

**ANALISIS *CREDIT SCORING* MENGGUNAKAN METODE  
*BAGGING K-NEAREST NEIGHBOR***



**SKRIPSI**

**Disusun oleh :**

**FATIMAH  
24010210120028**

**DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2016**

## HALAMAN PENGESAHAN I

Judul : Analisis *Credit Scoring* Menggunakan Metode *Bagging k-Nearest Neighbor*

Nama : Fatimah

NIM : 24010210120028

Departemen : Statistika

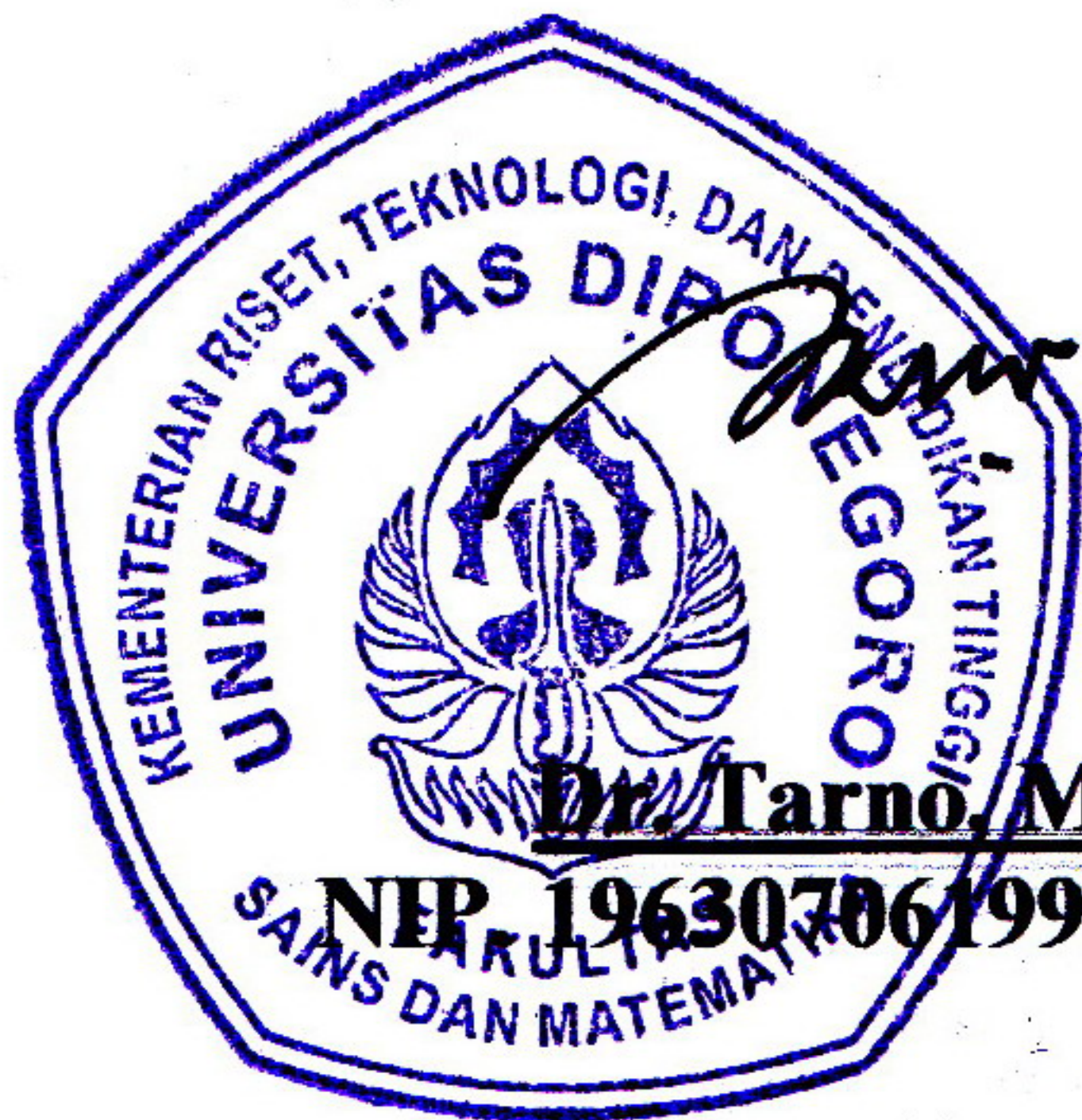
Telah diujikan pada sidang Tugas Akhir dan dinyatakan lulus pada tanggal 3 November 2016.

Semarang, November 2016

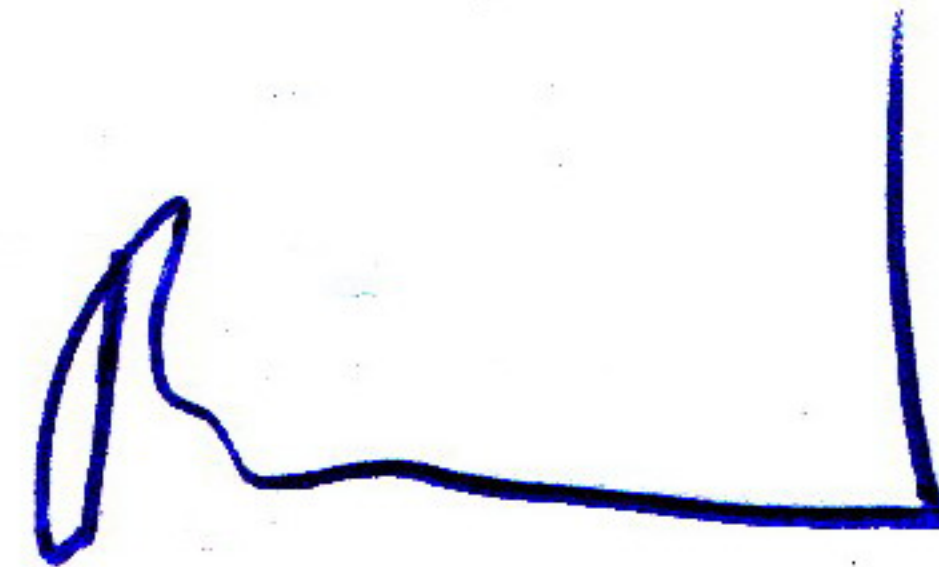
Mengetahui,

Ketua Departemen Statistika  
FSM UNDIP

Ketua Panitia Penguji  
Ujian Tugas Akhir,



Dr. Tarno, M.Si  
NIP. 196307061991021001



Prof. Drs. Mustafid, M.Eng, Ph.D  
NIP. 1955052819800310002

## HALAMAN PENGESAHAN II

**Judul** : *Analisis Credit Scoring Menggunakan Metode Bagging k-Nearest Neighbor*

**Nama** : Fatimah

**NIM** : 24010210120028

**Departemen** : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 3 November 2016

Semarang, 3 November 2016

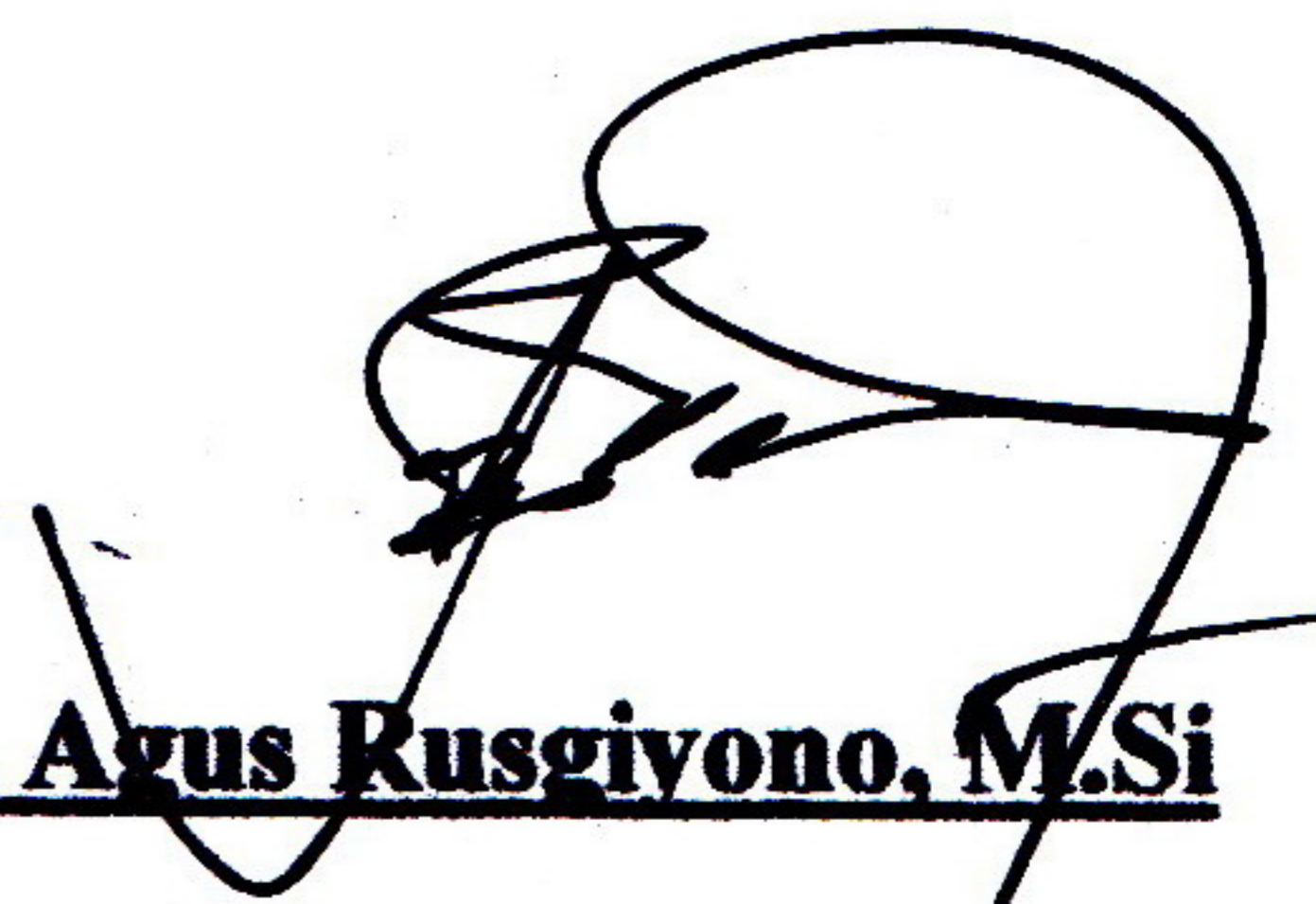
Pembimbing I



**Moch. Abdul Mukid, S.Si, M.Si**

**NIP. 197808172005011001**

Pembimbing II



**Drs. Agus Rusgiyono, M.Si**

**NIP. 196408131990011001**

## KATA PENGANTAR

Puji syukur penulis ucapkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “**Analisis Credit Scoring Menggunakan Metode *Bagging k-Nearest Neighbor***”. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Bapak Dr. Tarno, M.Si. selaku Ketua Departemen Statistika Fakultas Sains dan Matematika.
2. Bapak Moch. Abdul Mukid, S.Si, M.Si. dan Bapak Drs. Agus Rusgiyono, M.Si. sebagai dosen pembimbing yang telah memberikan bimbingan dan pengarahan dalam penulisan tugas akhir ini.
3. Bapak dan Ibu dosen Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro yang telah memberikan ilmu yang bermanfaat.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis menyelesaikan tugas akhir ini.

Penulis berharap tugas akhir ini dapat bermanfaat bagi civitas akademika di Universitas Diponegoro, khususnya Departemen Statistika dan masyarakat pada umumnya.

Semarang, November 2016

Penulis

## ABSTRAK

Menurut Melayu (2004) kredit adalah semua jenis pinjaman yang harus dibayar kembali bersama bunganya oleh peminjam sesuai dengan perjanjian yang telah disepakati. Untuk tetap menjaga kualitas kredit yang diberikan dan menghindari kegagalan keuangan bank akibat resiko kredit yang terlalu besar, maka dibutuhkan cara untuk mengidentifikasi nasabah berpotensi kredit macet yakni salah satunya dengan metode *Credit Scoring*. Satu diantara metode statistika yang dapat digunakan untuk memprediksi klasifikasi pada *Credit Scoring* adalah *Bagging k-Nearest Neighbor*. Metode ini menggunakan sejumlah  $k$ -objek tetangga terdekat antara data *testing* dengan data *training* yang di-*resampling* sebanyak  $B$  kali. Dalam tugas akhir ini, digunakan enam variabel independen yakni usia, lama kerja, pendapatan bersih, pinjaman lain, nominal akun, dan rasio hutang. Berdasarkan analisis, diperoleh nilai optimal parameternya adalah  $k = 1$  dan ketepatan prediksi klasifikasi status kredit menggunakan *Bagging k-Nearest Neighbor* adalah sebesar 66,67%.

Kata kunci : *Credit scoring*, Klasifikasi, *Bagging k-Nearest Neighbor*

## ABSTRACT

According to Melayu (2004) credit is all types of loans that have to be paid along with the interest by the borrower according to the agreed agreement. To keep the quality of loans and avoid financial failure of banks due to large credit risks, we need a method to identified any potentially customer's with bad credit status, one of the methods is *Credit Scoring*. One of Statistical method that can predict the classification for *Credit Scoring* called *Bagging k-Nearest Neighbor*. This Method uses *k*-object nearest neighbor between data testing to *B*-bootstrap of the training dataset. This classification will use six independence variables to predict the class, these are Age, Work Year, Net Earning, Other Loan, Nominal Account and Debt Ratio. The result determine  $k=1$  as the optimal *k*-value and show that *Bagging k-Nearest Neighbor's* accuracy rate is 66,67%.

Key word : Credit scoring, Classification, Bagging k-Nearest Neighbor

**ANALISIS *CREDIT SCORING* MENGGUNAKAN METODE  
*BAGGING K-NEAREST NEIGHBOR***



=====  
**SKRIPSI**  
=====

Disusun oleh :  
**FATIMAH**  
**24010210120028**

**DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2016**

**ANALISIS *CREDIT SCORING* MENGGUNAKAN METODE  
*BAGGING K-NEAREST NEIGHBOR***

**Disusun Oleh :**

**FATIMAH**

**24010210120028**

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains  
pada Departemen Statistika**

**DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2016**

## HALAMAN PENGESAHAN I

Judul : *Analisis Credit Scoring Menggunakan Metode Bagging k-Nearest*

*Neighbor*

Nama : Fatimah

NIM : 24010210120028

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir dan dinyatakan lulus pada tanggal 3 November 2016.

Semarang, November 2016

Mengetahui,

Ketua Departemen Statistika  
FSM UNDIP

Ketua Panitia Penguji  
Ujian Tugas Akhir,

**Dr. Tarno, M.Si**  
NIP. 196307061991021001

**Prof. Drs. Mustafid, M.Eng, Ph.D**  
NIP. 1955052819800310002

## HALAMAN PENGESAHAN II

Judul : Analisis *Credit Scoring* Menggunakan Metode *Bagging k-Nearest*

*Neighbor*

Nama : Fatimah

NIM : 24010210120028

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 3 November 2016

Semarang, 3 November 2016

Pembimbing I

Pembimbing II

**Moch. Abdul Mukid, S.Si, M.Si**

**NIP. 197808172005011001**

**Drs. AgusRusgiyono, M.Si**

**NIP. 196408131990011001**

## KATA PENGANTAR

Puji syukur penulis ucapkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “**Analisis Credit Scoring Menggunakan Metode Bagging k-Nearest Neighbor**”. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Bapak Dr. Tarno, M.Si. selaku Ketua Departemen Statistika Fakultas Sains dan Matematika.
2. Bapak Moch. Abdul Mukid, S.Si, M.Si. dan Bapak Drs. Agus Rusgiyono, M.Si. sebagai dosen pembimbing yang telah memberikan bimbingan dan pengarahan dalam penulisan tugas akhir ini.
3. Bapak dan Ibu dosen Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro yang telah memberikan ilmu yang bermanfaat.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis menyelesaikan tugas akhir ini.

Penulis berharap tugas akhir ini dapat bermanfaat bagi civitas akademika di Universitas Diponegoro, khususnya Departemen Statistika dan masyarakat pada umumnya.

Semarang, November 2016

Penulis

## ABSTRAK

Menurut Melayu (2004) kredit adalah semua jenis pinjaman yang harus dibayar kembali bersama bunganya oleh peminjam sesuai dengan perjanjian yang telah disepakati. Untuk tetap menjaga kualitas kredit yang diberikan dan menghindari kegagalan keuangan bank akibat resiko kredit yang terlalu besar, maka dibutuhkan cara untuk mengidentifikasi nasabah berpotensi kredit macet yakni salah satunya dengan metode *Credit Scoring*. Satu diantara metode statistika yang dapat digunakan untuk memprediksi klasifikasi pada *Credit Scoring* adalah *Bagging k-Nearest Neighbor*. Metode ini menggunakan sejumlah *k*-objek tetangga terdekat antara data *testing* dengan data *training* yang di-*resampling* sebanyak *B* kali. Dalam tugas akhir ini, digunakan enam variabel independen yakni usia, lama kerja, pendapatan bersih, pinjaman lain, nominal akun, dan rasio hutang. Berdasarkan analisis, diperoleh nilai optimal parameternya adalah  $k = 1$  dan ketepatan prediksi klasifikasi status kredit menggunakan *Bagging k-Nearest Neighbor* adalah sebesar 66,67%.

Kata kunci : *Credit scoring*, Klasifikasi, *Bagging k-Nearest Neighbor*

## ABSTRACT

According to Melayu (2004) credit is all types of loans that have to be paid along with the interest by the borrower according to the agreed agreement. To keep the quality of loans and avoid financial failure of banks due to large credit risks, we need a method to identified any potentially customer's with bad credit status, one of the methods is *Credit Scoring*. One of Statistical method that can predict the classification for *Credit Scoring* called *Bagging k-Nearest Neighbor*. This Method uses *k*-object nearest neighbor between data testing to *B*-bootstrap of the training dataset. This classification will use six independence variables to predict the class, these are Age, Work Year, Net Earning, Other Loan, Nominal Account and Debt Ratio. The result determine  $k = 1$  as the optimal *k*-value and show that *Bagging k-Nearest Neighbor's* accuracy rate is 66,67%.

Key word : Credit scoring, Classification, Bagging k-Nearest Neighbor

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN I.....	ii
HALAMAN PENGESAHAN II.....	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
BAB II TINJAUAN PUSTAKA	
2.1 Kredit.....	4
2.2 <i>Credit Scoring</i> .....	7
2.3 <i>Data Mining</i> .....	7
2.3.1 Klasifikasi.....	9
2.3.2 Pendekatan Teknik Klasifikasi.....	10
2.3.3 Analisis <i>k-Nearest Neighbor</i> .....	13
2.3.3.1 Jarak Euclidian.....	13
2.3.3.2 Penentuan Parameter.....	14
2.3.3.3 Algoritma <i>k-Nearest Neighbor</i> (kNN).....	15
2.3.3.4 Karakteristik Klasifikasi dengan <i>k-Nearest Neighbor</i> .....	16
2.4 Metode <i>Bagging (Bootstrap Aggregating)</i> .....	17

BAB III METODE PENELITIAN	
3.1 Sumber Data.....	20
3.2 Variabel Penelitian.....	20
3.3 Metode Penelitian.....	21
3.4 Diagram Alir.....	23
BAB IV HASIL DAN PEMBAHASAN	
4.1 Deskripsi Data.....	24
4.2 Pembagian Data.....	25
4.3 Penentuan Nilai Parameter Terbaik.....	26
4.4 Penentuan Klasifikasi Kelas untuk Observasi Pertama Data <i>Testing</i> ..	27
4.4.1 Perhitungan Jarak antara Data <i>Testing</i> dengan Data <i>Bootstrap</i> ..	27
4.4.2 Penentuan Tetangga Terdekat Sebanyak $k$ pada Setiap Data <i>Bootstrap</i> .....	30
4.4.3 Penentuan Kelas dari Observasi Pertama Data <i>Testing</i> Berdasarkan Hasil <i>Majority Vote</i> .....	31
4.5 <i>Accuracy</i> (Ketepatan) Hasil Prediksi Gabungan Menggunakan <i>Bagging k-Nearest Neighbor</i> .....	32
BAB V KESIMPULAN.....	35
DAFTAR PUSTAKA.....	36
LAMPIRAN.....	38

## DAFTAR GAMBAR

	Halaman
Gambar 1 Ilustrasi Model Klasifikasi.....	11
Gambar 2 Ilustrasi Metode Kombinasi Model.....	18

## DAFTAR TABEL

Halaman

Tabel 1 Deskripsi Variabel Usia, Lama Kerja, Pinjaman Lain, Pendapatan Bersih, Nominal Akun dan Rasio Hutang.....	24
Tabel 2 Deskripsi Data Usia, Lama Kerja, Pinjaman Lain, Pendapatan Bersih, Nominal Akun dan Rasio Hutang pada Data <i>Training</i> .....	25
Tabel 3 Deskripsi Data Usia, Lama Kerja, Pinjaman Lain, Pendapatan Bersih, Nominal Akun dan Rasio Hutang pada Data <i>Testing</i> .....	25
Tabel 4 Perbandingan Nilai <i>Error Rate</i> Hasil Klasifikasi Kelas Data <i>Training</i> dan Data <i>Testing</i> .....	26
Tabel 5 Data <i>Bootstrap</i> Pertama.....	28
Tabel 6 Observasi Pertama pada Data <i>Testing</i> .....	28
Tabel 7 Jarak Euclidian Data <i>Bootstrap</i> Pertama ke Data <i>Testing</i> .....	30
Tabel 8 Urutan Jarak dari Nilai yang Terkecil Hingga Terbesar untuk Data <i>Bootstrap</i> Pertama.....	31
Tabel 9 Hasil Prediksi Kelas untuk Status Kredit Berdasarkan Jarak dengan Data <i>Bootstrap</i> .....	32
Tabel 10 Hasil <i>Majority Vote</i> untuk Penentuan Kelas Observasi Pertama Data <i>Testing</i> .....	33

## DAFTAR LAMPIRAN

	Halaman
Lampiran1 Data <i>Training</i> , n = 900 .....	39
Lampiran 2 Data <i>Testing</i> , n = 99 .....	40
Lampiran 3 Syntax <i>Cross Validation</i> di SAS.....	41
Lampiran 4 Syntax <i>Bootstrap</i> di SAS.....	42
Lampiran 5 Hasil Klasifikasi Kelas Lengkap Berdasarkan Data <i>Bootstrap</i> Menggunakan <i>Majority Vote</i> .....	43



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Istilah kredit berasal dari bahasa Yunani yaitu '*credere*' yang berarti kepercayaan (Herprasetyo, 2012). Menurut Undang-undang RI nomor 10 tahun 1998, perbankan adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkan kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak. Dalam perundangan yang sama, pengertian kredit adalah suatu penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga.

Menurut Chapra (2000), kredit harus dialokasikan dengan tujuan membantu merealisasikan kemaslahatan sosial secara umum. Tujuan dari kesepakatan kredit menurut Kasmir (2002) adalah memberikan keuntungan bagi bank, membantu usaha nasabah dan membantu pemerintah. Lebih rincinya, bagi debitur yang tidak mempunyai cukup modal untuk membuka usaha, kegiatan kredit ini sangat membantu debitur untuk menjalankan usahanya. Sedangkan untuk pihak bank sendiri, pemberian kredit memperoleh peningkatan laba bagi bank yang berasal dari bunga yang dibayarkan oleh nasabah. Dan bagi pemerintah, semakin banyak kredit yang disalurkan oleh pihak perbankan, maka semakin banyak peningkatan pembangunan di berbagai sektor. Namun jika ditemui kondisi dimana nasabah sudah tidak sanggup membayar sebagian atau seluruh kewajibannya kepada bank

seperti yang telah diperjanjikan maka kondisi ini disebut dengan kredit macet (Kuncoro dan Suhardjono, 2002). Semakin banyak jumlah nasabah kredit macet dalam sebuah bank, akan menghambat kinerja bank dan melemahkan pendapatan dari usaha perbankan tersebut.

Untuk mencegah terjadinya hal tersebut, maka dibutuhkan metode untuk membantu pihak bank memprediksi lebih awal potensi kredit macet dari calon nasabah kredit tersebut. Sehingga dapat dilakukan pertimbangan atas persetujuan pengajuan kredit oleh calon nasabah tersebut ataukah tidak. Metode untuk menilai kelayakan dari calon nasabah kredit diantaranya adalah metode *Credit Scoring*.

Metode statistika yang dapat digunakan untuk mengklasifikasikan kondisi status kredit nasabah menggunakan *Credit Scoring* diantaranya adalah metode *k-Nearest Neighbor* (kNN). Metode *k-Nearest Neighbor*, dapat mengklasifikasikan apakah seorang nasabah kredit termasuk dalam kategori kredit lancar atau tidak, dengan catatan data-data yang dilibatkan terjamin akurasi. Metode *k-Nearest Neighbor* merupakan kelompok *instance-based learning*, dimana pengklasifikasian dilakukan dengan mencari kelompok *k* objek dalam data *training* yang paling dekat dengan objek pada data baru atau data *testing*.

Telah banyak dilakukan inovasi untuk meningkatkan keakuratan dalam hal memprediksi suatu klasifikasi, diantaranya seperti yang dikembangkan oleh Breiman (1996) yakni metode klasifikasi yang menggabungkan dua metode yakni *Bootstrap* dan juga *Aggregating*, dimana metode ini lebih dikenal sebagai metode *Bagging*.

Berdasarkan latar belakang masalah tersebut dilakukan penelitian dengan judul “**Analisis Credit Scoring Menggunakan Metode *Bagging k-Nearest Neighbor***”.

## **2.1 Perumusan Masalah**

Berdasarkan latar belakang yang telah dikemukakan tersebut dapat dirumuskan masalah yang diteliti adalah bagaimana menerapkan Metode *Bagging k-Nearest Neighbor* dalam mengklasifikasi kelompok status kredit macet dan lancar pada data nasabah bank ‘X’?

## **3.1 Pembatasan Masalah**

Pada penelitian ini terdapat beberapa batasan masalah yang bertujuan agar pembahasan lebih mudah dan lebih fokus. Batasan masalah tersebut adalah sebagai berikut:

- 1) Penelitian ini hanya dibatasi pada data tahun 2011 Laporan kondisi Nasabah Kredit di Bank ‘X’.
- 2) Masalah yang diteliti adalah, menganalisis hasil klasifikasi kelompok nasabah kredit lancar atau macet pada Bank ‘X’ dengan menggunakan metode *Bagging k-Nearest Neighbor*.

## **4.1 Tujuan Penelitian**

Tujuan dalam penulisan tugas akhir ini adalah memperoleh hasil prediksi klasifikasi status kredit macet dan lancar pada data Nasabah Bank ‘X’ menggunakan metode *Bagging k-Nearest Neighbor*.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Kredit**

Secara etimologi, kata kredit berasal dari bahasa Yunani yaitu “*Cradere*” yang berarti “Kepercayaan” (Herprasetyo, 2012). Menurut Kasmir (2001), kredit adalah pemberian prestasi (misalnya uang, barang) dengan balas prestasi (kontra prestasi) yang akan terjadi pada waktu yang akan datang. Menurut Melayu (2004), kredit adalah semua jenis pinjaman yang harus dibayar kembali bersama bunganya oleh peminjam sesuai dengan perjanjian yang telah disepakati. Pengertian kredit menurut UU No.10 Tahun 1998, yaitu meminjamkan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan perjanjian tertulis baik dibawah tangan maupun dihadapan notaris dari berbagai pengamanan maka debitur akan menyerahkan suatu jaminan baik yang berupa kebendaan maupun yang bukan kebendaan, dan pihak debitur berkewajiban untuk melunasi utangnya setelah jangka waktu tertentu di masa mendatang dengan balas prestasi yaitu berupa pemberian bunga.

Bedasarkan pengertian di atas dapat disimpulkan bahwa kredit adalah pemberian uang atau barang kepada pihak lain yang didasarkan atas kepercayaan disertai dengan balas jasa dan jangka waktu tertentu. Kredit bermasalah akan berdampak negatif baik bagi kelangsungan hidup bank itu sendiri maupun bagi perekonomian negara, menurut Mahmoedin (2002) kondisi tersebut akan mempengaruhi penghasilan bank, kepercayaan masyarakat kepada bank, menurunkan tingkat kesehatan bank dan juga mempengaruhi modal bank tidak dapat berkembang dengan baik.

Sebelum kredit diberikan, bank terlebih dahulu mengadakan analisis kredit. Tujuan analisis ini adalah agar bank yakin bahwa kredit yang diberikan benar-benar aman dalam arti uang yang disalurkan pasti kembali. Menurut Ismail (2010), analisis kredit adalah suatu proses analisis kredit yang dilakukan oleh bank untuk menilai suatu permohonan kredit yang telah diajukan oleh calon debitur.

Menurut Rivai dan Permata dalam bukunya *Credit Management Handbook* (2007), hal-hal yang terlebih dahulu harus dipenuhi adalah prinsip *6C's analysis*, yaitu sebagai berikut:

1. *Character*

*Character* adalah keadaan watak (sifat) dari nasabah, baik dalam kehidupan pribadi maupun dalam lingkungan usaha. Kegunaan dari penilaian terhadap karakter ini adalah untuk mengetahui sampai sejauh mana itikat atau kemauan nasabah untuk memenuhi kewajibannya (*willingness to pay*) sesuai dengan perjanjian yang telah ditetapkan.

2. *Capital*

*Capital* adalah jumlah dana (modal) sendiri yang dimiliki oleh calon nasabah. Semakin besar modal sendiri dalam perusahaan, tentu semakin tinggi kesungguhan calon nasabah dalam menjalankan usahanya dan bank akan semakin lebih yakin dalam memberikan kredit. Kemampuan modal sendiri merupakan benteng yang kuat agar tidak mudah mendapat guncangan dari luar, misalnya jika terjadi kenaikan suku bunga, komposisi modal sendiri ini perlu ditingkatkan. Penilaian atas besarnya modal sendiri merupakan hal yang penting mengingat kredit bank hanya sebagai

tambahan pembiayaan dan bukan untuk membiayai seluruh modal yang diperlukan.

### 3. *Capacity*

*Capacity* adalah kemampuan yang dimiliki calon nasabah dalam menjalankan usahanya guna memperoleh laba yang diharapkan. Kegunaan dari penilaian ini adalah untuk mengukur (untuk mengetahui) sampai sejauh mana calon nasabah mampu mengembalikan atau melunasi hutang-hutangnya (*ability to pay*) secara tepat waktu dari usaha yang diperolehnya.

### 4. *Collateral*

*Collateral* adalah barang-barang yang diserahkan nasabah sebagai agunan terhadap kredit yang diterimanya. *Collateral* tersebut harus dinilai oleh bank untuk mengetahui sejauh mana risiko kewajiban finansial nasabah kepada bank. Penilaian terhadap jaminan ini meliputi jenis, lokasi, bukti kepemilikan, dan status hukumnya.

### 5. *Condition of economy*

*Condition of economy*, yaitu situasi dan kondisi politik, sosial, ekonomi, budaya yang memengaruhi keadaan perekonomian pada suatu saat yang berpotensi memengaruhi kelancaran perusahaan calon debitur.

### 6. *Constraint*

*Constaint* adalah batasan dan hambatan yang tidak memungkinkan suatu bisnis untuk dilaksanakan pada tempat tertentu, misalkan pendirian suatu usaha pompa bensin yang disekitarnya banyak bengkel las atau pembakaran batu bata.

## 2.2 *Credit Scoring*

Metode *credit scoring* dapat diterjemahkan sebagai salah satu metode statistika (atau metode kuantitatif) yang digunakan untuk memprediksi peluang calon nasabah tidak sanggup membayar kembali pinjaman, ataupun bagi nasabah kredit yang sudah terdaftar berpeluang menjadi nasabah yang pembayaran pinjamannya menunggak (Mester, 1997). Penggunaan metode *credit scoring* ini dapat membantu pihak yang memberi pinjaman menentukan apakah pengajuan kredit seorang peminjam dapat disetujui (Morrison, 2004). Dari pernyataan tersebut dapat disimpulkan bahwa *credit scoring* atau penilaian kredit adalah sistem (cara) yang digunakan oleh bank atau lembaga pembiayaan lainnya yang berguna untuk menentukan apakah yang bersangkutan layak atau tidak mendapatkan pinjaman. Sedangkan untuk data nasabah yang dibutuhkan dalam penerapan *credit scoring*, diambil dari masing-masing data aplikasi pinjaman nasabah, selain dengan menggunakan program statistik yang berisi tentang sejarah pinjaman yang bersangkutan (Mester, 1997), data juga berisi antara lain mengenai siklus pembayaran tagihan apakah tepat waktu atau tidak, berapa banyak kredit yang masih atau pernah dimiliki yang bersangkutan.

## 2.3 *Data Mining*

Menurut Hermawan (2013), *data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang

dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Database (KDD)* adalah penerapan metode saintifik pada *data mining*.

Menurut Prasetyo (2014), *data mining* bertujuan untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna. Dengan kondisi adanya sisi data yang belum dieksplorasi, dan di sisi lain, kemajuan teknik komputasi dan ilmu komputer sudah tumbuh pesat, keterbutuhan akan ekplorasi informasi baru semakin meningkat. *Data mining* menjadi solusi penyelesaian pencarian informasi yang sebelumnya tidak dapat dideteksi secara tradisional dan hanya menggunakan kemampuan analisis manusia.

Proses dalam tahap *data mining* terdiri dari tiga langkah utama (Sumathi dan Sivanandam, 2006), yaitu:

a. *Data Preparation*

Pada langkah ini, data dipilih, dibersihkan (*cleaning*) dan dilakukan *preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

b. Algoritma *data mining*

Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi tambahan yang dibutuhkan.

### c. Fase analisis data

Keluaran dari *data mining* dievaluasi untuk melihat apakah *knowledge* domain ditemukan dalam bentuk *rule* yang telah diekstrak dari jaringan.

*Data mining* didefinisikan sebagai proses penemuan pola dalam data (Witten *et al.*, 2011). Berdasarkan tugasnya, *data mining* dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering* dan asosiasi (Larose, 2005).

#### 2.3.1 Klasifikasi

Teknik data mining yang digunakan dalam tugas akhir ini adalah teknik klasifikasi. Menurut Prasetyo (2014), klasifikasi adalah proses pembelajaran suatu fungsi tujuan (target)  $f$  yang memetakan tiap himpunan atribut  $x$  ke satu dari label kelas himpunan atribut  $y$  yang didefinisikan sebelumnya. Fungsi targetan disebut juga model klasifikasi. Algoritma klasifikasi yang digunakan dalam penulisan tugas akhir ini adalah *k-nearest neighbor*.

Menurut Han dan Kamber (2006) ada dua jenis Klasifikasi, yaitu:

##### 1. Pemodelan Deskriptif

Model klasifikasi yang dapat berfungsi sebagai suatu alat penjelasan untuk membedakan objek-objek dalam kelas-kelas yang berbeda.

##### 2. Pemodelan Prediktif

Model klasifikasi yang dapat digunakan untuk memprediksi label kelas *record* yang tidak diketahui.

Menurut Gorunescu (2011), proses klasifikasi data didasarkan pada empat komponen, yaitu:

a. Kelas

Variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objeknya. Contohnya: status kredit macet dan lancar.

b. *Predictor*

Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: usia, lama bekerja, total pendapatan bersih, ataupun total pinjaman lain yang dimiliki nasabah kredit.

c. *Training dataset*

Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

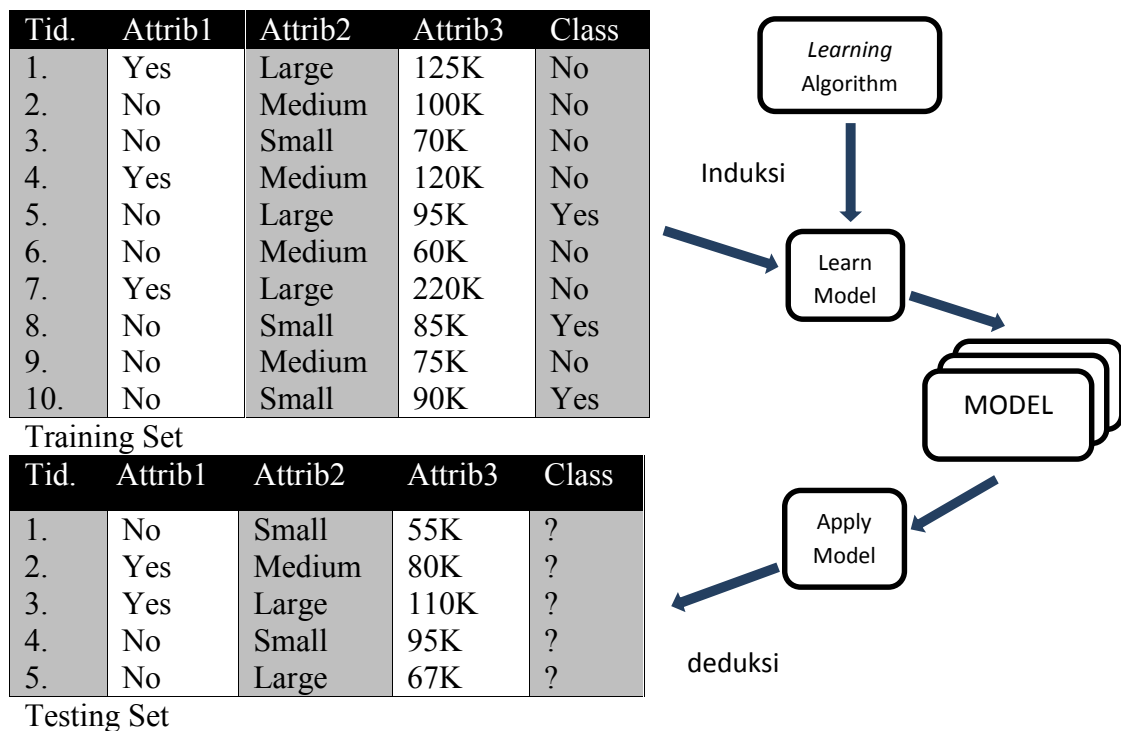
d. *Testing dataset*

Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

### 2.3.2 Pendekatan Teknik Klasifikasi

Teknik klasifikasi merupakan suatu pendekatan sistematis untuk membangun model klasifikasi dari suatu himpunan data masukan. Tiap teknik menggunakan suatu algoritma pembelajaran (*learning algorithm*) untuk mendapatkan suatu model yang paling memenuhi hubungan antara himpunan atribut dan label kelas dalam data

masukannya. Tujuan dari algoritma pembelajaran adalah untuk membangun model yang berkemampuan baik, yaitu model yang dapat memprediksi label kelas dari *record* yang tidak diketahui kelas sebelumnya dengan lebih akurat. Biasanya *dataset* yang diberikan dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk membangun model dan data *testing* digunakan untuk memvalidasi (Hermawati, 2013). Berikut ini merupakan ilustrasi model klasifikasi untuk dua kelas dengan menggunakan sepuluh data *training*, lima data *testing* yang dapat dilihat pada Gambar 1.



**Gambar 1.** Ilustrasi Model Klasifikasi

Pada Gambar 1, terdapat langkah-langkah dalam menerapkan model klasifikasi, dimana proses yang digunakan untuk membentuk model klasifikasi adalah dengan algoritma pembelajaran berdasarkan himpunan *training set* yang disebut dengan proses induksi. Sedangkan

proses penerapan model klasifikasi untuk memprediksi kelas tabel dari data dalam himpunan *testing set* disebut dengan proses deduksi.

Untuk mengevaluasi performa dari model yang dibangun, perlu dilakukan pengukuran performa, yaitu pengukuran akurasi (*accuracy*) atau tingkat kesalahan (*error rate*). Jika  $f_{ij}$  menotasikan jumlah *record* dari kelas  $i$  yang berada di kelas  $j$  pada saat pengujian (Hermawati, 2013), maka pengukuran akurasi dapat dituliskan sebagai berikut:

$$\text{Error Rate} = \frac{\text{jumlah\_prediksi\_yang\_salah}}{\text{jumlah\_prediksi\_keseluruhan}} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (1)$$

Dimana:

$f_{11}$  = jumlah data kelompok 1 yang benar diklasifikasikan ke kelompok 1

$f_{00}$  = jumlah data kelompok 0 yang benar diklasifikasikan ke kelompok 0

$f_{10}$  = jumlah data kelompok 1 yang salah diklasifikasikan ke kelompok 0

$f_{01}$  = jumlah data kelompok 0 yang salah diklasifikasikan ke kelompok 1

### 2.3.3 Analisis *k-Nearest Neighbor* (kNN)

Metode *k-nearest neighbor* (kNN) termasuk dalam kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. Metode kNN dilakukan dengan mencari kelompok  $k$  objek dalam data *training* yang paling dekat dengan objek pada data baru atau data *testing* (Wu dan Kumar, 2009).

### 2.3.3.1 Jarak Euclidian

Pada metode klasifikasi seperti metode *k-nearest neighbor* menggunakan suatu kuantitas yang disebut dengan kedekatan atau *proximity*. Secara geometris, pengukuran kedekatan akan diukur berdasarkan jarak antara data yang pertama terhadap data yang pertama terhadap data yang kedua. Semakin dekat jarak kedua data maka akan semakin besar kemiripannya, dan semakin jauh jarak kedua data maka semakin kecil kemiripannya (Prasetyo, 2014).

Jarak yang digunakan untuk mengukur jarak antar data pada tugas akhir ini adalah jarak Euclidian. Diketahui permisalan, obyek A dengan hasil pengukuran adalah  $(x_1, x_2, \dots, x_p)$  dan obyek B dengan hasil pengukuran adalah  $(y_1, y_2, \dots, y_p)$  maka jarak euclidian antara A dan B adalah sebagai berikut:

$$d(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (2)$$

### 2.3.3.2 Penentuan Parameter

Di dalam penggunaan algoritma kNN, diperlukan perhitungan jarak dari sebuah datum ke datum yang lain. Nilai jarak inilah yang digunakan sebagai nilai kedekatan antara data uji (data *testing*) dengan data latih (data *training*). Nilai  $k$  pada kNN berarti  $k$ -data terdekat dari data *testing*. Jika  $k=1$ , kelas dari satu data *training* sebagai tetangga terdekat (terdekat pertama) dari data *testing* tersebut akan diberikan sebagai kelas untuk data

*testing*, yaitu kelas 1. Jika  $k=2$ , akan diambil 2 tetangga terdekat dari data *training*. Begitu juga jika nilai  $k=3, 4, 5$  dan seterusnya. Jika dalam  $k$ -tetangga ada dua kelas yang berbeda, akan diambil kelas dengan jumlah data terbanyak (*voting majority*) (Prasetyo, 2012).

Langkah yang digunakan dalam metode *k-Nearest Neighbor*:

1. Tentukan parameter  $k$  (jumlah tetangga paling dekat), menggunakan metode *hold-out*.
2. Hitung jarak Euclidian antara data *testing* dengan data *training* yang ada.
3. Urutkan jarak tersebut dimulai dari kelompok yang memiliki jarak terkecil ke yang terbesar.
4. Tentukan jarak terdekat sebanyak nilai  $k$  yang telah ditentukan.
5. Dengan kategori *nearest neighbor* yang paling banyak, maka dapat ditetapkan kelas terbanyak tersebut sebagai kelas dari masing-masing objek data *testing*.

### 2.3.3.3 Algoritma *k-Nearest Neighbor* (k-NN)

Pada proses klasifikasi data, penentuan nilai parameter merupakan hal yang sangat memengaruhi ketepatan suatu metode dalam mengklasifikasikan data. Pada metode *k-nearest neighbor* ini, parameter yang dimaksud adalah nilai  $k$ . Nilai  $k$  yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Pemilihan

jumlah  $k$  yang paling tepat perlu dijajaki agar *error rate* bisa diperkecil. Nilai  $k$  ganjil untuk kelas berjumlah dua akan memudahkan *voting* karena dijamin tidak akan terjadi dua kelas yang mendapatkan suara *voting* yang sama. Akan tetapi jika  $k$  genap, akan ada kemungkinan dua kelas mendapatkan suara *voting* yang sama. Nilai  $k$  yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*.

Terdapat berbagai langkah dalam menetapkan  $k$  terbaik, diantaranya adalah proses *cross validation* dengan menggunakan metode *hold-out* (Bishop, 1995) sebagai berikut:

1. Secara acak, bagi data menjadi 2 bagian, yakni data *training*, dan data *testing*.
2. Tentukan nilai parameter *nearest neighbor*  $k = 1$ .
3. Hitung jarak Euclidian antara data *training* dengan data *testing*.
4. Urutkan nilai jarak tersebut, dari yang nilainya terkecil hingga yang terbesar.
5. Tentukan jarak terdekat sebanyak nilai  $k$  yang telah ditentukan.
6. Tetapkan kelas yang terbanyak muncul dari hasil data tetangga terdekat tersebut, sebagai kelas dari data *testing*.
7. Hitunglah *error rate* hasil klasifikasi kelas data *training* dan data *testing* untuk nilai  $k$  tersebut.

8. Kemudian tentukan nilai  $k=k+1$
9. Lakukan kembali langkah 3 – 7 hingga didapati nilai *error rate* yang paling rendah.
10. Lalu tentukan nilai  $k$  yang mempunyai nilai *error rate* yang paling rendah untuk kedua hasil klasifikasi data *training* dan data *testing* sebagai nilai  $k$  yang terbaik, yang akan digunakan untuk mengklasifikasikan data *testing* menggunakan *Bagging*.

#### 2.3.3.4 Karakteristik Klasifikasi dengan *k-Nearest Neighbor*

*k-Nearest Neighbor* (kNN) merupakan teknik klasifikasi yang sederhana tetapi mempunyai hasil yang cukup bagus. Meskipun begitu, kNN juga mempunyai kelebihan dan kekurangan. Beberapa karakteristik kNN adalah sebagai berikut:

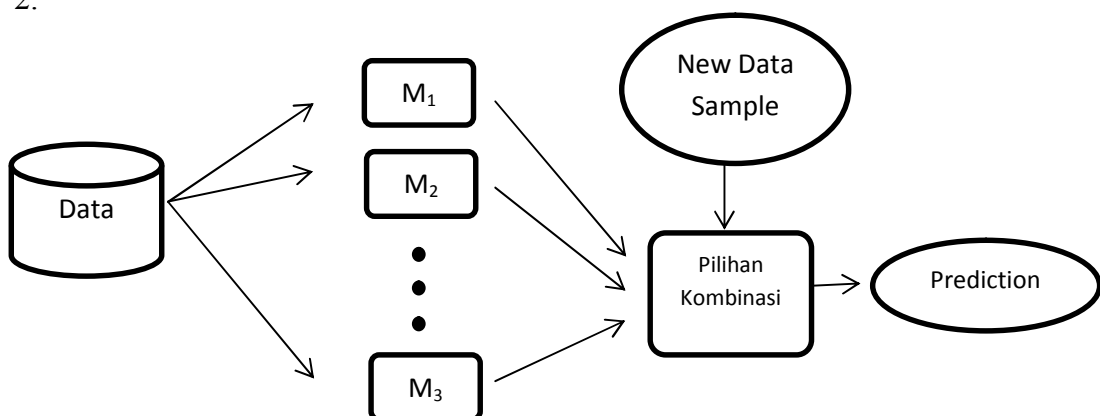
1. kNN merupakan algoritma yang menggambarkan seluruh data *training* untuk melakukan proses klasifikasi (*complete storage*). Hal ini mengakibatkan proses prediksi yang sangat lama untuk data dalam jumlah yang sangat besar. Tetapi semakin banyak data *training*, maka akan dapat dibuat batas keputusan yang semakin halus.
2. Algoritma kNN tidak membedakan setiap fitur dengan suatu bobot, semua bobot untuk fiturnya bernilai sama untuk satu sama lainnya.
3. Karena kNN masuk kategori *lazy learning* yang menyimpan sebagian atau semua data dan hampir tidak ada proses

pelatihan, kNN sangat cepat dalam proses pelatihan (karena memang tidak ada), tetapi sangat lambat dalam proses prediksi.

4. Hal yang rumit adalah menentukan nilai  $k$  yang paling sesuai karena kNN pada prinsipnya memilih tetangga terdekat, parameter jarak juga penting untuk dipertimbangkan sesuai dengan kasus datanya. Jarak Euclidian sangat cocok untuk menggunakan jarak terdekat (lurus) antara dua data.

#### 2.4 Metode *Bagging* (*Bootstrap Aggregating*)

Metode kombinasi model adalah metode yang masing-masingnya merupakan kombinasi barisan yang memuat sejumlah  $k$ -model yakni  $M_1, M_2, \dots, M_k$ , dengan tujuan untuk membuat sebuah peningkatan pada model gabungan,  $M^*$ . Secara umum terdapat dua metode kombinasi model, yakni *Bagging* dan *Boosting*. Keduanya dapat digunakan dalam proses klasifikasi dan prediksi (Han dan Kamber, 2006). Ilustrasi metode gabungan tersebut dapat dilihat pada Gambar 2.



**Gambar 2.** Ilustrasi Metode Kombinasi Model

Pada Gambar 2, terdapat gambaran langkah kerja dari metode kombinasi model, dimana metode tersebut bertujuan untuk meningkatkan ketepatan model yang terdiri dari metode *Bagging* dan *Boosting*. Metode ini menghasilkan sebuah model klasifikasi atau prediksi,  $M_1, M_2, \dots, M_k$ . Dimana strategi *voting* terhadap pilihan kombinasi yang muncul adalah strategi yang digunakan untuk menggabungkan prediksi untuk objek yang telah diberikan yang belum diketahui kategorinya.

Metode *Bagging* merupakan teori yang diusulkan oleh Breiman (1996), yang didasari oleh konsep teori *bootstrap* dan *aggregating* yang menggabungkan kedua manfaat teori tersebut menjadi satu. *Bootstrap* ini diterapkan berdasarkan pada teori *sampling* acak dengan pengembalian (Tibshirani dan Efron, 1993). Oleh karena itu *bootstrap*  $X^b = (X_1^b, X_2^b, \dots, X_n^b)$  (menerapkan pengambilan sampel acak dengan pengembalian) dari data *training*  $= (X_1, X_2, \dots, X_n)$ , dimana penggunaan metode ini dianggap dapat meminimalisir kesalahan dalam pengklasifikasian. Sebagai pengaruhnya, model pada teknik klasifikasi (*classifier*) yang telah dibentuk tersebut kemungkinan mempunyai kinerja yang lebih baik. Pengertian dari proses *aggregating* adalah menggabungkan beberapa *classifier* tersebut.

Terkadang *classifier* gabungan memberikan hasil yang lebih baik dibandingkan dengan satu *classifier* saja, dikarenakan adanya penggabungan kedua manfaat dari beberapa *classifier* tersebut pada akhir penyelesaiannya. Oleh karena itu *Bagging* bermanfaat untuk membantu membangun atau membentuk *classifier* yang lebih baik pada sampel data *training*.

Berikut ini merupakan tahapan penerapan teknik *Bagging*:

1. Lakukan replikasi *bootstrap*  $X^b$  sebanyak  $m$ , dari sejumlah  $n$  data *training*.  
Ulangi langkah ini untuk  $b = 1, 2, \dots, B$ . Dimana  $m$  adalah banyaknya data yang diambil dari data *training*,  $n$  adalah ukuran sampel dari data *training* dan  $B$  adalah banyaknya replikasi *bootstrap* yang dilakukan.
2. Dengan menggunakan *simple majority vote*, dipilih label yang paling banyak muncul dari hasil penilaian sebagai aturan untuk pengambilan keputusan terakhir.

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Sumber Data**

Data yang digunakan adalah data sekunder nasabah tahun 2011 yang diperoleh dari Bank “X” yang terletak di Provinsi Lampung, dimana terdapat data diri nasabah kredit macet dan kredit lancar di dalamnya.

#### **3.2. Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini yaitu satu variabel respon (Y) dan enam variabel prediktor (X).

1. Variabel respon (Y) merupakan status kredit dari nasabah kredit yang dikategorikan dengan notasi 0 untuk kredit lancar dan notasi 1 untuk kredit macet.
2. Variabel prediktor (X) merupakan data nasabah kredit yang terdiri dari:
  - a. Usia debitur dalam satuan tahun ( $X_1$ )
  - b. Lama bekerja dalam satuan tahun ( $X_2$ )
  - c. Pendapatan bersih dalam satuan rupiah ( $X_3$ )
  - d. Pinjaman lain dalam satuan rupiah ( $X_4$ )
  - e. Nominal yang tertera dalam akun tabungan yang dimiliki ( $X_5$ )
  - f. Rasio hutang terhadap pendapatan ( $X_6$ )

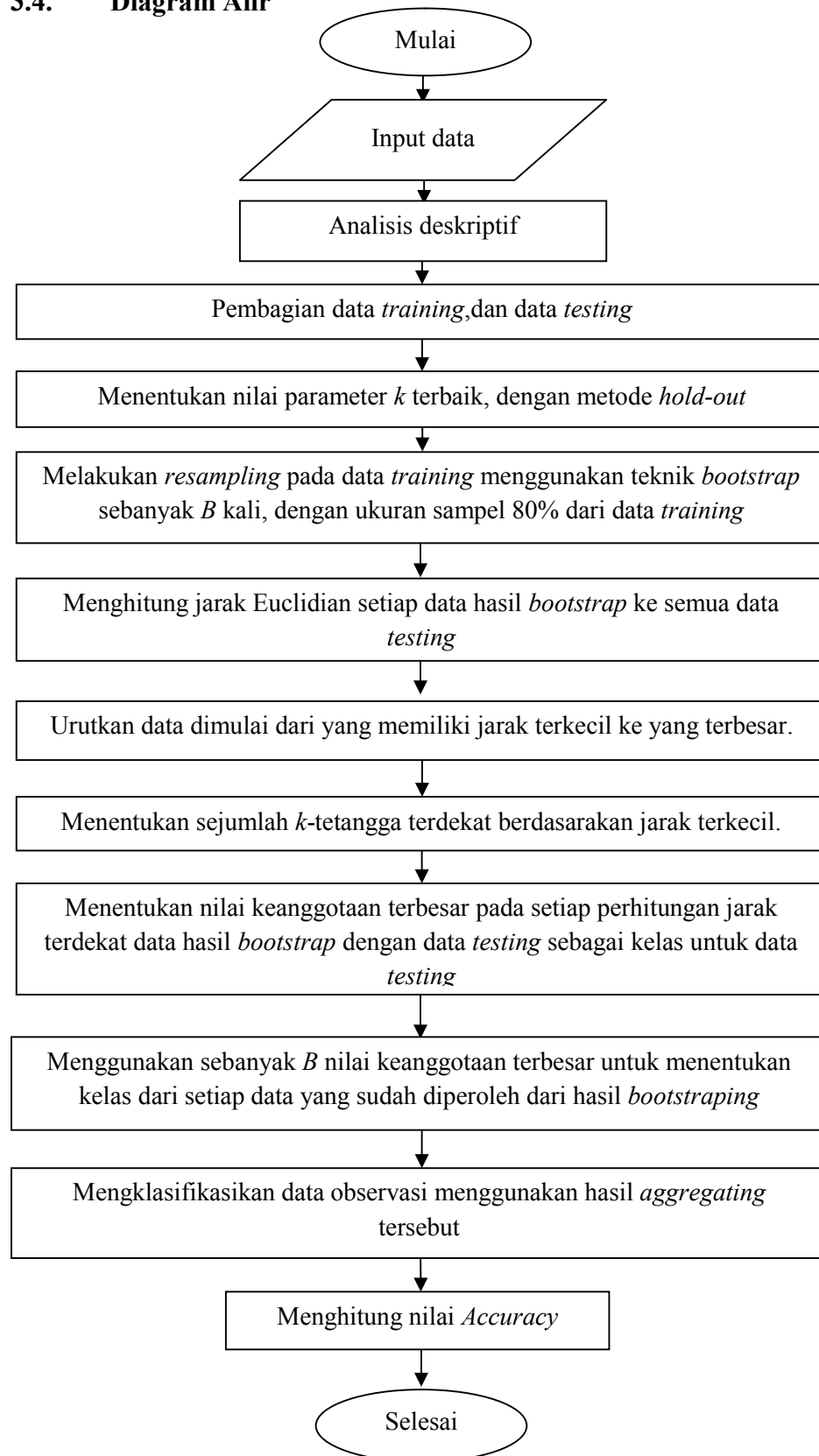
### 3.3. Metode Analisis

Software atau alat bantu yang digunakan dalam pengolahan data ini adalah Ms.Excel 2010 dan SAS 9.0. Setelah data diperoleh, maka langkah-langkah yang dilakukan dalam menganalisis data adalah:

1. Memasukkan data status kredit nasabah sebagai variabel repon sedangkan usia debitur, lama bekerja, jumlah pendapatan bersih, jumlah pinjaman lain, jumlah nominal dalam akun tabungan, dan rasio hutang terhadap pendapatan dimasukkan sebagai variabel prediktor.
2. Analisis deskriptif data debitur berdasarkan status kredit.
3. Membagi data menjadi data *training*, dan data *testing* dengan proporsi 90:10
4. Melakukan analisis *k-Nearest Neighbor* menggunakan *Bagging (Bootstrap Aggregating)* dengan langkah sebagai berikut:
  - a. Menentukan nilai  $k$  terbaik dengan *cross validation* dengan langkah sebagai berikut:
    - 1) Tentukan nilai parameter *nearest neighbor*  $k = 1$ .
    - 2) Hitung jarak Euclidian antara data *training* ke data *testing*
    - 3) Urutkan nilai jarak tersebut, dari yang nilainya terkecil hingga yang terbesar.
    - 4) Tentukan jarak terdekat sebanyak nilai  $k$  yang telah ditentukan.
    - 5) Hitunglah *error rate* hasil klasifikasi kelas data *training* dan data *testing* untuk nilai  $k$  tersebut.
    - 6) Kemudian tentukan nilai  $k = k + 1$
    - 7) Lakukan kembali langkah 2 – 5 hingga didperoleh nilai *error rate* yang paling rendah.

- 8) Lalu tentukan nilai  $k$  yang mempunyai nilai *error rate* yang paling rendah pada kedua hasil klasifikasi data *training* dan data *testing* sebagai nilai  $k$  yang terbaik, yang akan digunakan untuk mengklasifikasikan data *testing* dengan menggunakan data hasil *bootstrap*.
- b. Melakukan *resampling* pada data *training* menggunakan teknik *bootstrap* sebanyak  $B$  kali dengan ukuran sampel sebanyak  $m$ , dimana jumlah  $m$  adalah 80% dari  $n$  data *training*
  - c. Menghitung jarak Euclidian setiap data hasil *bootstrap* data *training* kesemua data *testing*
  - d. Urutkan objek–objek dimulai dari kelompok yang memiliki jarak terkecil ke yang terbesar.
  - e. Menentukan nilai keanggotaan terbesar pada tiap perhitungan jarak terdekat data hasil *bootstrap* dengan data *testing*.
  - f. Melakukan *aggregating* dengan aturan *majority vote*, menggunakan nilai keanggotaan dengan jumlah terbesar untuk menentukan kelas dari setiap data yang sudah diperoleh dari hasil *bootstrapping*, untuk menentukan kategori untuk setiap nilai observasi pada data *testing*.
  - g. Mengklasifikasikan data observasi menggunakan hasil *aggregating* tersebut.
  - h. Melakukan validasi hasil prediksi dengan menghitung nilai keakuratan hasil klasifikasi dengan menghitung nilai *accuracy* dari analisis tersebut.

### 3.4. Diagram Alir



**BAB IV**  
**HASIL DAN PEMBAHASAN**

**4.1 Deskripsi Data**

Data yang digunakan pada penelitian ini merupakan data pribadi nasabah kredit pada Bank 'X' di Provinsi Lampung yang terdiri dari enam variabel prediktor. Terdiri dari 999 data nasabah yang diambil sebagai sampel, terdapat 199 (19,9%) nasabah termasuk ke dalam status kredit yang macet, dan 800 (80,1%) nasabah lainnya termasuk ke dalam status kredit yang lancar. Deskripsi untuk enam variabel prediktor yakni usia, lama kerja, pinjaman lain, pendapatan bersih, nominal akun, rasio hutang, untuk setiap kelas pada status kredit nasabah dapat dilihat pada Tabel 1.

**Tabel 1.** Deskripsi Variabel Usia, Lama Kerja, Pinjaman Lain, Pendapatan Bersih, Nominal Akun, dan Rasio Hutang

Variabel Prediktor	Rata-rata		Standar Deviasi	
	Kredit Lancar	Kredit Macet	Kredit Lancar	Kredit Macet
Usia	37,75	37,02	7,60	8,49
Lama Kerja	7,63	6,24	6,78	6,65
Pinjaman Lain	347.000	223.000	1.170.186,822	990.688,178
Pendapatan Bersih	5.160.000	4.560.000	5.888.368,369	4.839.785,008
Nominal Akun Tabungan	14.501.000	6.944.900	121.433.000	20.374.500
Rasio Hutang	30,552	31,825	11,067	14,786

Dari data pada Tabel 1, terlihat bahwa nilai rata-rata pendapatan bersih pada nasabah berstatus kredit lancar adalah sebesar Rp 5.160.000, nilai ini lebih besar dari rata-rata pendapatan bersih pada nasabah kredit macet sebesar Rp 4.560.000. Begitu pula dengan rasio hutang, dimana pada nasabah kredit lancar rasio hutangnya memiliki rata-rata 30,552. Nilai tersebut bernilai lebih kecil daripada rata-rata rasio hutang pada nasabah kredit macet sebesar 31,825.

## 4.2 Pembagian Data

Penerapan metode *k-Nearest Neighbor* (kNN) pada klasifikasi nasabah kredit dilakukan dengan terlebih dahulu membagi data menjadi data *training* dan data *testing*. Dari sejumlah 999 data, diambil 900 data (90,09%) sebagai data *training* dan 99 data (9,91%) digunakan sebagai data *testing*. Deskripsi data untuk data *training* dan data *testing* dapat dilihat pada Tabel 2 dan Tabel 3.

**Tabel 2.** Deskripsi Data Usia, Lama Kerja, Pinjaman Lain, Pendapatan Bersih, Nominal Akun dan Rasio Hutang pada Data *Training*

Variabel Prediktor	Rata-rata		Standar Deviasi	
	Kredit Lancar	Kredit Macet	Kredit Lancar	Kredit Macet
Usia	37,74	36,70	7,599	8,472
Lama Kerja	7,77	6,17	6,77	6,642
Pinjaman Lain	358.000	228.000	1.215.002,674	1.036.672,932
Pendapatan Bersih	5.180.000	4.720.000	6.023.282,017	5.076.876,648
Nominal Akun Tabungan	15.246.000	6.950.500	127.663.000	20.525.900
Rasio Hutang	30,572	31,076	10,889	14,435

**Tabel 3.** Deskripsi Data Usia, Lama Kerja, Pinjaman Lain, Pendapatan Bersih, Nominal Akun dan Rasio Hutang pada Data *Testing*

Variabel Prediktor	Rata-rata		Standar Deviasi	
	Kredit Lancar	Kredit Macet	Kredit Lancar	Kredit Macet
Usia	37,87	39,67	7,72	8,404
Lama Kerja	6,35	6,81	6,70	6,897
Pinjaman Lain	245.000	156.000	612.130,647	246.476,771
Pendapatan Bersih	5.020.000	3.180.000	4.477.960,266	1.248.659,542
Nominal Akun Tabungan	7.598.700	6.898.000	18.781.700	19.521.100
Rasio Hutang	30,367	38,178	12,675	16,522

Pada Tabel 2 dan Tabel 3 dapat dicermati kondisi nilai rata-rata untuk variabel pendapatan bersih pada nasabah kredit yang lancar lebih besar dibandingkan dengan nasabah kredit macet. Begitupula pada nilai rasio hutang pada nasabah kredit lancar bernilai lebih kecil dibandingkan dengan nasabah

kredit macet. Pada Lampiran 1 dan Lampiran 2 dapat dilihat tabel dari data *training* dan data *testing* yang digunakan dalam tugas akhir ini.

### 4.3 Penentuan Nilai Parameter Terbaik

Pada metode *k-Nearest Neighbor*, sebelum menghitung jarak Euclidian terdekat dari data yang akan dicari klasifikasi kelasnya, diperlukan proses untuk menentukan nilai parameter terbaik *k* terlebih dahulu. Dalam tugas akhir ini akan digunakan metode *hold-out*. Proses penentuan parameter terbaik ini menggunakan data *training* dan data *testing* dengan bantuan *software* SAS 9.0, yang *listing code* nya dapat dilihat pada Lampiran 3.

**Tabel 4.**Perbandingan Nilai *Error Rate* Hasil Klasifikasi Kelas Data *Training* dan Data *Testing*

k	Error rate	
	Data <i>Training</i>	Data <i>Testing</i>
1	0	0,313131
2	0,0907	0,363636
3	0,1655	0,393939
4	0,2341	0,515152
5	0,2805	0,565657
6	0,2749	0,434343
7	0,2848	0,414141
8	0,3025	0,464646
9	0,3209	0,484848
10	0,3317	0,515152

Hasil penentuan nilai parameter terbaik *k*, dapat dilihat dari nilai *error rate* terkecil hasil klasifikasi kelas pada kedua data tersebut. Tabel 4 menunjukkan perbandingan *error rate* pada kedua data tersebut dengan nilai  $k = 1, 2, \dots, 10$ .

Berdasarkan Tabel 4, diperoleh nilai *error rate* terkecil untuk data *training* dan *testing* ada pada nilai  $k = 1$ , yakni *error rate* untuk klasifikasi data *training* adalah sebesar 0, sedangkan *error rate* untuk klasifikasi data *testing* adalah

sebesar 0,313131. Maka dapat diambil kesimpulan bahwa nilai  $k$  terbaik yang digunakan untuk melakukan klasifikasi menggunakan metode *Bagging k-Nearest Neighbor* untuk data *testing* adalah nilai  $k = 1$ .

#### **4.4 Penentuan Klasifikasi Kelas untuk Observasi Pertama Data *Testing***

Pada metode ini akan digabungkan dua metode untuk mengklasifikasikan kelas dari data *testing*, yakni metode *Bagging* dan *k-Nearest Neighbor*. Sebelum diterapkan metode *k-Nearest Neighbor* akan dilakukan terlebih dahulu metode *Bagging*.

Pada metode *Bagging (Bootstrap Aggregating)* nantinya akan ada dua proses. Pertama, proses *bootstrap* yakni pengambilan sampel ulang disertai pengembalian pada data *training* sebanyak 50 kali dengan jumlah data pada masing-masing hasil *bootstrap* sebanyak 720 data (80%) menggunakan bantuan *software SAS 9.0*. Kedua, proses *aggregating* yaitu pendugaan kelas secara gabungan berdasarkan dugaan data *testing* menggunakan data hasil *bootstrap* yang telah ada dengan aturan *majority vote* (suara terbanyak).

##### **4.4.1 Perhitungan Jarak Euclidian antara Data *Testing* dengan Data *Bootstrap***

Proses *bootstrap* dilakukan dengan bantuan *software SAS 9.0*, yang *listing code* nya dapat dilihat pada Lampiran 4, dimana pengambilan sampel acak disertai dengan pengembalian, sehingga data yang sudah terambil mempunyai kemungkinan untuk kembali terambil menjadi data *bootstrap*. Data *training* yang di-*resampling* sebanyak 50 kali menjadi data *bootstrap* dengan ukuran sampelnya 720 data.

Sebelum ditentukan kelas untuk data *testing* observasi pertama, data *bootstrap* inilah yang akan dijadikan sebagai data *training* yang akan dihitung *majority vote* nya untuk penentuan kelas tersebut. Pada Tabel 5 dapat dilihat data hasil *bootstrap* pertama.

**Tabel 5.** Data *Bootstrap* Pertama

No.	Usia	Lama kerja	Pinjaman lain	Pendapatan bersih	Nominal akun	Rasio hutang	Status kredit
1	40	7	0	3.920.069	164.824,2	24,027	Lancar
2	45	14	0	2.222.741	3.600.000	47,901	Macet
3	30	2	0	11.601.567	918.311,4	21,988	Lancar
4	47	28	0	3.809.983	10.261.638	39,554	Lancar
5	41	1	1.070.693	6.140.795	540.000	43,349	Lancar
.	.	.	.	.	.	.	.
716	36	2	456.303	5.913.697	900.000	36,736	Lancar
717	36	2	456.303	5.913.697	900.000	36,736	Lancar
718	40	15	913.296	4.129.724	14.098.273	55,464	Lancar
719	40	15	913.296	4.129.724	14.098.273	55,464	Lancar
720	40	15	913.296	4.129.724	14.098.273	55,464	Lancar

Pada Tabel 5 dapat dilihat karakteristik dari data *bootstrap* dimana terdapat sampel yang muncul lebih dari sekali dikarenakan *bootstrap* merupakan metode *resampling with replacement* (pengambilan sampel dengan pengembalian), yang tiap sampelnya memiliki peluang untuk terambil kembali walaupun sudah terpilih sebelumnya.

Pada Tabel 6 terdapat data observasi pertama dari data *testing* yang akan diklasifikasikan kelasnya berdasarkan jaraknya dengan data hasil *bootstrap*.

**Tabel 6.** Observasi Pertama, pada Data *Testing*.

Usia	Lama Kerja	Pinjaman Lain	Pendapatan Bersih	Nominal Akun	Rasio Hutang
32	4	461.221	2.293.779	6.419.734	46,2

Selanjutnya akan dihitung jarak Euclidian dari data *bootstrap* pertama ke data *testing* untuk observasi pertama, dengan perhitungan sebagai berikut ini:

- a. Berikut ini akan dihitung jarak dari observasi pertama data *bootstrap* pertama dengan data *testing* observasi pertama, dimana dilambangkan dengan  $d(\mathbf{x}_{11}, \mathbf{y}_1)$

$$\begin{aligned} d(\mathbf{x}_{11}, \mathbf{y}_1) &= \sqrt{(40 - 32)^2 + (7 - 4)^2 + (0 - 461.221)^2 + (3.920.069 - 2.293.779)^2 +} \\ &\quad (164.824,2 - 6.419.734)^2 + (24,027 - 46,2)^2} \\ &= 6.479.309 \end{aligned}$$

- b. Berikut ini akan dihitung jarak dari observasi kedua data *bootstrap* pertama dengan data *testing* observasi pertama, dimana dilambangkan dengan  $d(\mathbf{x}_{12}, \mathbf{y}_1)$

$$\begin{aligned} d(\mathbf{x}_{12}, \mathbf{y}_1) &= \sqrt{(45 - 32)^2 + (14 - 4)^2 + (0 - 461.221)^2 + (2.222.741 - 2.293.779)^2 +} \\ &\quad (3.600.000 - 6.419.734)^2 + (47,901 - 46,2)^2} \\ &= 2.858.089 \end{aligned}$$

- c. Berikut ini akan dihitung jarak dari observasi ketiga data *bootstrap* pertama dengan data *testing* observasi pertama, dimana dilambangkan dengan  $d(\mathbf{x}_{13}, \mathbf{y}_1)$

$$\begin{aligned} d(\mathbf{x}_{13}, \mathbf{y}_1) &= \sqrt{(30 - 32)^2 + (2 - 4)^2 + (0 - 461.221)^2 + (11.601.567 - 2.293.779)^2 +} \\ &\quad (918.311,4 - 6.419.734)^2 + (21,988 - 46,2)^2} \\ &= 10.821.890 \end{aligned}$$

- d. Berikut ini akan dihitung jarak dari observasi keempat data *bootstrap* pertama dengan data *testing* observasi pertama, dimana dilambangkan dengan  $d(\mathbf{x}_{14}, \mathbf{y}_1)$

$$\begin{aligned} d(\mathbf{x}_{14}, \mathbf{y}_1) &= \sqrt{(47 - 32)^2 + (28 - 4)^2 + (0 - 461.221)^2 + (3.809.983 - 2.293.779)^2 +} \\ &\quad (10.261.638,4 - 6.419.734)^2 + (39,554 - 46,2)^2} \\ &= 4.155.938 \end{aligned}$$

e. Berikut ini akan dihitung jarak dari observasi kelima data *bootstrap* pertama dengan data *testing* observasi pertama, dimana dilambangkan dengan  $d(x_{15}, y_1)$

$$d(x_{15}, y_1) = \sqrt{(41 - 32)^2 + (1 - 4)^2 + (1.070.693 - 461.221)^2 + (6.140.795 - 2.293.779)^2 + (540.000 - 6.419.734)^2 + (43,349 - 46,2)^2}$$

$$= 7.052.819$$

Langkah menghitung jarak Euclidian data *bootstrap* dengan data *testing* ini dilakukan hingga observasi ke 720 pada data *bootstrap* pertama. Sehingga didapati hasil perhitungan jarak Euclidian semua observasi seperti yang dapat dilihat pada Tabel 7.

**Tabel 7.** Jarak Euclidian data *Bootstrap* pertama ke data *testing*

No.	Usia	Lama kerja	Pinjaman lain	Pendapatan bersih	Nominal akun	Rasio hutang	Status kredit	Jarak
1	40	7	0	3.920.069	164.824,2	24,027	Lancar	6.479.309
2	45	14	0	2.222.741	3.600.000	47,901	Macet	2.858.089
3	30	2	0	11.601.567	918.311,4	21,988	Lancar	10.821.890
.	.	.	.	.	.	.	.	.
718	40	15	913.296	4.129.724	14.098.273	55,464	Lancar	7.907.909
719	40	15	913.296	4.129.724	14.098.273	55,464	Lancar	7.907.909
720	40	15	913.296	4.129.724	14.098.273	55,464	Lancar	7.907.909

#### 4.4.2 Penentuan Tetangga Terdekat Sebanyak $k$ pada Setiap data *Bootstrap*

Setelah melakukan perhitungan jarak Euclidian antar data, akan ditentukan tetangga terdekat sebanyak  $k$ , dimana  $k = 1$  adalah nilai terbaik parameter  $k$ , untuk memutuskan kelas dari observasi pertama data *testing* berdasarkan data hasil *bootstrap*. Namun sebelumnya, langkah yang harus dilakukan adalah mengurutkan nilai jarak antar data tersebut dari yang terkecil hingga yang terbesar.

Pada Tabel 8 dapat dilihat urutan jarak antar data dari yang terkecil ke yang terbesar pada data *bootstrap* pertama.

**Tabel 8.** Urutan Jarak dari Nilai yang Terkecil Hingga Terbesar untuk Data *Bootstrap* Pertama

No.	Usia	Lama kerja	Pinjaman lain	Pendapatan bersih	Nominal akun	Rasio hutang	Status Kredit	Jarak
715	38	9	408.412	2.173.496	6.391.048	23,647	Lancar	134.460,8
92	50	6	535.991	2.630.509	6.480.000	22,876	Lancar	350.156,6
288	32	1	0	2.581.924	6.215.350	29,853	Lancar	580.969,4
82	44	17	0	2.581.747	6.068.012	29,186	Lancar	647.579
477	43	13	0	2.687.930	6.819.995	19,117	Lancar	726.834,5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
384	53	17	493.090	3.843.320	1,77E+08	30,015	Macet	1,7E+08
294	35	10	0	4.328.601	2,04E+08	34,815	Lancar	1,97E+08
351	34	2	1.188.659	5.756.717	9,94E+08	30,676	Lancar	9,87E+08
352	34	2	1.188.659	5.756.717	9,94E+08	30,676	Lancar	9,87E+08
472	45	3	2.360.289	45.639.712	2,88E+09	13,339	Lancar	2,87E+09

Pada Tabel 8 tersebut, dapat dilihat untuk nilai  $k=1$ , jarak terdekat pertama dari data *bootstrap* pertama ke data *testing* adalah data observasi ke-715, dengan nilai jarak = 134.460,8. Sehingga prediksi kelas menggunakan data *bootstrap* pertama untuk observasi pertama data *testing* masuk ke dalam kelas kredit Lancar, sesuai dengan kelas dari data tetangga terdekat pertama observasi ke-715.

#### 4.4.3 Penentuan Kelas dari Observasi Pertama Data *Testing* Berdasarkan Hasil *Majority Vote*

Berdasarkan hasil dari penentuan tetangga terdekat di langkah sebelumnya, telah didapati 50 prediksi kelas untuk observasi pertama data *testing* dari perhitungan jarak dengan data hasil *bootstrap*. Langkah terakhir adalah *Aggregating*, dilakukan untuk menentukan kelas dari observasi pertama adalah dengan menghitung suara

terbanyak (*majority vote*) kelas yang muncul dari total 50 hasil prediksi kelas sebelumnya. Berikut ini dapat dilihat pada Tabel 9, hasil prediksi kelas status kredit dari 50 data hasil perhitungan jarak dengan data *bootstrap*.

**Tabel 9.** Hasil Prediksi Kelas untuk Status Kredit Berdasarkan Jarak dengan Data *Bootstrap*.

n	Hasil Prediksi berdasarkan jarak dengan Data Hasil <i>Bootstrap</i> ke-																				L	M	Majority Vote		
	1	2	3	4	5	6	7	8	9	10	...	40	41	42	43	44	45	46	47	48				49	50
1	L	L	L	L	L	L	L	L	L	L		L	L	L	L	L	L	L	L	L	L	L	50	0	L

Keterangan: L= Status Kredit Lancar, M = Status Kredit Macet

Setelah melihat hasil prediksi dari 50 data *bootstrap*, langkah selanjutnya adalah menghitung suara terbanyak dari kelas yang muncul tersebut yang telah direkap juga pada Tabel 9.

Hasilnya, untuk observasi pertama data *testing* diklasifikasikan ke dalam status kredit Lancar, karena hasil perhitungan kelas Lancar muncul lebih banyak yakni sebanyak 50 kali, dibandingkan dengan kelas Macet yang tidak muncul sama sekali atau nol.

Lakukan juga secara berulang untuk observasi kedua hingga terakhir pada data *testing*, yakni menentukan kelas dari observasi berdasarkan suara terbanyak dari kelas yang muncul pada hasil prediksi kelas status kredit berdasarkan data *bootstrap* tersebut.

#### 4.5 Accuracy (Ketepatan) Hasil Prediksi Gabungan Menggunakan *Bagging k-Nearest Neighbor*

Setelah selesai menentukan kelas untuk semua observasi data *testing*, langkah yang terakhir adalah menghitung ketepatan klasifikasi kelas status kredit

tersebut. Cara menghitungnya adalah dengan menghitung *error rate* terlebih dahulu seperti pada Persamaan (1).

Pada Tabel 9 telah disajikan data hasil penentuan kelas berdasarkan suara terbanyak untuk 99 observasi data *testing*. Dari 99 hasil prediksi kelas status kredit data *testing* tersebut selanjutnya akan dihitung *error rate* klasifikasi. Setelah di hitung terdapat 33 hasil prediksi yang berbeda dengan data asli kelas dari status kredit pada data *testing*, dan terdapat 66 data yang sesuai dengan data asli kelas dari status kredit tersebut. Hasil prediksi dari kelas status nasabah kredit selengkapnya dapat dilihat pada Lampiran 5.

**Tabel 10.** Hasil *Majority Vote* untuk Penentuan Kelas Observasi Pertama Data *Testing*

n	S t a t	Hasil Prediksi berdasarkan jarak dengan Data Hasil <i>Bootstrap</i>																				L	M	Majority Vote	
		1	2	3	4	5	6	7	8	9	10	...	41	42	43	44	45	46	47	48	49				50
1	L	L	L	L	L	L	L	L	L	L	L	...	L	L	L	L	L	L	L	L	L	L	50	0	Lancar
2	L	L	L	M	L	L	L	L	L	L	L	...	L	L	L	L	L	M	L	L	L	L	45	5	Lancar
3	L	L	L	L	L	L	L	L	L	L	L	...	M	L	L	L	L	L	L	L	L	L	45	5	Lancar
4	M	L	L	L	L	L	L	L	L	L	L	...	L	M	L	M	L	L	L	L	L	L	47	3	Lancar
5	L	M	M	M	M	M	M	M	M	M	L	...	L	M	M	M	L	M	M	M	M	L	12	3	Macet
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
95	L	L	L	L	M	L	L	L	L	L	L	...	L	L	L	L	L	L	L	L	L	L	47	2	Lancar
96	L	L	L	L	L	L	L	L	L	L	L	...	L	L	L	L	L	L	L	L	L	L	49	0	Lancar
97	M	L	L	L	L	L	L	L	L	L	L	...	L	L	L	L	L	L	L	L	L	L	49	0	Lancar
98	L	M	M	L	M	M	M	M	M	M	M	...	M	M	M	L	M	M	M	M	M	M	3	4	Macet
99	L	L	L	L	L	L	L	L	L	L	L	...	L	L	L	L	L	L	L	L	L	L	49	0	Lancar

Keterangan: L= Status Kredit Lancar, M = Status Kredit Macet

Selanjutnya cara menghitung *error rate* nya adalah sebagai berikut:

$$Error Rate = \frac{\text{Jumlah prediksi kelas yang salah}}{\text{jumlah data seluruhnya}} = \frac{33}{99} = 0,333333$$

Dengan diketahuinya *error rate* akan dapat dihitung nilai *accuracy* (ketepatan), yakni:

$$Accuracy = 1 - Error Rate = 1 - 0,333333 = 0,666667$$

Dengan nilai *accuracy* 0,666667 berarti ketepatan hasil klasifikasi dengan *Bagging k-Nearest Neighbor* adalah sebesar 0,666667 atau 66,7%

## **BAB V**

### **KESIMPULAN**

Berdasarkan hasil dari pembahasan, diperoleh kesimpulan bahwa metode *Bagging k-Nearest Neighbor* dapat digunakan untuk mengklasifikasi status nasabah kredit lancar dan macet pada Bank 'X' di provinsi Lampung. Hasil pengklasifikasian tersebut, selanjutnya digunakan untuk memprediksi calon nasabah yang akan menggunakan layanan jasa kredit. Dalam penelitian ini, diperoleh hasil ketepatan klasifikasi nasabah kredit dengan metode *Bagging k-Nearest Neighbor* sebesar 66,67%.

## DAFTAR PUSTAKA

- Breiman, L. 1996. *Bagging Predictors*. *Machine Learning* 24 123-140.
- Bishop, C. M. 1995. *Neural Networks For Pattern Recognition*. Clarendon Press: Oxford
- Chapra, M. U. 2000. *Sistem Moneter Islam*, terj. Ikhwan Abidin Basri, *Towards a Just Monetary System*. Jakarta: Gema Insani.
- Gorunescu, F. 2011. *Data Mining: Concepts, Model and Techniques*, Prof. Janusz Kacprzyk and Prof. Lakhmi C. Jain, Eds. Berlin. Jerman: Springer
- Han, J. dan Kamber, M. 2006. *Data Mining: Concept and Techniques*. San Frasisco: Morgan Kaufmann Publishers.
- Hermawati, F. A. 2013. *Data Mining*. Yogyakarta: ANDI.
- Herprasetyo, B. 2012. *Sukses Ubah Kartu Kredit jadi Modal Usaha*. Tulungagung: Adora media
- Ismail. 2010. *Manajemen Perbankan*. Surabaya: Kencana
- Kasmir. 2001. *Manajemen Perbankan*. Jakarta: Raja Grafindo Persada
- Kasmir. 2002. *Bank dan Lembaga Keuangan Lainnya. Edisi Revisi 2002*. Jakarta: PT. Rajagrafindo Persada.
- Kuncoro, M dan Suhardjono. 2002. *Manajemen Perbankan: Teori Dan Aplikasi*. Jogjakarta: BPFE
- Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- Mahmoedin. 2002. *Melacak Kredit Bermasalah*. Jakarta: Pustaka Sinar Harapan.
- Mester, L. J. 1997. *What's The Point of Credit Scoring? Business Review (September)* 3-16.

- Melayu, H. 2004. *Dasar-dasar Perbankan*. Jakarta: CV. Haji Masagung.
- Morrison, J. 2004. *Introduction to Survival Analysis in Business, The Journal of Business Forecasting Methods & Systems* 23 (1) 18-22.
- Prasetyo, E. 2012. *DATA MINING – Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: ANDI.
- Prasetyo, E. 2014. *DATA MINING: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI.
- Tibshirani, R. J dan Efron, B. 1993. *An Introduction to the Bootstrap*. London: Chapman dan Hall.
- Republik Indonesia. 1998. *Undang-Undang No.10 1998 tentang Perbankan*. \_\_\_\_\_ . Jakarta.
- Rivai, V dan Permata, A. 2007. *Credit Management Handbook*. Jakarta: PT.Elex Media Komputindo.
- Sumathi, dan Sivanandam, S. N. 2006. *Introduction to Data Mining and its Application*. Berlin Heidelberg. New York: Springer.
- Witten, I. H, Frank E, Hall, M. A. 2011. *Data Mining : Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher.
- Wu, X. dan Kumar, V. 2009. *The Top Ten Algorithms in Data Mining*. London: CRC Press Taylor & Analitic Machine Intelligence 13: 841-847.

## LAMPIRAN

Lampiran 1.Data Training, n = 900

Debitur	Usia	Lama Kerja	PinjamanLain	PendapatanBersih	Nominal Akun	Rasio Hutang	Status
Debitur371	40	7	0	3920069	164824,20	24,027	Lancar
Debitur1622	49	7	316045	2042409	39600000,00	39,111	Macet
Debitur2335	45	14	0	2222741	3600000,00	47,901	Macet
Debitur535	46	17	0	7902600	78203649,60	27,294	Lancar
Debitur163	30	2	0	11601567	918311,40	21,988	Lancar
Debitur520	47	28	0	3809983	10261638,00	39,554	Lancar
Debitur3050	25	2	0	2712000	180000,00	29,306	Macet
Debitur97	41	1	1070693	6140795	540000,00	43,349	Lancar
Debitur666	51	8	0	6677132	77540094,00	7,485	Lancar
Debitur2756	26	2	0	2054607	3576340,80	22,66	Macet
Debitur94	41	14	0	4282100	6072904,80	26,395	Lancar
Debitur493	40	1	0	0	15660000,00	17,318	Lancar
Debitur340	52	27	0	0	5363096,40	17,129	Lancar
Debitur329	40	15	0	2609000	90000,00	39,391	Lancar
Debitur650	36	5	1408000	4977000	90000,00	39,753	Lancar
Debitur313	27	6	0	3827397	6594397,20	19,687	Lancar
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
Debitur557	35	16	297084	3045884	241614,00	37,062	Lancar
Debitur1913	38	1	0	5764060	9540000,00	130,724	Macet
Debitur169	26	6	402778	1989021	949879,80	38,324	Lancar
Debitur1403	51	12	0	0	360000,00	26,279	Macet
Debitur1749	40	8	0	2808545	810000,00	52,903	Macet
Debitur133	26	2	0	5192867	180000,00	13,179	Lancar
Debitur687	30	6	600000	1744330	138007,80	47,513	Lancar
Debitur688	39	5	0	2777350	1,80	33,913	Lancar
Debitur824	33	9	0	2960575	9000000,00	31,814	Lancar
Debitur659	38	9	408412	2173496	6391047,60	23,647	Lancar
Debitur455	36	2	456303	5913697	900000,00	36,736	Lancar
Debitur76	40	15	913296	4129724	14098273,20	55,464	Lancar
Debitur518	31	8	0	2857109	2532598,20	32,966	Lancar
Debitur730	52	24	0	5203000	83225905,20	31,136	Lancar

**Lampiran 2.**Data Testing, n = 99

Debitur	Usia	Lama kerja	Pinjaman Lain	Pendapatan Bersih	Nominal Akun	Rasio Hutang	status
Debitur299	32	4	461221	2293779	6419734	46,2	Lancar
Debitur542	53	2	0	3159337	1,8	40,662	Lancar
Debitur476	44	21	0	2594667	180000	59,413	Lancar
Debitur336	39	3	261000	4666000	90000	40,432	Macet
Debitur63	33	4	466737	10734847	90000	28,318	Lancar
Debitur131	35	3	3375636	19867926	1,03E+08	37,801	Lancar
Debitur755	35	3	0	2197620	2250000	28,059	Lancar
Debitur1789	40	1	0	5000000	17156315	36,909	Macet
Debitur398	50	2	0	17593582	1,29E+08	58,414	Lancar
Debitur827	30	2	0	2873500	9000000	17,883	Macet
Debitur2563	40	2	110952	5669048	1800000	18,41	Macet
Debitur67	36	13	0	2544708	127526,4	48,675	Lancar
Debitur854	40	15	83333	3023064	180000	30,103	Macet
Debitur789	34	1	0	2662300	10016581	13,511	Lancar
Debitur50	27	4	0	2529350	180000	53,479	Lancar
Debitur727	44	16	0	6752159	9033156	29,934	Lancar
Debitur1771	45	2	0	2827025	540000	54,53	Macet
Debitur3822	29	5	541000	3379000	7156420	24,113	Macet
...	...	...	...	...	...	...	...
Debitur507	43	11	0	3088000	146340	24,401	Lancar
Debitur587	42	4	50000	3310000	2250353	41,667	Lancar
Debitur1601	25	1	425000	3780620	1630091	42,206	Macet
Debitur222	40	6	663705	2445565	2863098	39,37	Macet
Debitur134	35	5	975529	8332471	15033298	39,545	Lancar
Debitur443	29	4	0	3167413	1800000	8,245	Lancar
Debitur41	30	1	0	6000000	6266506	18,838	Lancar
Debitur475	53	3	0	3434698	180000	59,843	Lancar
Debitur107	50	22	0	3768304	3234233	34,993	Lancar
Debitur646	28	2	181732	3611225	5400000	27,637	Lancar
Debitur567	41	14	0	4285000	75243,6	37,881	Macet
Debitur722	29	1	0	5774046	309020,4	18,705	Lancar
Debitur110	32	2	0	3128900	5718600	30,102	Lancar

**Lampiran 3.** *Syntax cross validation* di SAS

```
Proc discrim data= train
method =npar k=1 pool=yes
metric= full testdata= test testout= testout;
class status;
var Usia Lama_kerja Pinjaman_lain Pendapatan_bersih Nominal_akun
Rasio_hutang ;
run;

proc print data=testout;
run;
```

**Lampiran 4.** *Syntax bootstrap* di SAS

```
Proc survey select data=train out=outboot
    seed=30459584
    method=urs
    samprate=80
    outhits
    rep=50;
run;

ods listing close;
proc univariate data=outboot;
var x;
by Replicate;
output out=outall kurtosis=curt;
run;
ods listing;

proc univariate data=outall;
var curt;
output out=final pctlpts=2,5, 97,5 pctlpre=ci;
run;

proc print data=outboot;
run;
```

**Lampiran 5.** Hasil klasifikasi kelas lengkap berdasarkan data *bootstrap* menggunakan *majority vote*  
Kode Status Kredit, 0= Lancar, 1=Macet

	n	Hasil Prediksi berdasarkan jarak dengan Data Hasil Bootstrap																																																		Lancar	Macet	Majority Vote					
		Status Asli	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49				50				
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0
2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	5	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	5	0
4	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	3	0	
5	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	12	38	1		
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0			
7	0	0	0	1	0	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	16	0		
8	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	5	0			
9	0	0	0	1	0	1	0	0	1	1	1	0	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	30	1		
10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0			
11	1	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	21	0		
12	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	1	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	30	1		
13	1	1	1	0	1	0	1	1	1	0	1	0	1	0	1	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	28	1	







68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	49	0	0						
69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	49	0	0						
70	1	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	29	29	20	0				
71	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	49	0	0				
72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	49	0	0				
73	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	45	4	0			
74	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0	1	1	0	1	1	1	1	1	1	1	0	0	1	0	1	1	1	14	14	35	1	
75	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43	43	6	0			
76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	44	44	5	0		
77	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	13	13	36	1		
78	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	49	0	0		
79	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	36	13	0	
80	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	38	11	0
81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	49	0	0	
82	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	6	6	43	1			
83	0	0	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	15	34	1
84	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	45	4	0
85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

