses maka akan muncul cetak data yang telah diisikan form input data sebelumnya dan akan diketahui lancar atau macet.	Sesuai/ Tidak Sesuai	Valid/ Tidak Valid

## **BAB IV**

### **PEMBAHASAN**

#### **4.1 Pengolahan Data**

Pada penelitian ini menggunakan data primer yang didapat dari Bank BRI Kantor Cabang Batang dengan jumlah 81 data nasabah kredit dengan atribut nama, nomor rekening, umur, status, penghasilan, angsuran, jangka waktu, dan label. Dari data tersebut diambil 16 data sample yang digunakan sebagai contoh perhitungan manual entropy dan gain, dimana pada data sample memiliki 7 data nasabah lancar dan 9 data nasabah macet.

Tabel 4.1 Keterangan Atribut Data Nasabah Kredit

Nama	Atribut yang difungsikan sebagai ID
Nomor Rekening	Atribut yang difungsikan sebagai ID
Umur	Atribut yang digunakan sebagai kelayakan pengambilan kredit
Status	Atribut yang menunjukkan status pernikahan dari nasabah
Penghasilan	Atribut yang menjelaskan penghasilan nasabah kredit
Angsuran	Atribut yang menjelaskan pembayaran angsuran dan bunga yang diambil nasabah kredit setiap bulannya
Jangka Waktu	Atribut yang digunakan untuk menentukan lama waktu pembayaran kredit yang diambil yang sudah ada kesepakatan dengan pihak bank
Label	Atribut yang digunakan untuk menjelaskan hasil dari kredit nasabah yang merupakan pembayaran lancar atau macet

Selanjutnya adalah pemilihan atribut data yang digunakan. Atribut yang berubah juga digunakan karena penggunaan atribut yang saling berkaitan.

Tabel 4.2 Tabel detail atribut Data Nasabah Kredit

Atribut	Detail penggunaan	
Nama	x	No
Nomor Rekening	x	No
Umur	√	Nilai Model
Status	√	Nilai Model
Penghasilan	√	Nilai Model
Angsuran	√	No
Jangka Waktu	√	Nilai Model
Label	√	Label Target

Atribut yang digunakan guna mendukung penelitian agar mendapatkan hasil yang maksimal. Pengolahan data menggunakan java netbean dan mysql.

## 4.2 Pengujian Data

Pada penelitian ini digunakan 81 data nasabah untuk menentukan akurasi klasifikasi. Terdapat 16 data nasabah yang digunakan sebagai data sample untuk perhitungan entropy dan gain secara manual. Berikut data testing yang digunakan :

Tabel 4.3 Tabel Data Testing

No	Umur	Status	Penghasilan	Angsuran	Jangka Waktu	Label
1.	dewasa	MENIKAH	rendah	$\leq 700000$	240	Macet
2.	muda	BELUM MENIKAH	tinggi	$> 1500000$	160	Lancar
3.	dewasa	MENIKAH	sedang	$\geq 800000$ - $1500000$	160	Macet
4.	muda	BELUM MENIKAH	rendah	$\leq 700000$	120	Macet
5.	tua	MENIKAH	rendah	$\leq 700000$	180	Macet
6.	muda	BELUM MENIKAH	sedang	$\geq 800000$ - $1500000$	160	Lancar
7.	muda	MENIKAH	tinggi	$> 1500000$	160	Lancar
8.	tua	MENIKAH	rendah	$\leq 700000$	120	Macet
9.	tua	MENIKAH	tinggi	$\geq 800000$ - $1500000$	120	Lancar
10.	muda	BELUM MENIKAH	sedang	$\leq 700000$	160	Lancar
11.	dewasa	MENIKAH	rendah	$\leq 700000$	180	Macet
12.	dewasa	MENIKAH	tinggi	$> 1500000$	240	Lancar
13.	muda	BELUM MENIKAH	sedang	$\geq 800000$ - $1500000$	180	Macet
14.	muda	BELUM MENIKAH	rendah	$\leq 700000$	160	Macet
15.	dewasa	MENIKAH	rendah	$\leq 700000$	240	Macet
16.	tua	MENIKAH	sedang	$\geq 800000$ - $1500000$	240	Lancar

#### 4.3 Pemodelan Datamining Menggunakan Algoritma C4.5

Berikut ini merupakan langkah-langkah perhitungan nilai Entrophy dan Gain pada setiap atribut yang memiliki label Lancar dan Macet :

### 1. Perhitungan Nilai Entrophy

Langkah awal algoritma C4.5 adalah mencari nilai entrophy. Pertama, tentukan terlebih dahulu nilai entrophy total dalam kasus. Rumusnya sebagai berikut :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i \quad (4.1)$$

Tabel 4.4 Tabel sample jumlah kasus tiap atribut

Simpul	Kasus	Kolektibilitas	
		Lancar	Macet
Jumlah Kasus	16	7	9
Jangka Waktu			
240	4	2	2
180	3	0	3
160	6	4	2
120	3	1	2
dst			

Maka :

Entropy Total (S)

$$= \left( -\left( \frac{jml.lancar}{totalkasus} \right) * \log_2 \left( \frac{jml.lancar}{totalkasus} \right) \right) + \left( -\left( \frac{jml.macet}{totalkasus} \right) * \log_2 \left( \frac{jml.macet}{totalkasus} \right) \right)$$

Ettotal (7,9)

$$= \left( -\left( \frac{7}{16} \right) * \log_2 \left( \frac{7}{16} \right) \right) + \left( -\left( \frac{9}{16} \right) * \log_2 \left( \frac{9}{16} \right) \right)$$

$$= \left( -(0,4375) * \left( \frac{-0,3590}{0,3010} \right) \right) + \left( -(0,5625) * \left( \frac{-0,2498}{0,3010} \right) \right)$$

$$\begin{aligned}
&= (-(0,4375) * (-1,1927) + (-(0,5625) * (-0,8299)) \\
&= 0,5218 + 0,4668 \\
&= 0,9886
\end{aligned}$$

Selanjutnya lakukan perhitungan entropy pada tiap atribut berdasarkan pada jumlah kasus per subset atribut, yaitu :

a. Perhitungan nilai subset pada atribut jangka waktu

Entropy 240 (2, 2)

$$\begin{aligned}
&= (-(2/4) * \log_2(2/4) + (-(2/4) * \log_2(2/4)) \\
&= (-(0,5) * (\frac{-0,3010}{0,3010}) + (-(0,5) * (\frac{-0,3010}{0,3010})) \\
&= (-(0,5) * (-1) + (-(0,5) * (-1)) \\
&= 0,5 + 0,5 = 1
\end{aligned}$$

Entropy 160 (4, 2)

$$\begin{aligned}
&= (-(4/6) * \log_2(4/6) + (-(2/6) * \log_2(2/6)) \\
&= (-(0,6667) * (\frac{-0,1760}{0,3010}) + (-(0,3333) * (\frac{-0,4471}{0,3010})) \\
&= (-(0,6667) * (-0,5847) + (-(0,3333) * (-1,5850)) \\
&= 0,3898 + 0,5282 \\
&= 0,918
\end{aligned}$$

Entropy 120 (1, 2)

$$\begin{aligned}
&= (-(1/3) * \log_2(1/3) + (-(2/3) * \log_2(2/3)) \\
&= (-(0,3333) * (\frac{-0,4771}{0,3010}) + (-(0,6667) * (\frac{-0,1760}{0,3010}))
\end{aligned}$$



$$\begin{aligned}
&= (-0,3333) * (-1,5850) + (-0,6667) * (-0,5847) \\
&= 0,5282 + 0,3898 \\
&= 0,918
\end{aligned}$$

b. Perhitungan nilai subset pada atribut status  
Ebelummenikah (3,3)

$$\begin{aligned}
&= (- (3/6) * \log_2 (3/6) + (- (3/6) * \log_2 (3/6) ) \\
&= (- (0,5) * ( \frac{-0,3010}{0,3010} ) + (- (0,5) * ( \frac{-0,3010}{0,3010} ) ) \\
&= (- (0,5) * (-1) + (-0,5) * (-1) ) \\
&= 0,5 + 0,5 \\
&= 1
\end{aligned}$$

Emenikah (4,6)

$$\begin{aligned}
&= (- (4/10) * \log_2 (4/10) + (- (6/10) * \log_2 (6/10) ) \\
&= (- (0,4) * ( \frac{-0,3979}{0,3010} ) + (- (0,6) * ( \frac{-0,2218}{0,3010} ) ) \\
&= (- (0,4) * (-1,3219) + (-0,6) * (-0,7368) ) \\
&= 0,5287 + 0,4420 \\
&= 0,9707
\end{aligned}$$

c. Perhitungan nilai subset pada atribut penghasilan

Esedang (3,2)

$$= (- (3/5) * \log_2 (3/5) + (- (2/5) * \log_2 (2/5) )$$

$$= (- (0,6) * ( \frac{-0,2218}{0,3010} ) + (- (0,4) * ( \frac{-0,3979}{0,3010} ) )$$

$$= (- (0,6) * (- 0,7368 ) + (- (0,4) * (- 1,3219) )$$

$$= 0,5287 + 0,4420$$

$$= 0,9707$$

d. Perhitungan nilai subset pada atribut angsuran

$E \leq 7000000$  (1,7)

$$= (- (1/8) * \log_2 (1/8) + (- (7/8) * \log_2 (7/8) )$$

$$= (- (0,125) * ( \frac{-0,9030}{0,3010} ) + (- (0,875) * ( \frac{-0,0579}{0,3010} ) )$$

$$= (- (0,125) * (- 3 ) + (- (0,875) * (- 0,1923) )$$

$$= 0,375 + 0,1682$$

$$= 0,5432$$

$E > 8000000-15000000$  (3,2)

$$= (- (3/5) * \log_2 (3/5) + (- (2/5) * \log_2 (2/5) )$$

$$= (- (0,6) * ( \frac{-0,2218}{0,3010} ) + (- (0,4) * ( \frac{-0,3979}{0,3010} ) )$$

$$= (- (0,6) * (- 0,7368 ) + (- (0,4) * (- 1,3219) )$$

$$= 0,5287 + 0,4420$$

$$= 0,9707$$

e. Perhitungan nilai subset pada atribut umur

Edewasa (1,4)

$$= (- (1/5) * \log_2 (1/5) + (- (4/5) * \log_2 (4/5) )$$

$$= (- (0,2) * ( \frac{-0,6989}{0,3010} ) + (- (0,8) * ( \frac{-0,0969}{0,3010} ) )$$

$$= (- (0,2) * (- 0,8073) + (- (0,8) * (- 0,3219) )$$

$$= 0,4643 + 0,2575$$

$$= 0,7218$$

Emuda (4,3)

$$= (- (4/7) * \log_2 (4/7) + (- (3/7) * \log_2 (3/7) )$$

$$= (- (0,5714) * ( \frac{-0,2430}{0,3010} ) + (- (0,4285) * ( \frac{-0,3679}{0,3010} ) )$$

$$= (- (0,5714) * (- 0,8073 ) + (- (0,4285) * (- 1,2222) )$$

$$= 0,4612 + 0,5897$$

$$= 1,0059$$

Etua (2,2)

$$= (- (2/4) * \log_2 (2/4) + (- (2/4) * \log_2 (2/4))$$

$$= (- (0,5) * ( \frac{-0,3010}{0,3010} ) + (- (0,5) * ( \frac{-0,3010}{0,3010} ) )$$

$$= (- (0,5) * (-1) + (- (0,5) * (-1))$$

$$= 0,5 + 0,5$$

$$= 1$$

## 2. Perhitungan Nilai Gain

Setelah semua perhitungan entropy pada masing-masing subset atribut selesai, kemudian lakukan perhitungan nilai gain, dengan rumus :

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (4.2)$$

### a. Gain Jangka Waktu

$$\begin{aligned} &= 0,9886 - (4/16 * 1) + (3/16 * 0) + (6/16 * 0,918) + (3/16 * 0,918) \\ &= 0,9886 - (0,25 * 1) + (0,1875 * 0) + (0,375 * 0,918) + (0,1875 * 0,918) \\ &= 0,9886 - (0,25) + (0) + (0,3442) + (0,1721) \\ &= 0,9886 - 0,7663 \\ &= 0,2223 \end{aligned}$$

### b. Gain Status

$$\begin{aligned} &= 0,9886 - (10/16 * 0,9707) + (6/16 * 1) \\ &= 0,9886 - (0,625 * 0,9707) + (0,375 * 1) \\ &= 0,9886 - (0,6066) + (0,375) \\ &= 0,9886 - 0,9816 \\ &= 0,007 \end{aligned}$$

### c. Gain Penghasilan

$$= 0,9886 - (4/16 * 0) + (5/16 * 0,9707) + (7/16 * 0)$$

$$= 0,9886 - (0) + (0,3125 * 0,9707) + (0)$$

$$= 0,9886 - (0) + (0,3033)$$

$$= 0,9886 - 0,3033$$

$$= 0,6853$$

d. Gain Angsuran

$$= 0,9886 - (8/16 * 0,5432) + (5/16 * 0,9707) + (3/16 * 0)$$

$$= 0,9886 - (0,5 * 0,5432) + (0,3125 * 0,9707) + (0)$$

$$= 0,9886 - (0,2716) + (0,3033) + (0)$$

$$= 0,9886 - 0,5749$$

$$= 0,4137$$

e. Gain Umur

$$= 0,9886 - (5/16 * 0,7218) + (7/16 * 1,0059) + (4/16 * 1)$$

$$= 0,9886 - (0,3125 * 0,7218) + (0,4375 * 1,0059) + (0,25 * 1)$$

$$= 0,9886 - (0,2255) + (0,4400) + (0,25)$$

$$= 0,9886 - 0,3492$$

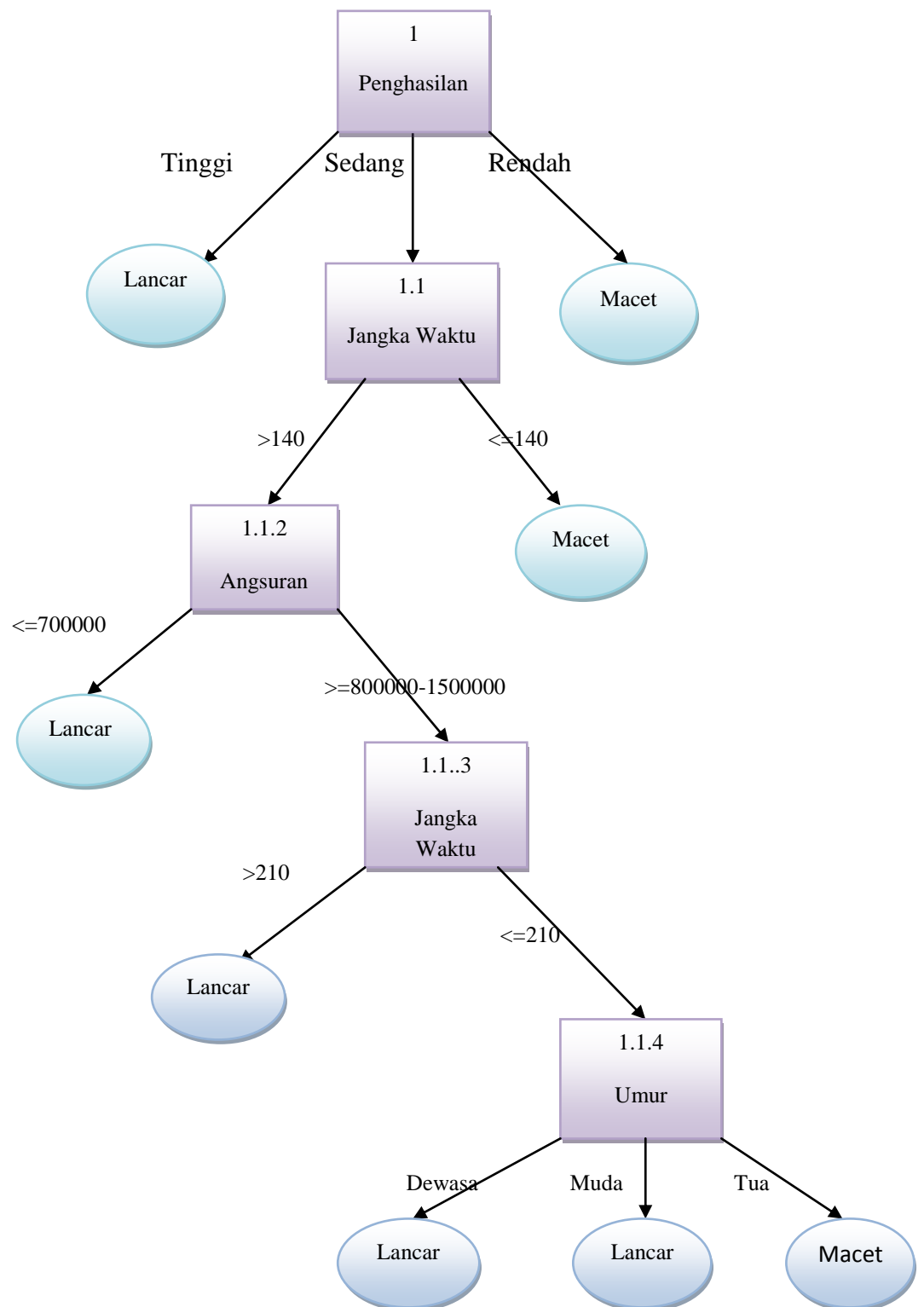
$$= 0,6394$$

Berikut adalah tampilan hasil hitungan entropy dan gain yang selengkapnya dalam bentuk tabel :

Tabel 4.5 Tabel Hasil Perhitungan Nilai Entrophy dan Gain

	Simpul	Jumlah Kasus	Kolektibilitas		Entropy	Gain
			Lancar	Macet		
Total		16	7	9	0,9886	
Jangka Waktu						0,2223
	240	4	2	2	1	
	180	3	0	3	0	
	160	6	4	2	0,918	
	120	3	1	2	0,918	
Status						0,007
	Menikah	10	4	6	0,9707	
	Belum Menikah	6	3	3	1	
Penghasilan						0,6853
	Tinggi	4	4	0	0	
	sedang	5	3	2	0,9707	
	rendah	7	0	7	0	
Angsuran						0,4137
	<=700000	8	1	7	0,5432	
	>=800000 -1500000	5	3	2	0,9707	
	>1500000	3	3	0	0	
Umur						0,6394
	Dewasa	5	1	7	0,7218	
	Muda	7	4	3	1,0059	
	Tua	4	2	2	1	

Nilai gain tertinggi dijadikan node akar pertama. Untuk node akar selanjutnya, dilakukan perhitungan lagi sampai semua tupel terpenuhi.



Gambar 4.1 Pohon Keputusan

#### 4.4 Hasil Pengujian Sistem

Dari langkah-langkah pengujian yang sudah dilakukan sebelumnya, kemudian akan dilakukan pengujian pada sistem penerapan data mining nasabah kredit Bank BRI Batang menggunakan metode *Black-box testing*, sebagai berikut:

Tabel 4.6 Hasil Uji Balck-box pada sistem

No	Skenario Pengujian	Test-case	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
1.	Mengisi semua isian data login dengan sembarangan	Username: bri Password: batang	Sistem akan menolak akses login dan menampilkan username dan password tidak valid	Tidak Sesuai	Tidak Valid
2.	Mengisi semua isian data login dengan benar	Username: riris Password: oktober	Sistem menerima akses login dan kemudian langsung menampilkan halaman utama sistem	Sesuai	Valid
3.	Tidak mengisikan salah satu dari form yang tersedia	Nama lengkap: Ayu Rizqi No.Rek: 1234567890 Usia: 23 Penghasilan: 3000000 Angsuran: 800000 Jangka Waktu:-	Saat data di save dan direfresh tidak ada muncul prediksi lancar atau macet.	Sesuai	Valid



4	Mengisikan secara lengkap form yang terdapat pada input data nasabah	Nama lengkap: Ayu Rizqi No.Rek: 1234567890 Usia: 23 Penghasilan: 3000000 Angsuran: 800000 Jangka Waktu: 160	Ketika data di save dan di refresh, maka akan muncul prediksi lancar atau macet.	Sesuai	Valid
5.	Isikan atau copy kan nomor rekening pada form pencarian untuk proses cetak data	Form Pencarian: 1234567890	Setelah klik proses maka akan muncul cetak data yang telah diisikan form input data sebelumnya dan akan diketahui lancar atau macet.	Sesuai	Valid

#### 4.5 Tampilan Program

Pada penjelasan ini akan diimplementasikan antarmuka program yang dibuat untuk simulasi penerapan datamining pola nasabah Bank BRI Batang, yaitu:

- a. Tampilan awal program ketika dijalankan, user diminta untuk melakukan pengisian username dan password terlebih dahulu. Hanya admin atau bahkan user yang diijinkan mengakses saja.



Gambar 4.2 Gambar halaman Login Sistem

- b. Tampilan halaman utama setelah admin atau user berhasil login, setelah itu pilih input data untuk melakukan proses penerapan datamining.



Gambar 4.3 Gambar Halaman Utama

- c. Tampilan input data nasabah, yang perlu diinputkan oleh user adalah nama, nomor rekening, usia, status, penghasilan, jumlah angsuran, dan jangka waktu. Tampilan ini digunakan untuk proses prediksi.

**Input Data Nasabah**

Input Data   Cetak Data   Data Training

Cari Data Berdasarkan Nomor Rekening

Nama Lengkap

No Rekening

Usia  Tahun

Status ☐ Menikah ☐ Belum Menikah

Penghasilan  IDR

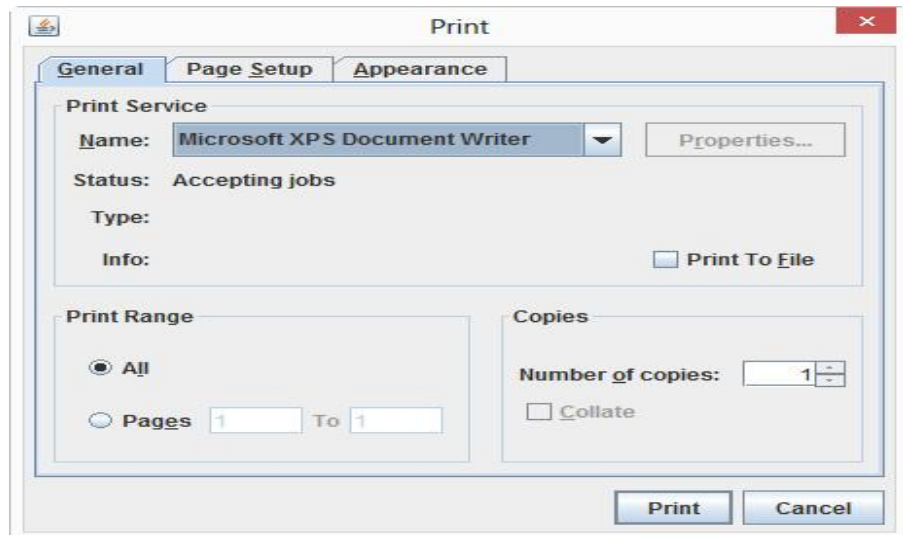
Angsuran  IDR

Jangka Kredit  Bulan

nama	rekening	usia	status	penghasilan	angsuran	waktu	label
miftah	332104090...	22	Belum Meni...	3000000	1000000	180	Macet

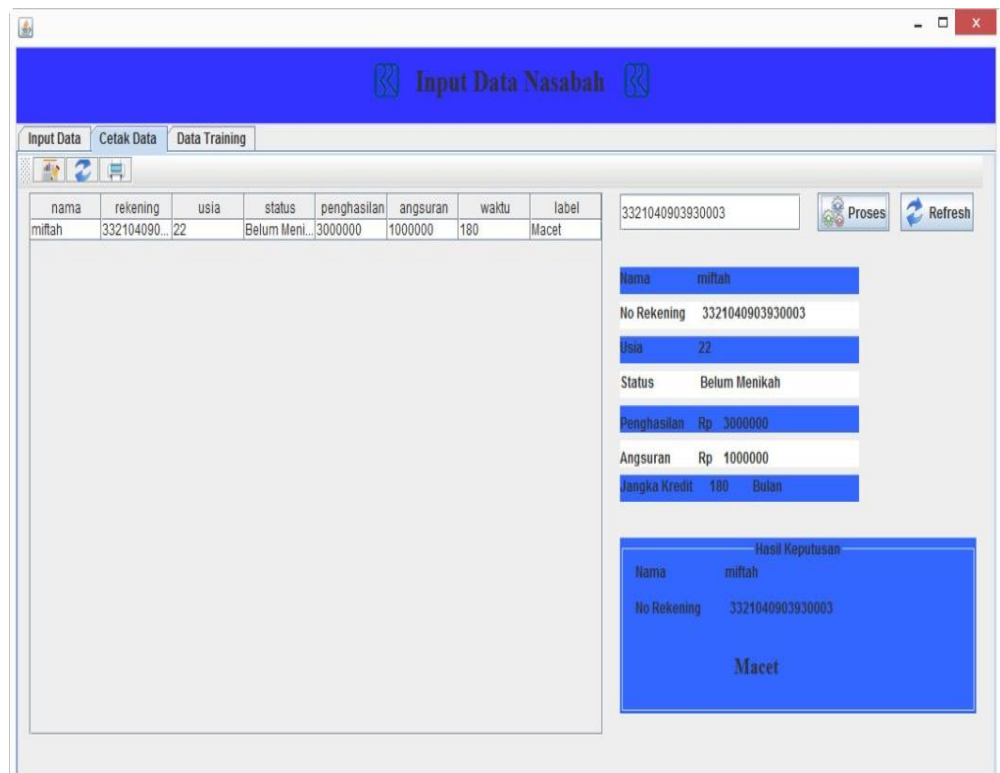
Gambar 4.4 Gambar Tampilan Input Data Nasabah

d. Tampilan antarmuka untuk proses print data.



Gambar 4.5 Gambar proses print data

e. Tampilan cetak data untuk inputan data nasabah dari data prediksi, disana ditampilkan hasil proses data kredit nasabah tersebut macet atau lancar.



Gambar 4.6 Gambar Tampilan Cetak Proses Data

#### 4.6 Hasil Pengujian

- a. Dari hasil pengujian sistem menggunakan bahasa pemrograman java netbean dengan metode decision tree C4.5 dengan jumlah data sebanyak 81 data record, maka menghasilkan 73 data teridentifikasi benar dan 8 data teridentifikasi salah. Untuk menentukan akurasi kinerja dari sistem ini maka nilai hasil pengujian akan dimasukkan ke dalam persamaan berikut:

##### Akurasi

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample data testing yang diuji.

$$= \frac{\text{hasil benar}}{\text{jumlah data}} \times 100\%$$

$$= \frac{73}{81} \times 100\% = 90,1\%$$

##### Precision

Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data yang bernilai positif (*True Positive*) dan data salah yang bernilai (*False Positive*).

$$= \frac{tp}{tp + fp} \times 100\%$$

$$= \frac{34}{34 + 1} \times 100\% = 97,1\%$$

##### Recall

Sedangkan *recall* dihitung dengan cara membagi data benar yang bernilai positif (*tp*) dengan hasil penjumlahan dari data benar yang bernilai positif (*tp*) dan data salah yang bernilai negatif (*fn*).

$$= \frac{tp}{tp+fn} \times 100\%$$

$$= \frac{34}{34+39} \times 100\% = 46,5\%$$

- b. Dari hasil pengujian menggunakan software rapidminer dengan metode decision tree C4.5 dengan jumlah data sebanyak 81 data record, maka menghasilkan 72 data teridentifikasi benar dan 9 data teridentifikasi salah. Berikut adalah hasil akurasi, precision, dan recall:

Akurasi

$$= \frac{72}{81} \times 100\% = 88,89\%$$

Precision

$$= \frac{35}{35+4} \times 100\% = 87,5\%$$

Recall

$$= \frac{35}{35+37} \times 100\% = 48,6\%$$

Dari kedua hasil percobaan diatas, menggunakan software rapidminer dan pemrograman java netben dapat diambil rata-rata sebagai berikut:

$$= \frac{\text{jumlah akurasi rapid} + \text{jumlah akurasi sistem}}{2}$$

$$= \frac{88,89+90,1}{2}\%$$

$$= 89,5 \%$$

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Setelah melakukan analisis, perancangan, dan pengujian maka dapat diperoleh kesimpulan sebagai berikut :

- a. Berdasarkan evaluasi bab sebelumnya dan dari pengujian yang telah dilakukan, penerapan data mining klasifikasi pola nasabah menggunakan algoritma C4.5 pada Bank BRI Batang menghasilkan akurasi yang tinggi, baik menggunakan dengan aplikasi maupun software rapidminer, yaitu 89,5%.
- b. Berdasarkan evaluasi yang telah dilakukan, kriteria macet yang diperoleh adalah yang berpenghasilan rendah atau yang berpenghasilan sedang dan jangka waktu  $\leq 140$ , dan atau yang jangka waktunya  $\leq 240$  dan berumur sudah tua.

#### **5.2 Saran**

Untuk menyempurnakan analisis penerapan datamining yang telah dibuat, maka penulis memberikan saran sebagai berikut :

- a. Dapat digunakan dan dikembangkan atau dibandingkan dengan teknik datamining lainnya.
- b. Karena jumlah data record yang terbatas, nantinya dapat dikembangkan untuk jumlah data yang lebih banyak agar rules yang terbentuk lebih baik lagi dan hasil akurasi yang didapat juga lebih tinggi.
- c. Untuk penelitian selanjutnya sebaiknya uji data sampai 3 kali atau lebih dengan presentase yang berbeda untuk mengetahui perbandingannya.

## DAFTAR PUSTAKA

- [1] Raditya Anindika. 2011. *“Analisis Faktor-Faktor yang Mempengaruhi Besarnya Pengambilan Kredit oleh Masyarakat pada Perum Pegadaian”*. Skripsi. FE, Ekonomi Pembangunan, Universitas Negeri Sebales Maret Surakarta.
- [2] Hoggson, N.F. 1926. *Banking Through the Ages*. New York: Dodd, Mead & Company
- [3] Republik Indonesia. 1998. Undang-Undang Perbankan. Jakarta: Sekretariat Negara.
- [4] Astiko, Sunardi. 1996. *“Pengantar Manajemen Perkreditan”*. Yogyakarta: Andi Yogyakarta.
- [5] Prasetyo, Eko. 2014. *“DataMining Mengolah Data Menjadi Informasi Menggunakan Matlab”*. Yogyakarta: Andi Yogyakarta.
- [6] Andriani, Anik. 2012. *“PenerapanAlgoritma C4.5 pada Program Klasifikasi Mahasiswa Dropout”*. Jakarta: AMIK BSI.
- [7] Indah Sari, Rina Dewi dan Yuwono Sindunata. 2014. *“Penerapan Datamining untuk Analisa Pola Perilaku Nasabah dalam Pengkreditan Menggunakan Metode C4.5 Studi Kasus pada KSU Insan Kamil Demak”*. Jurnal Ilmiah Teknologi dan Informasi Asia. Vol.8 No.2.
- [8] Sari, Greydi Normala. 2013. *“Faktor-Faktor yang Mempengaruhi Penyaluran Kredit Bank Umum di Indonesia (Periode 2008.1-2012.2)”*. Jurnal EMBA. Vo.1 No.3 Hal. 931-941.
- [9] Kamagi, David Hartanto dan Seng Hansun. 2014. *“Implementasi Datamining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa”*. Ultimaticks. Vol. 6 No. 1.



- [10] Suyatno, Thomas *et al.* 1993. “*Dasar-Dasar Perkreditan (edisi ketiga)*”. Jakarta: PT. Gramedia Pustaka Utama.
- [11] Maburur, Angga Ginanjar dan Riani Lubis. 2012. “*Penerapan Datamining untuk Memprediksi Kriteria Nasabah Kredit*”. Jurnal Komputer dan Informatika. Edisi 1 Vol. 1.
- [12] Rohman, Ibnu Fatchur. 2015. “*Penerapan Algoritma C4.5 pada Kepuasan Pelanggan Perum DAMRF*”. Semarang: Universitas Dian Nuswantoro.
- [13] Amri, M. Kroiril. 2013. “*Penerapan Datamining untuk Menentukan Kriteria Calon Nasabah Potensial pada AJB Bumiputera 1912 Palembang*”. Skripsi. FIK, Teknik Informatika, Universitas Bina Darma Palembang.
- [14] Y. L. Prihartanto, “*Sistem Informasi Manajemen Agenda pada Badan Pelayanan Perijinan Terpadu Kabupaten Karanganyar*”. Indonesian Journal on Computer Science Speed (IJCSS). Vol.9, No.3, pp 1-6, 2012.
- [15] D. T. Wahyuni, T.Sutojo dan A. Luthfiarta, “*Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naive Bayes Algoritma Genetika Sebagai Fitur Seleksi,*” pp. 1-14, 2010.
- [16] Puspita, Ari dan Mochamad Wahyudi. 2015. “*Algoritma C4.5 Berbasis Decision Tree untuk Prediksi Kelahiran Bayi Prematur*”. Konferensi Nasional Ilmu Pengetahuan dan Teknologi (KNIT).
- [17] Zulkifli. “*Metode Prediksi Berbasis Neural Network Untuk Pengujian Perangkat Lunak Metode black-box*”. Seminar Nasional ATI, 15 Juni 2013

## LAMPIRAN

No	Umur	Status	Penghasilan	Angsuran	Jangka Waktu	Label
1.	dewasa	MENIKAH	rendah	<=700000	240	Macet
2.	muda	BELUM MENIKAH	tinggi	>1500000	160	Lancar
3	dewasa	MENIKAH	sedang	>=800000-1500000	160	Macet
4.	muda	BELUM MENIKAH	rendah	<=700000	120	Macet
5.	tua	MENIKAH	rendah	<=700000	180	Macet
6.	muda	BELUM MENIKAH	sedang	>=800000-1500000	160	Lancar
7.	muda	MENIKAH	tinggi	>1500000	160	Lancar
8.	tua	MENIKAH	rendah	<=700000	120	Macet
9.	tua	MENIKAH	tinggi	>=800000-1500000	120	Lancar
10.	muda	BELUM MENIKAH	sedang	<=700000	160	Lancar
11.	dewasa	MENIKAH	rendah	<=700000	180	Macet
12.	dewasa	MENIKAH	tinggi	>1500000	240	Lancar
13.	muda	BELUM MENIKAH	sedang	>=800000-1500000	180	Macet
14.	muda	BELUM MENIKAH	rendah	<=700000	160	Macet
15.	dewasa	MENIKAH	rendah	<=700000	240	Macet
16.	tua	MENIKAH	sedang	>=800000-1500000	240	Lancar
17.	dewasa	MENIKAH	tinggi	>1500000	160	Lancar
18.	muda	BELUM MENIKAH	rendah	<=700000	160	Macet
19.	tua	MENIKAH	tinggi	>=800000-1500000	120	Lancar
20.	tua	MENIKAH	sedang	<=700000	180	Lancar
21.	tua	MENIKAH	rendah	<=700000	160	Macet

22.	tua	MENIKAH	tinggi	>1500000	160	Lancar
23.	muda	BELUM MENIKAH	sedang	>=800000- 1500000	120	Macet
24.	dewasa	MENIKAH	rendah	<=700000	120	Macet
25.	tua	MENIKAH	rendah	<=700000	160	Macet
26.	muda	BELUM MENIKAH	sedang	>=800000- 1500000	180	Lancar
27.	dewasa	MENIKAH	tinggi	>1500000	240	Lancar
28.	tua	MENIKAH	rendah	<=700000	180	Macet
29.	muda	BELUM MENIKAH	tinggi	>=800000- 1500000	160	Lancar
30.	muda	BELUM MENIKAH	sedang	<=700000	240	Lancar
31.	tua	MENIKAH	rendah	<=700000	240	Macet
32.	tua	MENIKAH	tinggi	>1500000	160	Lancar
33.	muda	MENIKAH	sedang	>=800000- 1500000	160	Macet
34.	tua	MENIKAH	rendah	<=700000	120	Macet
35.	tua	MENIKAH	rendah	<=700000	180	Macet
36.	muda	BELUM MENIKAH	sedang	>=800000- 1500000	160	Lancar
37.	muda	BELUM MENIKAH	tinggi	>1500000	160	Lancar
38.	dewasa	MENIKAH	rendah	<=700000	120	Macet
39.	tua	MENIKAH	tinggi	>=800000- 1500000	120	Lancar
40.	muda	BELUM MENIKAH	sedang	<=700000	160	Lancar
41.	muda	BELUM MENIKAH	rendah	<=700000	180	Macet
42.	dewasa	MENIKAH	tinggi	>1500000	240	Lancar
43.	tua	MENIKAH	sedang	>=800000- 1500000	180	Macet
44.	muda	BELUM MENIKAH	rendah	<=700000	160	Macet
45.	muda	BELUM MENIKAH	rendah	<=700000	240	Macet
46.	tua	MENIKAH	sedang	>=800000- 1500000	240	Lancar
47.	tua	MENIKAH	tinggi	>1500000	160	Lancar

48.	muda	MENIKAH	rendah	<=700000	160	Macet
49.	tua	MENIKAH	tinggi	>=800000-1500000	120	Lancar
50.	tua	MENIKAH	sedang	<=700000	180	Lancar
51.	muda	BELUM MENIKAH	rendah	<=700000	160	Macet
52.	muda	BELUM MENIKAH	tinggi	>1500000	160	Lancar
53.	dewasa	MENIKAH	sedang	>=800000-1500000	120	Macet
54.	tua	MENIKAH	rendah	<=700000	120	Macet
55.	muda	BELUM MENIKAH	rendah	<=700000	160	Macet
56.	dewasa	MENIKAH	sedang	>=800000-1500000	180	Lancar
57.	tua	MENIKAH	tinggi	>1500000	240	Lancar
58.	muda	BELUM MENIKAH	rendah	<=700000	180	Macet
59.	tua	MENIKAH	tinggi	>=800000-1500000	160	Lancar
60.	tua	MENIKAH	sedang	<=700000	240	Lancar
61.	muda	MENIKAH	rendah	<=700000	240	Macet
62.	tua	MENIKAH	tinggi	>1500000	160	Lancar
63.	tua	MENIKAH	sedang	>=800000-1500000	160	Macet
64.	muda	BELUM MENIKAH	rendah	<=700000	120	Macet
65.	muda	BELUM MENIKAH	rendah	<=700000	180	Macet
66.	dewasa	MENIKAH	sedang	>=800000-1500000	160	Lancar
67.	tua	MENIKAH	tinggi	>1500000	160	Lancar
68.	muda	BELUM MENIKAH	rendah	<=700000	120	Macet
69.	tua	MENIKAH	tinggi	>=800000-1500000	120	Lancar
70.	dewasa	MENIKAH	sedang	<=700000	160	Lancar
71.	tua	MENIKAH	rendah	<=700000	180	Macet
72.	muda	BELUM MENIKAH	tinggi	>1500000	240	Lancar

73.	tua	MENIKAH	sedang	>=800000-1500000	180	Macet
74.	tua	MENIKAH	rendah	<=700000	160	Macet
75.	muda	BELUM MENIKAH	rendah	<=700000	240	Macet
76.	tua	MENIKAH	sedang	>=800000-1500000	240	Lancar
77.	tua	MENIKAH	tinggi	>1500000	160	Lancar
78.	tua	MENIKAH	rendah	<=700000	160	Macet
79.	muda	BELUM MENIKAH	tinggi	>=800000-1500000	120	Lancar
80.	dewasa	MENIKAH	sedang	<=700000	180	Lancar
81.	tua	MENIKAH	rendah	<=700000	160	Macet