

17. 453/H/03



TUGAS AKHIR

**STUDI KLASIFIKASI PARTISIPASI USIA KERJA
DALAM ANGKATAN KERJA DI KABUPATEN KEDIRI
DENGAN PENDEKATAN METODE REGRESI LOGISTIK,
KERNEL DAN ARTIFICIAL NEURAL NETWORK**

Disusun Oleh :

MONITA VERA SANTY
1398 100 018

RSST
519.536
San
S-1
2002



PAPERTERAKAN	
Tgl. Terima	14/01/2003
Tujuan	H
No. Agenda Frp.	21.4903/03

**JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2002**

LEMBAR PENGESAHAN
TUGAS AKHIR

**STUDI KLASIFIKASI PARTISIPASI USIA KERJA
DALAM ANGKATAN KERJA DI KABUPATEN KEDIRI
DENGAN PENDEKATAN METODE REGRESI LOGISTIK,
KERNEL DAN ARTIFICIAL NEURAL NETWORK**

Disusun Oleh :

MONITA VERA SANTY
1398 100 018

Diajukan sebagai syarat untuk kelulusan Program Sarjana
Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember Surabaya

Surabaya, Nopember 2002

Menyetujui,
Pembimbing Tugas Akhir

Bambang Widjanarko Otok, S.Si, MSi.
NIP. 132 125 684

Mengetahui,
Ketua Jurusan Statistika

Drs. Nur Iriawan, M.Ikom, Ph.D
NIP. 131 782 011

Kupersembahkan Tugas Akhir ini kepada :

*Mama, Papa, serta kedua adikku Mega dan Ayu,
Atas segala curahan cinta, kasih sayang dan
perhatiannya untukku.....*

Serta mereka.....

*Orang - orang yang aku kasihi dan yang
mencintaiku*

ABSTRAK

Negara-negara berkembang mempunyai distribusi umur yang muda. Tingginya tingkat kelahiran kasar menyebabkan persentase penduduk berusia dibawah 15 tahun jauh lebih tinggi. Proses transisi demografi yang berkaitan dengan meningkatnya usia harapan hidup penduduk akan merubah distribusi umur dari muda menjadi tua. Proses ini terjadi berkat adanya perbaikan dibidang kesehatan bersamaan dengan meningkatnya teknologi kedokteran dan meningkatnya kesadaran masyarakat (Suwartapradja, 2000). Hal tersebut diatas memicu meningkatnya jumlah pengangguran disuatu negara apabila tidak diimbangi dengan lapangan kerja yang memadai sesuai dengan tingkat pendidikan yang dimiliki.

Dengan memperhatikan kondisi diatas, maka sangat menarik untuk diteliti serta dikaji lebih lanjut mengenai partisipasi seseorang dalam angkatan kerja dengan menggunakan metode klasifikasi yang tidak menggunakan asumsi data harus berdistribusi normal. Hal ini dilakukan karena banyak kasus pengklasifikasian yang datanya tidak normal. Metode yang diterapkan adalah regresi logistik, kernel dan *Artificial Neural Network* (ANN), dengan pembagian dua data yaitu data training dan data testing untuk analisanya. Data yang digunakan adalah data hasil Susenas 2000 untuk daerah kabupaten Kediri.

Hasil analisa data dari ketiga metode terlihat bahwa, untuk data training metode kernel dengan asumsi nilai varians sama memiliki ketepatan klasifikasi yang lebih baik dibandingkan dengan regresi logistik maupun ANN menurut kriteria *MCC* dan *PCC*. Sedangkan untuk data testing ternyata Metode ANN memberikan hasil yang lebih tepat dalam pengklasifikasian dibandingkan dengan logistik maupun kernel. Hal ini didukung pula oleh uji kestabilan model yang menggunakan *Press-Q*.

KATA PENGANTAR

Alhamdulillah, puji syukur kami panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga atas karunianya pula penulis dapat menyelesaikan tugas akhir ini dengan judul **“Studi Klasifikasi Usia Kerja dalam Angkatan Kerja di Kabupaten Kediri dengan Pendekatan Metode Regresi Logistik, Kernel dan Artificial Neural Network (ANN) “**

Tugas akhir tersebut merupakan salah satu prasyarat untuk memperoleh gelar sarjana di jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember Surabaya.

Penulis mengharapkan tugas akhir ini dapat memberi sumbangan yang bermanfaat terhadap pemikiran dan pengembangan ilmu pengetahuan pada umumnya dan khususnya bermanfaat bagi penulis sendiri.

Penulis menyadari sepenuhnya, bahwa tugas akhir ini masih jauh dari sempurna. Kritik dan saran dari pembaca sangat diharapkan sebagai masukan yang berharga untuk penelitian selanjutnya.

Surabaya, Nopember 2002

Penulis

UCAPAN TERIMA KASIH

Di dalam proses penggeraan tugas akhir ini penulis menyadari tidak sedikit bantuan dan dorongan yang diberikan baik secara material maupun spiritual sehingga tugas akhir ini dapat terselesaikan dengan baik. Oleh sebab itu dengan segala kerendahan hati penulis menyampaikan ucapan terima kasih kepada :

1. Ir. Nur Iriawan, M.Ikom, Ph.D selaku Ketua Jurusan Statistika
2. Bambang Widjarnako Otok, Ssi, Msi, selaku Dosen Pembimbing Tugas Akhir.
3. Drs. I. Nyoman Latra, MS selaku Dosen Wali dan Dosen Penguji Tugas Akhir.
4. Ir. Sri Pingit Wulandari, Msi selaku Dosen Penguji Tugas Akhir.
5. Bp. Hartono, atas segala masukan yang diberikan.
6. Mama, Papa atas segala lautan cinta dan kasih sayangnya padaku.
7. Pak Warno, Mbak Santy atas segala bantuan dan senyum yang diberikan waktu aku lagi bingung ria.
8. Teman seperjuangan Intan dan Ratna, akhirnya kita bisa make' toga bareng-bareng.
9. My big Brothers mas Hendy dan mas Wima yang udah banyak banget bantu aku buatin program dan masukan.
10. Mbak Wiwin atas pinjaman komputernya, sorry aku sering nganggu kegiatanmu.
11. Mbak Ida, Mbak Ana, Linda, Laila, linux, Beti, Mbak Narmi, Mbak Widji, Mbak Ika, Iis dan temen kost lainnya atas gangguan pas aku lagi pusing ria.

12. Sie Baby yang selalu tanpa bosen-bosennya ngasih dukungan, bantuan serta perhatiannya padaku, walaupun aku lagi marah atawa pas aku jutek banget.
Makasih ya semua perhatiannya.
13. Semua temen-temen angkatan '98 yang selalu memberikan senyumnya dan kebersamaan yang kita ukir bareng-bareng di Statistika.
14. Mas Beki atas pinjaman TA-nya.
15. Serta semua pihak yang telah membantu yang tidak bisa penulis sebutkan namanya satu persatu, makasih atas semua bantuannya.

DAFTAR ISI

Halaman

HALAMAN JUDUL

Lembar Pengesahan	
Abstrak	i
Kata Pengantar	ii
Ucapan Terima Kasih.....	iii
Daftar Isi	v
Daftar Tabel	ix
Daftar Gambar	xii
Daftar Lampiran.....	xii

BAB I PENDAHULUAN

1.1. Latar Belakang	1
1.2. Perumusan Masalah	2
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.5. Batasan Masalah	3

BAB II TINJAUAN PUSTAKA

2.1. Tinjauan Kependudukan.....	4
2.1.1. Teori Ketenagakerjaan	4
2.1.2. Faktor-Faktor yang Menjadi Latar Belakang Usia Produktif Bekerja.....	6
2.2. Metode Regresi Logistik	7

2.2.1. Model Regresi Logistik	8
2.2.2. Pembentukan Model Logistik dan Estimasi Parameter	9
2.2.3. Pengujian Kelayakan Model	13
2.2.4. Interpretasi Model	14
2.3. Metode Kernel	15
2.3.1. Taksiran Mean dan Varians	17
2.3.2. Bias	18
2.3.3. Pemilihan Bandwidth	20
2.3.4. Pemilihan Fungsi Kernel	22
2.4. Metode Artificial Neural Network	23
2.4.1. Feed Forward Artificial Neural Network	24
2.4.2. Metode Optimasi Error	31
2.5. Perbandingan Hasil Klasifikasi Metode Regresi Logistik, Kernel dan Artifisial Neural Network.....	33

BAB III METODOLOGI PENELITIAN

3.1. Sumber Data	36
3.2. Variabel Penelitian dan Definisi Operasional	37
3.3. Pengolahan Data	42

BAB IV ANALISA DAN PEMBAHASAN

4.1. Deskriptif Statistik	46
4.2. Metode Regresi Logistik	51
4.2.1. Metode Regresi Logistik Data Training	51

4.2.1.1. Pemodelan Univariat	51
4.2.1.2. Pemodelan Multivariat.....	53
4.2.1.3. Ketepatan Klasifikasi	56
4.2.1.4. Uji Keakuratan	57
4.2.2. Metode Regresi Logistik Data Testing	58
4.2.2.1. Ketepatan Klasifikasi	58
4.2.2.2. Uji Keakuratan dan Kestabilan	58
4.3. Metode Kernel	60
4.3.1. Pemilihan Fungsi Kernel dan Bandwith Optimal	60
4.3.2. Analisa Data Training Metode Kernel	61
4.3.2.1. Ketepatan Klasifikasi Asumsi Varians Sama	61
4.3.2.2. Uji Keakuratan Asumsi Varians Sama	62
4.3.2.3. Ketepatan Klasifikasi Asumsi Varians Tidak Sama	63
4.3.2.4. Uji Keakuratan Asumsi Varians Tidak Sama	63
4.3.3. Metode Kernel Data Testing	64
4.3.3.1. Ketepatan Klasifikasi Asumsi Varians Sama	64
4.3.3.2. Uji Keakuratan dan Kestabilan Asumsi	
Varians Sama	65
4.3.3.3. Ketepatan Klasifikasi Asumsi Varians Tidak Sama	66
4.3.3.4. Uji Keakuratan dan Kestabilan Model Asumsi	
Varians Tidak Sama	67
4.4. Metode Artificial Neural Nework	68
4.4.1. Metode ANN Data Training	69
4.4.1.1. Ketepatan Klasifikasi	69

4.4.1.2. Uji Keakuratan	70
4.4.2. Metode ANN Data Testing	71
4.4.2.1. Ketepatan Klasifikasi	71
4.4.2.2. Uji Keakuratan dan Kestabilan	72
4.5. Perbandingan Hasil Klasifikasi Metode Regresi Logistik, Kernel dan ANN	73
4.5.1 Data Training	73
4.5.2 Data Testing	74

BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan	75
5.2. Saran	76
Daftar Pustaka	77
Lampiran	79

DAFTAR TABEL

	Halaman
Tabel 2.1 Nilai Model Regresi Logistik	15
Tabel 2.2 Macam-Macam Fungsi Kernel	23
Tabel 3.1 Ringkasan Variabel Respon dan Prediktor	41
Tabel 4.1 Status Usia Kerja menurut Hubungan dengan Kepala Rumah Tangga	46
Tabel 4.2 Status Usia Kerja menurut Jenis Kelamin	47
Tabel 4.3 Status Usia Kerja menurut Kelompok Umur	48
Tabel 4.4 Status Usia Kerja menurut Status Perkawinan	48
Tabel 4.5 Status Usia Kerja menurut Status Kesehatan	48
Tabel 4.6 Status Usia Kerja menurut Tempat Tinggal	48
Tabel 4.7 Status Usia Kerja menurut tingkat pendidikan	49
Tabel 4.8 Status Usia Kerja menurut Jumlah Anggota Rumah Tangga	49
Tabel 4.9 Status Usia Kerja menurut Jumlah Anggota Rumah Tangga yang Bekerja	50
Tabel 4.10 Status Usia Kerja menurut Pengeluaran Perkapita Sebulan	50
Tabel 4.11 Hasil Metode Regresi Logistik Univariat	52
Tabel 4.12 Hasil Metode Regresi Logistik Multivariat	54
Tabel 4.13 Hasil Pemodelan Metode regresi logistik multivariat	54
Tabel 4.14 Hasil Akhir Metode Regresi Logistik Multivariate	55
Tabel 4.15 Hasil Klasifikasi Metode Regresi Logistik Data Training	57
Tabel 4.16 Hasil Klasifikasi Metode Regresi Logistik Data Testing.....	58

Tabel 4.17 Hasil Klasifikasi Metode Kernel Asumsi Varians Sama Data Training	62
Tabel 4.18 Hasil Klasifikasi Metode Kernel Asumsi Varians Tidak Sama Data Training	63
Tabel 4.19 Hasil Klasifikasi Metode Kernel Asumsi Varians Sama Data Testing	65
Tabel 4.20 Hasil klasifikasi Metode Kernel Asumsi Varians Tidak Sama Data Testing	67
Tabel 4.21 Hasil Metode ANN dengan Hidden Node 1-5	69
Tabel 4.22 Hasil Klasifikasi Metode ANN Data Training	70
Tabel 4.23 Hasil Klasifikasi Metode ANN Data Testing	71
Tabel 4.24 Perbandingan Ketepatan Klasifikasi Tiga Metode Data Training	73
Tabel 4.25 Hasil Perbandingan Ketepatan Klasifikasi Tiga Metode Data Testing	74

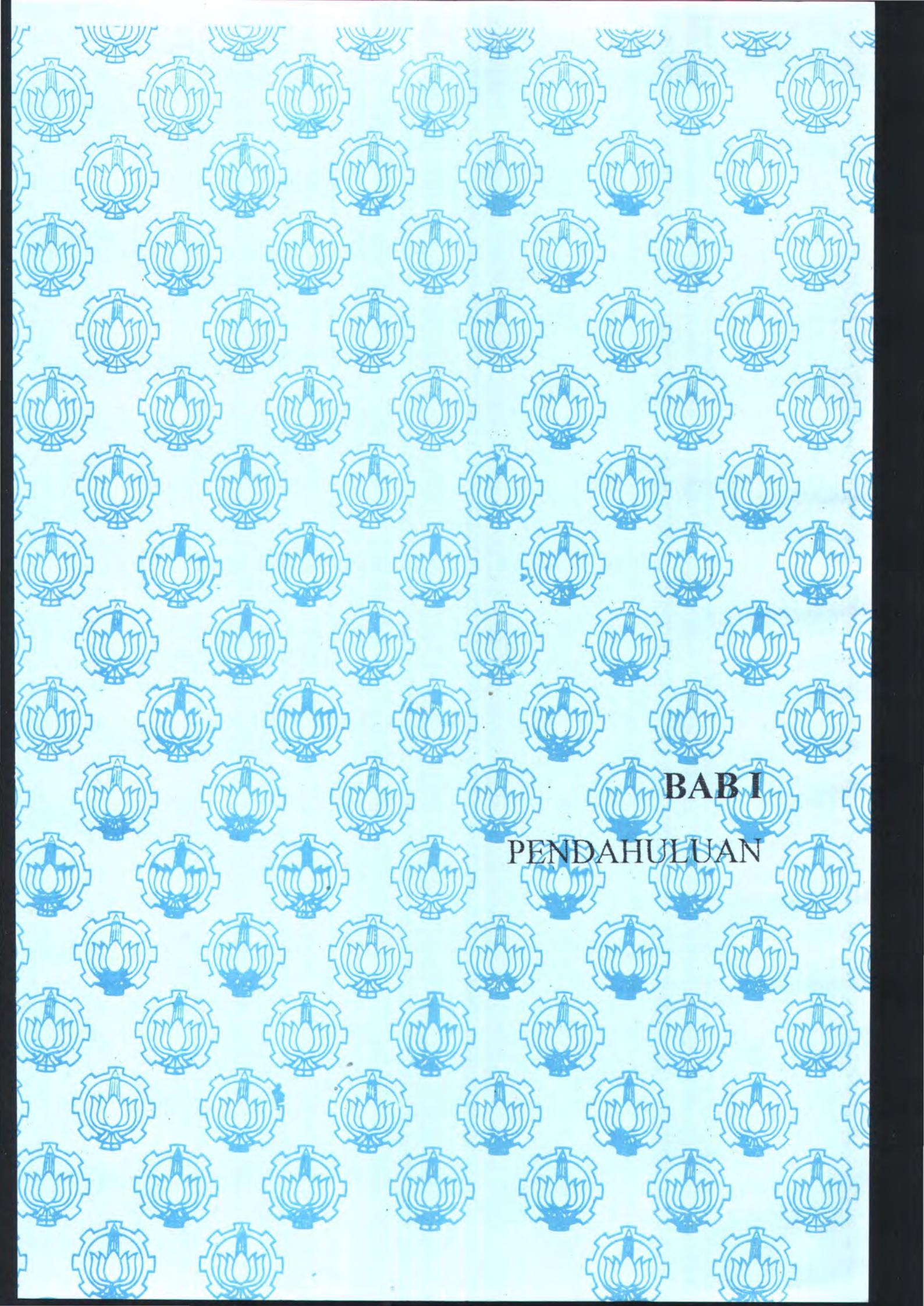
DAFTAR GAMBAR

Halaman

Gambar 2.1 Konsep Dasar Angkatan Kerja.....	5
Gambar 2.2 Feed Forward Networks dengan Hidden Layer Tunggal	26
Gambar 2.3 Kriteria Keakuratan Pengelompokan	34
Gambar 3.1 Bagan Pengolahan Data Secara Umum	45

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Deskriptif Statistik Usia Kerja Kabupaten Kediri.....	79
Lampiran 2 Metode Regresi Logistik Univariat	86
Lampiran 3 Metode Regresi Logistik Multivariat	93
Lampiran 4 Program ANN dan Regesi Logistik dalam Syntax S-Plus.....	98
Lampiran 5 Output Program ANN dan Regresi Logistik.....	100
Lampiran 6 Program Kernel dengan Syntax SAS	103
Lampiran 7 Output Metode Kernel Data Training	104
Lampiran 8 Output Metode Kernel Data Testing	106



BABI
PENDAHULUAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Negara-negara berkembang mempunyai distribusi umur yang muda, selain itu tingginya tingkat kelahiran kasar menyebabkan persentase penduduk berusia dibawah 15 tahun jauh lebih tinggi. Proses transisi demografi yang berkaitan dengan meningkatnya usia harapan hidup penduduk akan merubah distribusi umur dari muda menjadi tua. Proses ini terjadi berkat adanya perbaikan dibidang kesehatan bersamaan dengan meningkatnya teknologi kedokteran dan meningkatnya kesadaran masyarakat (Suwartapradja, 2000). Hal tersebut diatas memicu meningkatnya jumlah pengangguran disuatu negara apabila tidak diimbangi dengan lapangan kerja yang memadai sesuai dengan tingkat pendidikan yang dimiliki.

Dengan memperhatikan kondisi diatas, maka sangat menarik untuk diteliti serta dikaji lebih lanjut mengenai partisipasi seseorang dalam angkatan kerja dengan menggunakan metode klasifikasi yang tidak menggunakan asumsi data harus berdistribusi normal. Hal ini dilakukan karena banyak kasus pengklasifikasian yang datanya tidak normal.

Metode parametrik yang sering digunakan dan lebih disukai untuk masalah klasifikasi adalah analisis regresi logistik yang digunakan untuk analisis data respon kategorik dengan peubah-peubah bebas kontinu dan kategorik.

Pembentukan model terutama untuk pembedaan kelas dengan menghitung probabilitas masing-masing kelas. (Agresti, 1990).

Beberapa metode untuk analisis data kategorik adalah metode Kernel dan *Artificial Neural Network* (ANN). Karena pada penelitian ini selain menggunakan metode parametrik, juga ingin dicoba menggunakan metode non parametrik. Dimana semua metode yang digunakan memiliki asumsi bahwa data tidak harus berdistribusi normal. Metode Kernel memberikan hasil yang lebih baik dibanding fungsi linear diskriminan pada pola distribusi tidak normal. Yuli Arifiani (2000), metode kernel memberikan tingkat kesalahan lebih kecil dibandingkan dengan analisis diskriminan pada kasus identifikasi warna obyek menggunakan kamera. Fauset (1994) memperkenalkan algoritma yang popular digunakan untuk proses training adalah back-propagation dalam masalah klasifikasi.

1.2. Perumusan Masalah

Pada penelitian kali ini permasalahan yang timbul adalah :

1. Bagaimana menentukan faktor-faktor apa saja yang mempengaruhi variabel respon yaitu partisipasi usia kerja dikategorikan bekerja atau tidak bekerja.
2. Bagaimana ketepatan pengelompokan atau kesalahan klasifikasi pada metode klasifikasi, khususnya metode regresi logistik, kernel dan ANN dengan kriteria *Maximum Chance Criterion (MCC)* dan *Proportional Chance Criterion (PCC)*.

1.3. Tujuan penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Menetukan faktor-faktor apa saja yang mempengaruhi variabel respon yaitu partisipasi usia kerja dikategorikan bekerja atau tidak bekerja.
2. Membandingkan ketepatan pengelompokan atau kesalahan klasifikasi pada metode klasifikasi, khususnya metode regresi logistik, kernel dan ANN dengan kriteria *Maximum Chance Criterion (MCC)* dan *Proportional Chance Criterion (PCC)*.

1.4. Manfaat

Manfaat umum yang diperoleh dari penelitian ini adalah dapat menentukan pengelompokan secara tepat pada data respon dengan pendekatan parametrik (regresi logistik), serta dapat mengetahui ketepatan pengelompokan dengan pendekatan nonparametrik (kernel dan *Artificial Neural Network*). Sedangkan manfaat khusus yang diperoleh yaitu sebagai acuan untuk mengelompokkan partisipasi usia kerja dikatakan bekerja atau dikatakan tidak bekerja, dengan indikator-indikator yang sudah ditentukan dengan menggunakan metode regresi logistik dan kernel, serta untuk ketepatan pengelompokannya digunakan metode *Artificial Neural Network*.

1.5. Batasan Masalah

Dalam penelitian ini digunakan data sekunder dari data Survei Sosial Ekonomi Nasional tahun 2000 (Susenas 2000) untuk daerah kabupaten Kediri.

BAB II
TINJAUAN PUSTAKA

BAB II

TINJAUAN PUSTAKA

2.1. TINJAUAN KEPENDUDUKAN

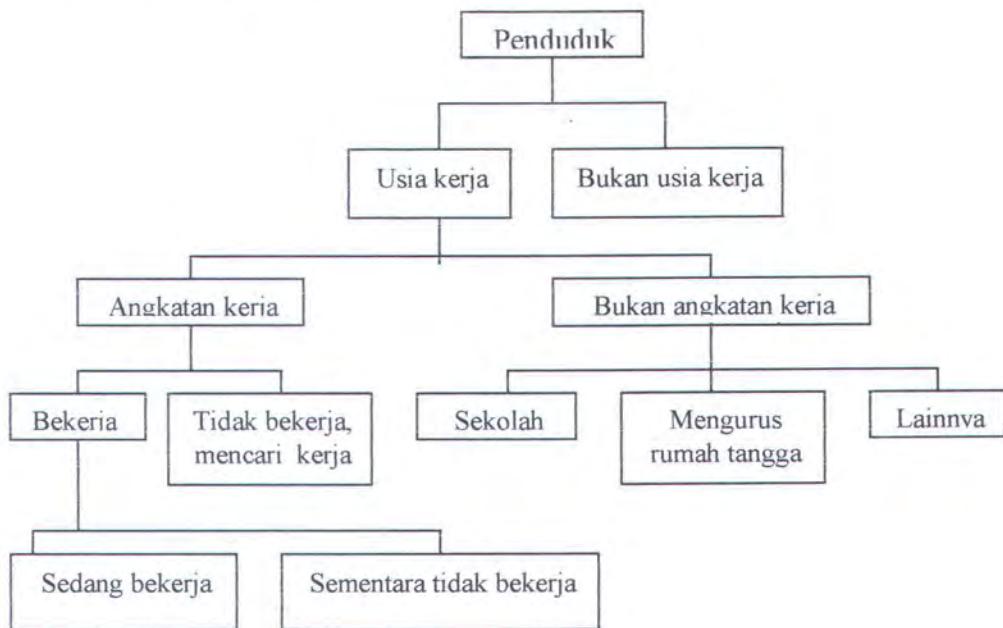
2.1.1. Teori Ketenagakerjaan

Pendekatan teori ketenagakerjaan yang digunakan dalam penelitian ini adalah konsep dasar Angkatan Kerja (*Standard Labour Force Concept*), yaitu seperti yang ditunjukkan pada gambar 2.1.

Penduduk dikelompokkan menjadi penduduk usia kerja dan penduduk bukan usia kerja. Penduduk usia kerja biasanya ditentukan berdasarkan batasan umur. Batasan umur ini berbeda-beda untuk tiap negara. Perbedaan ini kebanyakan disebabkan oleh kondisi masing-masing negara, tetapi pada umumnya yang dijadikan pertimbangan adalah tingkat perekonomian dan situasi tenaga kerja. Semakin maju perekonomian suatu daerah atau negara, batas umur yang ditentukan untuk usia kerja minimum semakin tinggi.

Pengertian tentang tenaga kerja adalah berbeda menurut kultur masing-masing negara atau daerah. Oleh sebab itu, badan kependudukan PBB telah mengeluarkan rekomendasi tentang konsep pengumpulan data ketenagakerjaan yang saat ini juga dianut oleh Indonesia. Batasan umur dalam pengumpulan data ketenagakerjaan di Indonesia masih terfokus pada penduduk usia 10 tahun keatas sebagai usia kerja. Meskipun demikian karena adanya program wajib belajar 9 tahun maka untuk kepentingan pengambilan kebijakan, akhir-akhir ini usia penduduk yang digunakan adalah 15 tahun atau lebih. Oleh karena jenis kegiatan

dari kelompok tersebut berbeda-beda, maka secara umum penduduk usia 10 tahun keatas dibagi kedalam dua kelompok yaitu Angkatan Kerja (AK) dan Bukan Angkatan Kerja (BAK).



Gambar 2.1 Konsep Dasar Angkatan Kerja (*Standard Labor Force Concept*)

Angkatan Kerja adalah penduduk yang bekerja dan penduduk yang mencari pekerjaan. Sedang Bukan Angkatan Kerja adalah mereka yang mengurus rumah tangga, sekolah dan lainnya atau yang tidak melakukan aktivitas ekonomi (pensiun, penerima transfer atau kiriman, penerima deposito/bunga bank, jompo atau alasan lain. Pengukuran didasarkan pada periode tertentu (*time reference*), yaitu kegiatan yang dilakukan selama seminggu yang lalu (BPS 1998).

Yang dimaksud dengan bekerja adalah kegiatan melakukan pekerjaan dengan tujuan memperoleh atau membantu memperoleh nafkah paling sedikit

selama satu jam secara terus-menerus selama seminggu yang lalu. Kegiatan bekerja ini mencangkup baik yang sedang bekerja maupun yang punya pekerjaan tetapi dalam seminggu yang lalu sementara tidak aktif bekerja, misal karena cuti, sakit dan sejenisnya (BPS, 2001).

2.1.2. Faktor-Faktor yang Menjadi Latar Belakang Usia Produktif Bekerja.

Salah satu faktor yang mendorong seseorang untuk bekerja adalah untuk meningkatkan tingkat kesejahteraan hidupnya. Susenas dalam mengukur tingkat kesejahteraan masyarakat adalah melalui pendekatan pengeluaran, karena pada umumnya masyarakat sulit mengungkap secara jujur mengenai pendapatan/penerimannya (Laporan Eksekutif Susenas 2000).

Menurut Suwatapradja (2000), pendidikan akan memperbaiki status, kemampuan dan keahlian seseorang. Hal ini meningkatkan kemampuan bersaing dan meningkatkan permintaan terhadap jasa-jasa di pasar tenaga kerja. Daerah tempat tinggal diduga turut berpengaruh terhadap partisipasi kerja seseorang, hal ini mungkin disebabkan adanya perbedaan tingkat perkembangan ekonomi antara daerah pedesaan dan perkotaan. Diperkotaan terdapat kesempatan ekonomi yang lebih luas dibandingkan di pedesaan. Didaerah pedesaan yang mayoritas penduduknya sebagai petani sebagian besar istri ialah bekerja membantu suami sebagai petani penggarap, dan buruh tani.

Selain itu berdasarkan hasil survei “*Aspek Kehidupan Rumah Tangga Indonesia*” (SAKERTI) dalam majalah *Warta Demografi* (2001), analisis dilakukan dengan menggunakan regresi logistik terhadap beberapa model,

termasuk pemodelan interaksi antara dua variabel yang dianggap mempunyai pengaruh terhadap status bekerja antara lain adalah variabel jenis kelamin berhubungan dengan variabel hubungan dengan kepala rumah tangga, dan antara jenis kelamin dengan status perkawinan. Hasil analisis menunjukkan bahwa yang menjadi faktor penentu adalah: jenis kelamin, hubungan dengan kepala rumah tangga, status perkawian, pendidikan, daerah tempat tinggal dan konsumsi perkapita.

2.2. METODE REGRESI LOGISTIK

Analisis regresi logistik adalah analisis yang digunakan untuk melihat hubungan antara peubah respon kategorik dengan peubah-peubah bebas kategorik maupun kontinu. Peubah respon dalam regresi logistik dapat berbentuk dikhotom (biner) maupun polykhotom (ordinal atau nominal). Metode yang sering digunakan untuk peubah respon berskala ordinal adalah dengan membentuk fungsi logit dari peluang kumulatif. (Agresti, 1990).

Untuk memperoleh penduga kemungkinan maksimum bagi parameternya digunakan metode kuadrat terkecil terboboti secara iteratif. Sedangkan untuk menguji model regresi logistik yang dibentuk digunakan uji nisbah kemungkinan dan aturan klasifikasi berdasarkan nilai-nilai peluang klasifikasi satu individu ke dalam populasi. (Hosmer & Lomeshow, 1989).

Menurut Hosmer (1989) pada metode *Regresi Logistik Berganda*, bentuk hubungan yang dikehendaki adalah :

$$E[Y|x] = \pi(x)$$

$$\pi(x) = \frac{e^{(\beta_0 + \beta' x)}}{1 + e^{(\beta_0 + \beta' x)}} \quad \dots \dots \dots \quad (2.2.1)$$

Dimana $E[Y|x]$ adalah nilai harapan Y pada nilai x tertentu,

$\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ dan $x = (x_1, x_2, \dots, x_p)$ dan p = jumlah parameter.

2.2.1 Model Regresi Logistik.

Melalui transformasi logit terhadap model regresi logistik diperoleh :

$$\begin{aligned}\pi(x) &= \frac{e^{(\beta_0 + \beta'X)}}{1 + e^{(\beta_0 + \beta'X)}} \\ \pi(x)(1 + e^{(\beta_0 + \beta'X)}) &= e^{(\beta_0 + \beta'X)} \\ \pi(x) + \pi(x)e^{(\beta_0 + \beta'X)} &= e^{(\beta_0 + \beta'X)} \\ \pi(x) &= e^{(\beta_0 + \beta'X)} - \pi(x)e^{(\beta_0 + \beta'X)} \\ \pi(x) &= e^{(\beta_0 + \beta'X)} \cdot (1 - \pi(x)) \\ \\ \frac{\pi(x)}{1 - \pi(x)} &= e^{(\beta_0 + \beta'X)} \\ \\ \ln\left\{\frac{\pi(x)}{1 - \pi(x)}\right\} &= \ln\left\{e^{(\beta_0 + \beta'X)}\right\} \\ \\ g(x) &= \beta_0 + \beta'x\end{aligned}$$

Dimana :

Selanjutnya $g(x)$ disebut dengan *Model Logit*.

Pada model regresi linier berganda diasumsikan bahwa $Y = E[Y|x] + \varepsilon$

dimana ϵ adalah galat dan menunjukkan selisih nilai obyek pengamatan terhadap

nilai harapannya. Galat diasumsikan berdistribusi normal dengan rataan nol dan variansi tetap (konstan) terhadap variabel pengamatannya. $\varepsilon \approx N(0, \sigma^2)$.

Sementara pada regresi logistik berganda dinyatakan :

$$Y = \pi(x) + \varepsilon$$

Sehingga dengan obyek pengamatan yang dikotomik dinyatakan, jika

$$Y = 0, \text{ maka } \varepsilon = -\pi(x) \text{ dengan probabilitas } \pi(x)$$

$$Y = 1, \text{ maka } \varepsilon = 1 - \pi(x) \text{ dengan probabilitas } \pi(x)$$

Dan dapat dinyatakan bahwa ε memiliki $E(\varepsilon) = 0$ dan $\text{var}(\varepsilon) = \pi(x).[1 - \pi(x)]$ yang mengikuti distribusi Binomial (Hosmer, 1989)

2.2.2 Pembentukan Model Regresi Logistik dan Estimasi Parameter

Dalam regresi linier, metode yang sering digunakan untuk mengestimasi parameter-parameter yang belum diketahui adalah metode *Ordinary Least Square*. Dengan metode ini dapat ditentukan nilai penduga parameter β_0 dan β_j yang meminimumkan jumlah kuadrat deviasi (error) nilai Y pengamatan. Tetapi cara *Least Square* ini tidak dapat diterapkan pada model dengan variabel respon dikotom. Nilai estimasi parameternya berbeda dengan nilai estimasi yang diperoleh dari regresi linier.

Metode pendugaan yang dipakai adalah Pendugaan *Likelihood Maksimum* (*Maximum Likelihood Estimation atau MLE*) dengan menetapkan asumsi distribusi Bernoulli dan obyek pengamatan yang saling bebas, $E(\varepsilon_i, \varepsilon_j) = 0, i \neq j$.

Pada pasangan pengamatan (x_i, y_i) fungsi likehood yang dimaksimumkan adalah:

$$\zeta(x_i) = \pi(x_i)^{y_i} \{1 - \pi(x_i)\}^{n_i - y_i} \quad \dots \dots \dots \quad (2.2.3)$$

Karena pengamatan diasumsikan bersifat independen, maka likelihood pengamatan merupakan perkalian dari fungsi likelihood masing-masing, misal dinyatakan $I(\beta)$

$$\begin{aligned} I(\beta) &= \prod_{i=1}^I \zeta(x_i) = \prod_{i=1}^I \pi(x_i)^{y_i} \{1 - \pi(x_i)\}^{n_i - y_i} \\ &= \prod_{i=1}^I \pi(x_i)^{y_i} \frac{\{1 - \pi(x_i)\}^{n_i}}{\{1 - \pi(x_i)\}^{y_i}} \\ &= \prod_{i=1}^I \left[\frac{\pi(x_i)}{(1 - \pi(x_i))} \right]^{y_i} \{1 - \pi(x_i)\}^{n_i} \quad \dots \dots \dots \quad (2.2.4) \end{aligned}$$

Dimana y_i dengan $i = 1, 2, 3, \dots, I$ adalah variabel random Binomial yang saling bebas dengan $E(y_i) = n_i \cdot \pi(x_i)$ dan $\sum_i n_i = N$.

Karena $z = \exp(\ln z)$, maka untuk $z = \left[\frac{\pi(x_i)}{\{1 - \pi(x_i)\}} \right]^{y_i}$ persamaan 2.2.4 dapat ditulis menjadi :

$$\begin{aligned} &\left[\prod_{i=1}^I \{1 - \pi(x_i)\}^{n_i} \right] \left[\prod_{i=1}^I \exp \left[\ln \left\{ \frac{\pi(x_i)}{(1 - \pi(x_i))} \right\}^{y_i} \right] \right] \\ &\left[\prod_{i=1}^I \{1 - \pi(x_i)\}^{n_i} \right] \exp \left[\sum_{i=1}^I y_i \cdot \ln \left\{ \frac{\pi(x_i)}{\{1 - \pi(x_i)\}} \right\} \right] \quad \dots \dots \dots \quad (2.2.5) \end{aligned}$$

Apabila dilakukan transformasi logit terhadap model regresi logistik persamaan 2.2.5, maka didapatkan :

$$\exp \left[\sum_{i=1}^I y_i \cdot \ln \left\{ \frac{\pi(x_i)}{(1 - \pi(x_i))} \right\} \right] = \exp \left[\sum_i y_i \left(\sum_j \beta_j x_{ij} \right) \right] = \exp \left\{ \left(\sum_i y_i x_{ij} \right) \beta_j \right\}$$



Dan

$$\begin{aligned}
 \{1 - \pi(x_i)\} &= 1 - \frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} \\
 &= \frac{1 + \exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} - \frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} \\
 &= \frac{1 + \exp\left(\sum_j \beta_j x_{ij}\right) - \exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} \\
 &= \frac{1}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} = \left\{1 + \exp\left(\sum_j \beta_j x_{ij}\right)\right\}^{-1}
 \end{aligned}$$

Sehingga persamaan 2.2.5 dapat ditulis menjadi :

$$\left[\sum_{i=1}^I \left\{ 1 + \exp\left(\sum_j \beta_j x_{ij}\right) \right\}^{-n_i} \right] \left[\exp\left\{ \sum_j (y_i x_{ij}) \beta_j \right\} \right] \dots \dots \dots \quad (2.2.6)$$

Prinsip dari MLE untuk mendapatkan nilai taksiran β adalah dengan memaksimalkan fungsi likelihood. Nilai maksimum dari fungsi persamaan 2.2.6 diperoleh melalui transformasi log sebagai berikut :

$$\begin{aligned}
 L(\beta) &= \ln(I(\beta)) \\
 &= \left\{ \sum_j \left(\sum_i y_i x_{ij} \right) \beta_j \right\} - \sum_i n_i \ln \left\{ 1 + \exp\left(\sum_j \beta_j x_{ij}\right) \right\} \dots \dots \quad (2.2.7)
 \end{aligned}$$

Nilai β diperoleh melalui turunan parsial pertama $L(\beta)$ terhadap β yang disamadengarkan nol, sehingga persamaan 2.2.7 akan menjadi :

$$\begin{aligned} \frac{\partial L(\beta)}{\partial L\beta_a} &= \sum_i y_i x_{ia} - \sum_i n_i x_{ia} \left[1 + \exp \left(\sum_j \beta_j x_{ij} \right) \right]^{-1} \cdot x_{ia} \left(\sum_j \beta_j x_{ij} \right) \\ &= \sum_i y_i x_{ia} - \sum_i n_i x_{ia} \left[\frac{\exp \left(\sum_j \beta_j x_{ij} \right)}{1 + \exp \left(\sum_j \beta_j x_{ij} \right)} \right] \\ &= \sum_i y_i x_{ia} - \sum_i n_i x_{ia} \pi_i = 0 \dots \dots \dots \quad (2.2.8) \end{aligned}$$

$a = 0, 1, 2, \dots, p$ parameter.

Dimana : $\pi_i = \frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)}$ menyatakan taksiran likelihood maksimum dari $\pi(x_i)$.

Sedang terhadap varians diperoleh dari turunan parsial kedua persamaan 2.2.8 dari hasil penurunan tersebut diperoleh :

$$\frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b} = - \sum_i n_i x_{ia} x_{ib} \cdot \frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{\left[1 + \exp\left(\sum_j \beta_j x_{ij}\right)\right]^2} \quad \dots \quad (2.2.9)$$

$$\text{Karena } \pi_i = \frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} \text{ dan } 1 - \pi_i = \frac{1}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)}$$

Maka persamaan 2.2.9 dapat ditulis menjadi :

2.2.3 Pengujian Kelayakan Model

Tahap berikutnya setelah melakukan pendugaan parameter model adalah menguji kelayakan model. Digunakan uji dan hipotesa-hipotesa statistik untuk menentukan apakah variabel-variabel prediktor dalam model mempunyai hubungan yang nyata atau signifikan dengan variabel responnya. Pengujian dilakukan secara parsial maupun secara serentak.

1. Pengujian hipotesis secara Parsial

Hipotesis pengujian ini adalah :

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Statistik uji yang dipakai adalah *Statistik Uji Wald*

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim N(0,1)$$

Pada tingkat kepercayaan α , hipotesa nol (H_0) ditolak bila nilai $W_j > Z\alpha/2$ atau nilai p-value $< \alpha$; dimana Z menunjukkan nilai variabel

random pada tabel distribusi normal standard dan didapatkan nilai p-value = $P(|Z| > W_j)$.

2. Pengujian Hipotesis secara Serentak

Hipotesis pengujian ini adalah :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad j = 1, 2, \dots, p$$

$$H_1 : \text{paling sedikit ada satu } \beta_j \text{ yang tidak sama dengan nol.}$$

Pengujian ini disebut juga *Uji Chi-Square*. Statistik uji yang dipakai adalah *Statistik Uji G (Uji Perbandingan Likelihood)*.

$$G^2 = -2 \cdot \ln \left(\frac{L_1}{L_0} \right)$$
$$G^2 = -2(Ln(L_1) - Ln(L_0))$$

dimana L_1 = Likelihood tanpa variabel prediktor tertentu

L_0 = Likelihood dengan variabel prediktor tertentu

Pada tingkat kepercayaan α , hipotesa nol (H_0) ditolak bila didapatkan nilai p-value $< \alpha$ atau nilai $G > \chi_{\alpha,b}^2$ dimana $\chi_{\alpha,b}^2$ menunjukkan nilai variabel random pada tabel distribusi chi-square pada derajat bebas b .

2.2.4 Interpretasi Model

Setelah model dibentuk dan layak baik ditinjau dari hasil pengujian statistik maupun kepentingan bidang ilmu yang dikaji maka diperlukan

interpretasi terhadap koefisien pembentuk model tersebut. Dalam menginterpretasikan koefisien parameter digunakan yang *odds ratio* (ψ). Untuk maksud tersebut akan dilakukan pembahasan terhadap beberapa keadaan variabel pengamatan.

Tabel 2.1. Tabel Nilai Model Metode Regresi Logistik

	$X=1$	$X=0$
$Y=1$	$\pi(1)=\frac{e^{\beta_0 + \beta_1}}{1+e^{\beta_0 + \beta_1}}$	$\pi(0)=\frac{e^{\beta_0}}{1+e^{\beta_0}}$
$Y=0$	$1-\pi(1)=\frac{1}{1+e^{\beta_0 + \beta_1}}$	$1-\pi(0)=\frac{1}{1+e^{\beta_0}}$

2.3. METODE KERNEL

Metode Kernel merupakan salah satu metode nonparametrik dimana tidak memerlukan asumsi-asumsi tertentu, selain itu metode kernel sebenarnya tidak memerlukan asumsi mengenai bentuk distribusi data yang akan dianalisis (Seber 1984). Jika distribusi masing-masing grup tidak memiliki asumsi tertentu atau suatu asumsi kenormalan dilanggar, maka metode nonparametrik ini dapat digunakan untuk menduga peluang grup tersebut dan metode nonparametrik lebih fleksibel dalam berbagai kasus yang terjadi dalam kehidupan sehari-hari (Holmstrom dan Sain 1997).

Suatu karakteristik dasar yang menunjukkan sifat dari variabel acak X adalah fungsi kepadatan peluangnya (Hindle, 1990). Dalam praktiknya, bentuk fungsi kepadatan peluang dari suatu variabel acak tidak selalu diketahui. Jika

diberikan sejumlah n pengamatan yang diasumsikan *independent* dan merupakan variabel acak yang berdistribusi identik f , maka tujuannya adalah menduga fungsi kepadatan peluang berdasar pengamatan tersebut. Ada dua pendekatan yang bisa dilakukan, yaitu pendekatan parametrik dan nonparametrik. Dalam pendekatan parametrik, menduga fungsi kepadatan peluang sama artinya dengan menduga parameter. Sedangkan dalam pendekatan nonparametrik sama artinya dengan menduga fungsi kepadatan secara keseluruhan.

Kelebihan taksiran Kernel adalah bentuknya lebih fleksibel dan bentuk matematisnya lebih mudah disesuaikan (Härdle, 1990). Secara umum fungsi Kernel didefinisikan sebagai :

$$Kh(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \text{ untuk } -\infty < x < \infty \quad \dots \dots \dots \quad (2.3.1)$$

serta memenuhi :

1. $K(x) \geq 0$, untuk semua x
 2. $\int_{-\infty}^{\infty} K(x) dx = 1$
 3. $\int_{-\infty}^{\infty} K^2(x) dx < \infty$

Maka taksiran fungsi kepadatan kernelnya adalah :

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad \dots \quad (2.3.2)$$

Dimana :

K : Fungsi Kernel

h : Derajat penghalusan (*Bandwidth*)

n : Jumlah observasi pada salah satu grup

2.3.1. Taksiran Mean dan Varians $\hat{f}_h(x)$

Taksiran fungsi kernel bergantung pada dua parameter, yaitu :

- *Fungsi Kernel K*
- *Bandwidth h*

Selanjutnya diteliti apakah ekspektasi $\hat{f}_h(x)$ merupakan taksiran yang *unbiased*

$$\begin{aligned} E\{\hat{f}_h(x)\} &= \frac{1}{n} \sum_{i=1}^n E\{K_h(x - x_i)\} \\ &= E\{K_h(x - x_i)\} \\ &= \int K_h(x - \mu) f(\mu) d\mu \\ &= \int K(s) f(x + sh) ds \\ &= \int K(s) \left(f(x) + \frac{sh}{1!} f'(x) + \frac{sh^2}{2!} f''(x) + \dots \right) ds \\ &= f(x) \int K(s) ds + o(h) \int K(s) ds \\ &= f(x) \quad , \quad h \rightarrow 0 \end{aligned}$$

Oleh karena $E(\hat{f}_h(x)) = f(x)$ pada saat *bandwidth* h mendekati nol maka $\hat{f}_h(x)$ secara asimtotik adalah taksiran yang tidak bias. Sedangkan taksiran variansinya adalah sebagai berikut :

$$\begin{aligned}
\text{Var}(\hat{f}_h(x)) &= \frac{1}{n^2} \text{Var} \left\{ \sum_{i=1}^n K_h(x - x_i) \right\} \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[K_h(x - x_i)] \\
&= \frac{1}{n} \text{Var}[K_h(x - x_i)] \\
&= \frac{1}{n} \left\{ E[K_h^2(x - X)] - [E[K_h(x - x_i)]]^2 \right\} \\
&= \frac{1}{n} \left\{ \left(\frac{1}{h^2} \int K^2\left(\frac{x-\mu}{h}\right) f(\mu) d\mu \right) - (f(x) + o(h))^2 \right\} \\
&= \frac{1}{n} \left\{ \left(\frac{1}{h} \int K^2(s) f(x+sh) ds \right) - (f(x) + o(h))^2 \right\} \\
&= \frac{1}{n} \left\{ \left(\frac{1}{h} \int K^2(s) ds (f(x) + o(h)) \right) - (f(x) + o(h))^2 \right\} \\
&= \frac{1}{n} \left\{ \left(\frac{1}{h} \|K\|_2^2 (f(x) + o(h)) \right) - (f(x) + o(h))^2 \right\} \\
&= \frac{1}{n} \|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right) \quad nh \rightarrow \infty \dots \dots \dots \quad (2.3.3)
\end{aligned}$$

Varians di atas mendekati proporsional pada $\frac{1}{nh}$ apabila dipilih h yang besar akan menghasilkan varians yang kecil.

2.3.2 Bias $\hat{f}_h(x)$

Pada pembahasan sebelumnya diketahui bahwa $E(\hat{f}_h(x))$ mendekati $f_h(x)$ jika h mendekati nol dan $n \rightarrow \infty$. Sehingga untuk n yang terbatas $\hat{f}_h(x)$

merupakan taksiran yang bias dari $f_h(x)$. Selanjutnya dibahas tentang bias tersebut.

$$\begin{aligned}\text{Bias} \left(\hat{f}_h(x) \right) &= \int K(s) \int (x + sh) ds - f(x) \\ &= \int K(s) \left[f(x) + sh f'(x) + \frac{h^2 s^2}{2} f''(x) + o(h^2) \right] ds - f(x) \\ &= f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) - f(x)\end{aligned}$$

Jika K simetris disekitar nol, maka $\int sK(s) h^f(x) ds = 0$, sehingga :

Terlihat bahwa bias terjadi pada saat h kuadrat, sehingga dipilih h yang kecil untuk memperkecil bias.

Jelaslah bahwa bias dari x proporsional pada turunan kedua dari $f(x)$. sehingga dapat diambil kesimpulan bahwa :

1. Jika nilai $bandwidth h$ semakin kecil, bias semakin kecil maka varians akan bertambah besar.
 2. Jika nilai $bandwidth h$ bertambah besar, bias bertambah besar maka varians akan semakin kecil.

Hal ini berarti bahwa :

- Derajat penghalusan yang kecil memberikan bias yang kecil tetapi varians yang besar.
 - Derajat penghalusan yang besar memberikan bias yang besar tetapi varians yang kecil.

Untuk mengkompromikan kedua efek tersebut di atas digunakan *Mean Squared Error MSE* $(\hat{f}_h(x))$ atau *Mean Integrated Square Error MISE* $(\hat{f}_h(x))$.

Dimana :

$$\begin{aligned} \text{MSE}(\hat{f}_h(x)) &= \text{var}(\hat{f}_h(x)) + [\text{bias}(\hat{f}(x))]^2 \\ &= \frac{1}{nh} f(x) \left\| K \right\|_2^2 \frac{h^4}{4} (f'(x) \mu_2(K))^2 + o\left(\frac{1}{nh}\right) + o(h^4) \\ h \rightarrow 0, \quad nh \rightarrow \infty. \end{aligned} \quad (2.3.5)$$

$\text{MSE}(\hat{f}_h(x)) \rightarrow 0$, jika $h \rightarrow 0$ dan $nh \rightarrow \infty$, sehingga fungsi padat probabilitas kernel konsisten.

2.3.3 Pemilihan Bandwidth h

Masalah terpenting yang berhubungan dengan penggunaan taksiran kepadatan Kernel adalah pemilihan *bandwidth* yang optimal. Nilai optimal dari h tergantung pada kriteria yang digunakan untuk mengukur keseluruhan akurasi dari $f_h(x)$. Kriteria yang biasa digunakan adalah *Mean Square Error*.

Ada beberapa cara untuk menyelesaikan masalah ini selain cara di atas, diantaranya :

1. Mencoba beberapa nilai h , misalnya h_1, h_2, \dots dan kemudian menghitung $f'(x)$ untuk masing-masing nilai h . Kemudian dapat dipilih bentuk $f'(x)$ yang kelihatan sebagai bentuk yang paling masuk akal. (Asumsikan bahwa kita mengetahui dengan baik bagaimana kehalusan (*smooth*) dan fungsi sebenarnya).
2. Menggunakan metode *Cross Validation*.

Berdasarkan taksiran kepadatan \hat{f}_h dan ingin diuji H_0 khusus, maka hipotesisnya adalah :

$$H_0 : \hat{f}_h(x) = f(x)$$

$$H_1: \hat{f}_h(x) \neq f(x)$$

Statistik ujinya adalah : $\frac{f(x)}{\hat{f}_h(x)}$

Bandwidth yang baik adalah bila statistik uji tersebut nilainya mendekati satu.

Cara lain untuk memilih parameter penghalusan h adalah dengan memilih sebuah nilai yang mengoptimalkan kriteria yang ada. Grup yang berbeda mungkin memiliki sekumpulan nilai optimal yang berbeda. Diasumsikan bahwa fungsi kepadatan yang tidak diketahui tersebut memiliki batasan dan turunan kedua yang kontinu, dan kernel adalah sebuah fungsi kepadatan peluang yang simetris. Salah satu kriteria adalah untuk meminimalkan perkiraan $MISE$ dari fungsi kepadatan terduga. Nilai optimal h yang dihasilkan tergantung pada fungsi kepadatan dan kernel. Pilihan yang berasalan untuk parameter penghalusan h adalah untuk mengoptimalkan kriteria dengan mengasumsikan bahwa grup t berdistribusi normal dengan matriks kovarians V_t . Maka nilai optimal h yang dihasilkan pada grup t yaitu :

$$(A(K_t)/n_t)^{1/(p+4)} \dots \quad (2.3.6)$$

dimana konstanta optimal A (K_t) tergantung pada kernel K_t . Untuk beberapa kernel yang berguna. Konstanta A (K_t) diberikan pada :

$$A(K_t) = \frac{2^{p+1}(p+2)\Gamma(p/2)}{p}$$

Dengan kernel *Uniform*.

Dengan kernel Normal.

$$A(K_l) = \frac{2^{p+1} p^2 (p+2)(p+4)\Gamma(p/2)}{2p+1}$$

Dengan kernel *Epanechnikov*.

Nilai h yang optimal diperoleh dengan mensubstitusikan nilai $A(K_t)$ ke dalam persamaan:

$$h = (A(Kt)/nt)^{1/(p+4)}$$

dimana :

t : sebuah subskrip untuk membedakan grup

n_t : jumlah pengamatan dalam grup t

p : banyaknya grup

Γ : fungsi gamma

2.3.4 Pemilihan Fungsi Kernel

Kriteria pemilihan Kernel yang baik berdasarkan pada resiko kernel minimum yang diperoleh dari kernel-kernel optimal atau kernel-kernel dengan variansi minimum. Secara analitik sifat dari $\hat{f}_h(x)$ bergantung pada penentuan nilai *bandwidth*. Macam-macam fungsi Kernel dapat dijelaskan pada tabel 2.2 berikut :

Tabel 2.2 Macam-Macam Fungsi Kernel (dengan I adalah fungsi indikator)

Kernel	$K(\mu)$
<i>Uniform</i>	$\frac{1}{2}I(\mu \leq 1)$
<i>Triangle</i>	$(1 - \mu)I(\mu \leq 1)$
<i>Epanechnikov</i>	$\frac{3}{4}(1 - \mu^2)I(\mu \leq 1)$
<i>Quartie/ Biweight</i>	$\frac{15}{16}(1 - \mu^2)^2 I(\mu \leq 1)$
<i>Triweight</i>	$\frac{35}{32}(1 - \mu^2)^3 I(\mu \leq 1)$
<i>Gaussian</i>	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mu^2\right)$
<i>Cosinus</i>	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}\mu\right)I(\mu \leq 1)$

Sumber : Hardle, 1990

Dalam praktek biasanya dipilih Kernel *Gaussian* (Seber,1984) dikarenakan lebih halus dibandingkan fungsi kernel yang lain.

2.4. METODE ARTIFICIAL NEURAL NETWORK (ANN)

Artificial Neural Network (ANN) merupakan suatu sistem analisis yang proses kerjanya diilhami dari aktifitas jaringan syaraf pada manusia yang terdiri dari sekumpulan neuron-neuron atau unit-unit yang saling berinteraksi. Pada neuron syaraf manusia, proses alami mengatur bagaimana sinyal input pada dendrit diproses dan kemudian diterjemahkan dalam aktivitas *axon*. Sedangkan pada neuron buatan, proses learning mengatur input-input yang digunakan untuk pemetaan outputnya. ANN hanya tergantung pada arsitektur, training, testing dan algoritmanya. Proses training merupakan proses pembelajaran dari ANN yang

mengatur input – input yang digunakan dan bagaimana pemetaannya pada output hingga diperoleh model ANN serta training terjadi pada saat pengaturan weight dan bias. Sedangkan proses testing merupakan proses pengujian ketelitian dari model yang telah diperoleh dari proses training. Algoritma yang populer digunakan untuk proses training adalah *backpropagation* (Fauset, 1994).

Struktur ANN menurut sistem kerjanya terdiri dari tiga lapisan yaitu lapisan input (*input layer*), lapisan antara lapisan tersembunyi (*hidden layer*) dan lapisan output (*output layer*). Masing-masing lapisan diberikan pembobot (*weight*) yang akan mentransformasi nilai input menjadi nilai output. Setiap layer terdiri dari beberapa neuron dan antar neuron-neuron akan terhubung dengan neuron-neuron lain pada *layer* terdekat. (Ripley, 1996).

2.4.1 Feed Forward Artificial Neural Network (FFANN)

ANN juga merupakan unit dasar dari *Feed Forward Layered Neural Network* (FFANN) dalam neuron yang formal. FFANN merupakan suatu jaringan dimana mempunyai neuron-neuron yang hanya mempunyai satu jalur hubungan dengan yang lainnya. Setiap neuron dapat diberi label atau nomor mulai angka kecil pada input sampai angka terbesar pada output, sehingga setiap neuron hanya terhubung dengan neuron yang mempunyai angka yang lebih tinggi. Secara notasi yang lebih sederhana, setiap neuron menjumlahkan input dan menambahkan sebuah konstanta (bias) untuk membentuk input total dan menerapkan fungsi aktivasi pada input total untuk mendapatkan suatu output. Tiap jalur tersebut dihubungkan oleh sebuah pembobot. (Fauset, 1994).

Jaringan pembobotan didasarkan pada perbedaan target dengan hasil output. Kesalahan dari *output layer* akan diinteraksikan mundur oleh jaringan “*link weight*”. Prosedur ini akan berulang sampai diperoleh nilai yang konvergen dengan nilai error yang minimum. Menurut Ripley (1996) hubungan dari input dan output atau dapat dinyatakan sebagai berikut :

$$y_k = f_k \left[\alpha_k + \sum_{j \rightarrow k} w_{jk} f_j \left[\alpha_j + \sum_{i \rightarrow j} w_{ij} x_i \right] \right] \dots \dots \dots \quad (2.4.1)$$

dimana :

x : signal input

y : signal output

f_k : fungsi aktivasi

α_k : nilai bias untuk output ke-k

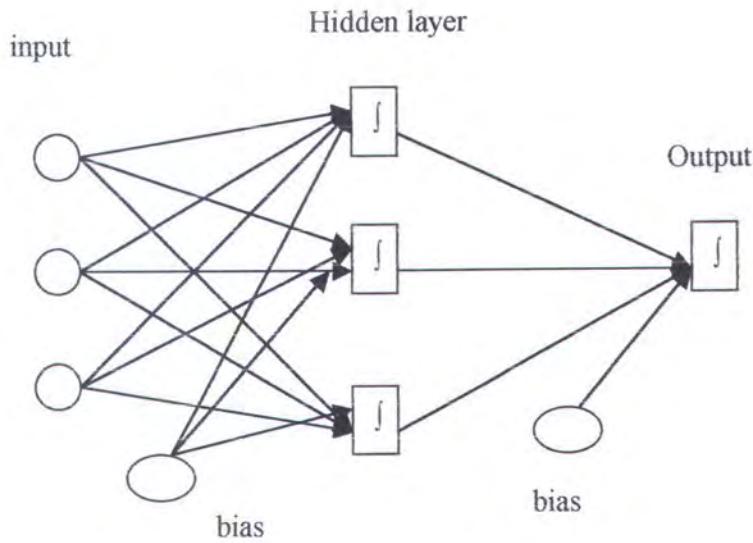
α_j : nilai bias untuk hidden note ke-j

$i = 1, 2, \dots, l; j = 1, 2, \dots, m$ dan $k = 1, 2, \dots, n$.

$\sum_{j \rightarrow k} w_{jk}$: jumlah bobot dari hidden ke output

$\sum_{i \rightarrow j} w_{ij}$: jumlah bobot dari input ke hidden.

Arsitektur ini juga disebut *Multi Layer Perceptron* (MLP) dengan satu *hidden layer*. Gambar dibawah ini menunjukkan sebuah fungsi arsitektur ANN dengan *hidden layer* tunggal.



Gambar 2.2 Feed Forward Networks dengan Hidden Layer Tunggal

Dalam sebuah arsitektur network penggunaan fungsi aktivasi tidak selalu sama tergantung dari permasalahan dan teori pendukungnya. Dalam penerapannya, permasalahan utama yang sering muncul adalah penentuan parameter atau pembobotan serta pengoptimalan jumlah *layer* dan jumlah neuron dari suatu arsitektur *network*.

Secara umum *Neural Network* merupakan suatu arsitektur yang terdiri dari satu atau beberapa *hidden layer* dimana semua neuron dalam sebuah *layer* mempunyai fungsi yang sama. Arsitektur NN dengan fungsi logistik pada output dapat dipandang sebagai bentuk non linier dari metode regresi logistik.

Proses dalam ANN terbagi menjadi tiga tahap utama yaitu *feed forward*, *back propagation* dan *update* nilai bobot dimana penjelasan dari tahapan-tahapan tersebut adalah sebagai berikut :

a. *feed forward*

tahap *feed forward* dilakukan proses dari input sampai diperoleh hasil output. Dalam proses ini input masuk pada input layer yang berupa data. Nilai yang berada pada *node* input dilanjutkan ke semua *node* pada *hidden layer* (kecuali bias), selanjutnya nilai tersebut diteruskan ke semua output *layer*. Nilai-nilai yang dikirim akan diaktifasi. Ada beberapa fungsi aktivasi yang dapat digunakan, antara lain :

$$\text{Hard Limit} \quad y = \begin{cases} f(x) = 1, & \text{jika } x > 0 \\ f(x) = 0, & \text{jika } x \leq 0 \end{cases} \quad \dots \dots \dots \quad (2.4.2)$$

$$\text{Linear} \quad y = W^T X + \alpha \quad \dots \dots \dots \quad (2.4.3)$$

$$\text{Logistic Sigmoid} \quad y = \frac{1}{1+e^{-n}} = \frac{1}{1+e^{-(W^T X + \alpha)}} \quad \dots \dots \dots \quad (2.4.4)$$

$$\text{Dimana } n = W^T X + \alpha$$

$$\text{Hyperbolic Tangen} \quad y = \frac{e^n - e^{-n}}{e^n + e^{-n}} = \tanh(W^T x + \alpha) \quad \dots \dots \dots \quad (2.4.5)$$

Gaussian Radial Basic Function:

$$y = \exp\left(-\frac{n^2}{2\sigma^2}\right) = \exp\left(-\frac{(W^T x + \alpha)^2}{2\sigma^2}\right) \quad \dots \dots \dots \quad (2.4.6)$$

Setiap fungsi diatas, digunakan dalam arsitektur ANN untuk beberapa tujuan yang berbeda.

Algoritma dalam proses *feed forward* ini adalah sebagai berikut :

1. Penentuan inisialisasi bobot, yaitu dengan random atau ditentukan nilainya.
2. Masukkan input signal x_i , $i = 1, 2, \dots, l$ pada node input.
3. Kalikan signal yang masuk dengan nilai bobot link dan berikan hasilnya ke *hidden layer*.
4. Nilai yang diterima oleh hidden node (z_j , $j = 1, 2, \dots, m$) dijumlahkan dan dilakukan suatu proses aktivasi.
5. Lakukan proses seperti tahap 3 dalam penyampaian signal dari *hidden layer* ke *output layer*.
6. Bandingkan hasilnya dengan nilai target pada tahap *backpropagation*.

b. Backpropagation

Pada tahap *backpropagation* ini dilakukan proses pembandingan nilai output dari tahap *feed forward* dengan nilai target yang ditentukan, kemudian dilanjutkan sampai ke depan sampai *input layer* sehingga diperoleh nilai error. Dalam metode *backpropagation*, *weight-weight* tersebut diatur untuk meminimalisasikan nilai kuadrat beda antar output model dan output taksiran (*Sum Square Error*). Nilai error tersebut akan digunakan sebagai dasar untuk meng-update nilai bobot yang menghubungkan antara *hidden layer* dan *output layer*. Proses yang ada pada tahap ini sama dengan proses pada tahap *feed forward*, hanya saja pada tahap *back propagation ini*, proses dilakukan dari *output layer* ke *input layer*.

Error pada sebuah unit keluaran didefinisikan oleh :

dimana t_k adalah harga keluaran yang diinginkan, y_k adalah keluaran aktual dari unit ke k atau nilai output pada *layer* terakhir, dengan "k" merupakan unit keluaran ke-k.

Dalam proses *backpropagation*, bentuk fungsi error yang dieliminasi adalah sebagai berikut :

$$EP = \frac{1}{2} \sum_{k=1}^r (t_k - y_k)^r (t_k - y_k) \dots \dots \dots (2.4.8)$$

Nilai vektor selisih yang dievaluasi dengan perulangan mundur

$$\tau = \tau - 1, \tau - 2, \dots, 1$$

adalah :

$$\delta_k^{\tau}(i) = \left[\sum_{\lambda=1}^{\tau+1} \delta_k^{\tau+1}(\lambda) W^{\tau+1}(\lambda i) \right] \frac{\partial f^{\tau}}{\partial n_k^{\tau}(i)} \dots \quad (2.4.9)$$

Secara umum, algoritma dalam proses *backpropagation* adalah sebagai berikut :

1. Bandingkan setiap nilai output yang dihasilkan ($y_k, k = 1, 2, \dots, m$) dengan nilai target (t_k).
 2. Hitung nilai koreksi untuk nilai bobot yang menghubungkan *hidden node* dan *output node*, serta *output node* dan bias.
 3. Nilai error yang diperoleh dari output layer dilanjutkan ke seluruh *node* yang ada pada *hidden layer*, kemudian dikalikan dengan nilai bobot yang ada pada jalur yang dilaluinya. Setiap nilai yang diterima oleh *hidden node*

akan dijumlahkan dan dikalikan dengan turunan dari fungsi aktivasi untuk memperoleh nilai error.

4. Hitung nilai koreksi untuk nilai bobot antara *input node* dan *hidden node* serta antara *hidden node* dan bias.

Proses estimasi dengan menggunakan backpropagation, terdapat beberapa permasalahan yang sering dijumpai antara lain pemilihan nilai awal, lokal minimum, jumlah neuron pada lapisan *hidden* dan lain-lain. Meskipun telah ada beberapa metode untuk memperkecil peluang munculnya permasalahan tersebut, tetapi tidak menjamin bahwa estimasi *backpropagation* akan memberikan hasil yang lebih baik (Fauset, 1994).

c. Update nilai bobot

Proses ini merupakan tahap terakhir dari ANN. Pada tahap ini dilakukan

—
—
—
—
—

- Met een duidelijk datentrek kunnen node en output node worden gescheiden.

$$(3.4.10)$$

- Menghubungkan titik output node dan hidden node

(24.11)

Ketiga tahap tersebut di atas akan terus berulang (berjalan) sampai diperoleh suatu kondisi yang diinginkan, dimana kriteria ini ditentukan oleh peneliti, dan biasanya berupa nilai SSE atau BIC (*Bayesian Information Criterion*) yang minimal atau jumlah iterasi maksimal.

Dalam kriteria pemilihan model terbaik dapat dilakukan dengan menghitung nilai dari BIC yang dapat dirumuskan sebagai berikut :

Dimana :

n : Banyak observasi efektif dalam estimasi parameter

p : Banyaknya parameter yang ditaksir dalam model

S : Nilai dari SSE (*Sum Square Error*)

Penentuan arsitektur optimal dilakukan dengan *trial and error* sampai dengan jumlah hidden node yang ditentukan peneliti. Berdasarkan prinsipnya maka akan terpilih jumlah hidden node dimana nilai BIC terkecil karena kriteria tersebut sangat peka terhadap besarnya observasi efektif.

Nilai SSE dianggap kurang mewaliki karena nilai SSE cenderung akan semakin kecil dengan semakin bertambahnya jumlah parameter.

2.4.2. Metode Optimasi Error

Metode yang dipakai dalam optimasi error (nilai SSE) adalah metode optimasi orde kedua (matrik Hessian). Metode matrik Hessian ini digunakan dalam algoritma *Quasi-Newton* yang merupakan metode optimasi dengan pendekatan gradien.

Metode *Quasi-Newton* ini didasarkan pada pendekatan lokal kuadratik yang merupakan suatu fungsi error dari deret Taylor tentang $E(w)$ disekitar beberapa titik \hat{w} , sehingga akan didapatkan :

$$E(w) = E(\hat{w}) + (w - \hat{w})^T b + \frac{1}{2} (w - \hat{w})^T H (w - \hat{w}) \quad \dots \quad (2.4.13)$$

dimana b didefinisikan sebagai gradien dari E yang telah dikoreksi oleh w .

Matrik Hessian H didefinisikan sebagai :

$$(H)_{ij} = \frac{\partial^2 E}{\partial w_i \partial w_j} \Big|_w \quad \dots \dots \dots \quad (2.4.15)$$

dari rumusan diatas akan didapatkan suatu hubungan pendekatan lokal dari gradien yang dinyatakan dengan :

Jika $\nabla E = 0$ dan w^* sebagai nilai minimum dari fungsi error, maka dari rumus 2.4.13 dapat dirubah menjadi :

dimana : w^* : nilai minimum dari fungsi error

H : matrik Hessian

g : gradien untuk setiap nilai w

$$g : \nabla E = H(w - w^*) \quad \dots \dots \dots \quad (2.4.19)$$

$H^{-1}g$: newton direction atau newton-step (bentuk dasar dari strategi optimisasi)

2.5. PERBANDINGAN HASIL KLASIFIKASI METODE REGRESI LOGISTIK, KERNEL, DAN ANN.

Dari data testing yang telah dianalisa pengelompokan dengan tiga metode yaitu regresi logistik , kernel dan ANN, maka selanjutnya sebagai kriteria masalah klasifikasi dilihat prosentasi kesalahan pengelompokan yang kecil berdasarkan kriteria MCC (*Maximum Chance criterion*) dan PCC (*proportional chance criterion*).

Ketepatan prediksi awal pengelompokan dengan hasil pengelompokan dengan hasil pengelompokan berdasarkan analisis metode regresi logistik, kernel dan *artificial neural network* dapat dilihat melalui nilai *hit ratio*, yang dirumuskan sebagai berikut :

$$\text{Hit Ratio} = \frac{\text{Jumlah individu yang tepat diklasifikasikan}}{\text{Jumlah individu yang dijadikan sampel}} \times 100\%$$

Akhirnya untuk menentukan apakah pengelompokan usia kerja dalam angkatan kerja *bekerja* atau *tidak bekerja* mempunyai tingkat akurasi yg tinggi, maka guna membuktikan atas hipotesis penelitian, dihitung *Change Model* (C_{pro}), yang dirumuskan sebagai berikut:

$$C_{\text{pro}} = p^2 + q^2 \quad \dots \dots \dots \quad (2.5.1)$$

$$C_{\text{max}} = (n_{\text{max}} / N) \times 100\% \quad \dots \dots \dots \quad (2.5.2)$$

Keterangan :

C_{pro} : *The proportional change criterion*

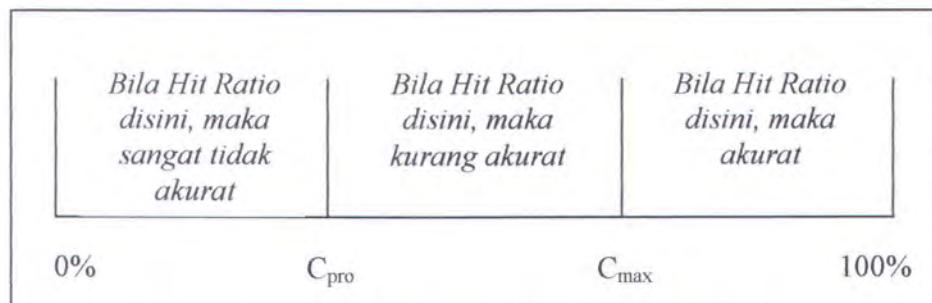
p : Proporsi jumlah sampel yang tepat diklasifikasikan di kelompok 1

q : Proporsi jumlah sampel yang tepat diklasifikasikan di kelompok 2

n_{\max} : Jumlah sampel terbesar pada salah satu kelompok

N : Jumlah sampel secara keseluruhan

Keakuratan pengelompokan (Dwi Budi, 1998) dapat digambarkan pada gambar 2.2 yang menjelaskan bagaimana sistem pengklasifikasian pada tiap kelompok.



Gambar 2.2 Kriteria Keakuratan Pengelompokan

Tahap selanjutnya adalah menguji kestabilan pengelompokan usia kerja dalam angkatan kerja dengan mencobakan kepada penduduk usia kerja diluar analisis sampel, yaitu dengan mengujikannya melalui *Validation Sample* atau *Holdout Sampel* dengan cara menghitung *Press's-Q* (Nareshi K. Maholtra, 1993) yang diformulasikan sebagai berikut :

$$Press's-Q = \frac{(N - (nxK))^2}{N(K-1)} \dots \dots \dots (2.5.3)$$

Keterangan :

N : Total sampel

n : Jumlah individu yang tepat diklasifikasikan

K : Jumlah dari grup (kelompok)

Dengan hipotesis :

H0 : Fungsi tidak stabil dalam pengelompokan

H1 : Fungsi stabil dalam pengelompokan

Nilai *Press's-Q* kemudian dibandingkan dengan nilai *Chi-Square* (χ^2_{k-1})

tabel. Jika nilai *Press's-Q* > (χ^2_{k-1}) maka tolak H0 jadi fungsi tersebut mempunyai kestabilan dalam melakukan pengelompokan, dan sebaliknya jika *Press's-Q* < (χ^2_{k-1}) maka gagal tolak H0 jadi fungsi tersebut tidak mempunyai kestabilan dalam melakukan pengelompokan.

BAB III
METODOLOGI PENELITIAN

BAB III

METODOLOGI PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data Survei Sosial Ekonomi Nasional (Susenas) 2000 yang telah dilakukan oleh BPS. Susenas merupakan survei yang dirancang untuk mengumpulkan data sosial kependudukan yang relatif sangat luas. Data yang dikumpulkan antara lain menyangkut bidang-bidang pendidikan, kesehatan/gizi, perumahan/lingkungan hidup, kriminalitas, kegiatan sosial budaya, konsumsi dan pendapatan rumah tangga, perjalanan, dan pendapat masyarakat mengenai kesejahteraan rumah tangganya.

Pada tahun 1992, sistem pengumpulan data Susenas diperbarui, yaitu informasi yang digunakan untuk menyusun indikator kesejahteraan rakyat (kesra) yang terdapat dalam modul (keterangan yang dikumpulkan tiga tahun sekali) ditarik kedalam KOR (kelompok keterangan yang dikumpulkan tiap tahun). Sejak itu setiap tahun dalam Susenas tersedia perangkat data yang dapat digunakan untuk memantau taraf kesejahteraan masyarakat, merumuskan program pemerintah yang khusus ditujukan untuk meningkatkan kesejahteraan sektor-sektor tertentu dalam masyarakat, dan menganalisis dampak berbagai program peningkatan kesejahteraan penduduk.

Susenas 2000 dilaksanakan diseluruh wilayah Indonesia dengan ukuran sampel 208.672 rumah tangga tersebar diseluruh propinsi baik daerah perkotaan maupun pedesaan. Jumlah rumah tangga yang akan dicacah dengan kuisioner kor

saja adalah 143.008 rumah tangga dan yang dicacah dengan kuisioner kor-modul social, budaya dan pendidikan sebanyak 65.664 rumah tangga. Jumlah rumah tangga yang dicacah dengan kuisioner modul kependudukan dan garam yodium sebanyak 208.672 rumah tangga (BPS, 2000).

Sesuai dengan objek penelitian, dengan menggunakan data Susenas 2000 jumlah populasi penduduk kabupaten Kediri adalah sebanyak 1.397.543 orang dengan jumlah penduduk laki-laki sebanyak 703635 dan penduduk perempuan sebanyak 693908, dari jumlah tersebut didapatkan jumlah angkatan kerja produktif untuk wilayah kabupaten Kediri adalah sebanyak 3539 orang, dan setelah dilakukan seleksi data yang sesuai dengan kriteria skala asumsi-asumsi penelitian, maka data yang digunakan adalah sebanyak 2696 orang

3.2. Variabel Penelitian dan Definisi Operasional

Berdasarkan tinjauan pustaka kependudukan mengenai faktor-faktor yang menjadi latar belakang seseorang berada dalam masa produktif memasuki pasar kerja dan adanya penyesuaian menggunakan data sekunder, maka variabel respon dan variabel prediktor pada penelitian ini terbatas pada data hasil Susenas, yaitu :

1. Variabel respon (Y) : partisipasi seseorang dalam angkatan kerja mempunyai dua kategori jawaban, yaitu :

1 = Berpartisipasi, berarti sedang bekerja ataupun yang punya pekerjaan tetapi sementara tidak aktif bekerja, misal karena cuti, sakit, mogok dan sejenisnya selama seminggu sebelum waktu survei.

0 = Tidak berpartisipasi, berarti yang tidak bekerja ataupun yang sedang mencari kerja selama seminggu sebelum waktu survei.

2. X1= Hubungan dengan kepala rumah tangga, mempunyai 8 kategori jawaban, yaitu :

1 = Kepala rumah tangga (Krt), yaitu seorang dari sekelompok anggota rumah tangga yang bertanggung jawab atas kebutuhan sehari-hari rumah tangga, atau orang yang dianggap ditunjuk sebagai kepala rumah tangga.

2 = Istri / suami dari KRT

3 = Anak

4 = Menantu

5 = Cucu

6 = Orang tua / mertua

7 = Famili lain

8 = Lainnya

3. X2 = Jenis kelamin, yaitu:

1 = laki-laki

2 = perempuan

4. X3 = Umur, menurut BPS 1998 dibagi menjadi :

1 = 10 – 24 tahun

2 = 25 - 44 tahun

3 = 45 – 64 tahun

5. X4 = Status perkawinan, mempunyai tiga kategori jawaban, yaitu :
- 1 = Belum kawin
 - 2 = Kawin, yaitu mempunyai istri / suami pada saat pencacahan, baik tinggal bersama maupun terpisah.
 - 3 = Cerai, yaitu berpisah sebagai suami / istri karena bercerai ataupun ditinggal mati dan belum kawin lagi.
6. X5 = Status kesehatan, mempunyai dua kategori jawaban, yaitu :
- 1 = tidak sakit/sehat, yaitu bila dalam satu bulan terakhir sebelum saat pencacahan tidak mempunyai keluhan kesehatan, antara lain : panas/demam, batuk, pilek, asma, napas sesak/cepat, diare, campak, telinga berair, sakit kuning, sakit kepala berulang, kejang-kejang, lumpuh, kecelakaan, sakit gigi dan lain-lain.
 - 2 = Sakit, yaitu bila dalam satu bulan terakhir sebelum saat pencacahan mempunyai keluhan kesehatan baik yang menyebabkan terganggunya pekerjaan, sekolah atau kegiatan sehari-hari ataupun tidak.
7. X6 = Daerah tempat tinggal, mempunyai dua kategori jawaban yaitu :
- 1 = Kota,
 - 2 = desa
8. X6 = Pendidikan yang ditamatkan, mempunyai empat kategori jawaban, yaitu:
- 1 = tidak tamat SD atau tidak sekolah
 - 2 = SD / sederajat

- 3 = SLTP / sederajat
- 4 ≥ SLTA / sederajat atau diatasnya.

- 9. X7 = Banyaknya anggota rumah tangga
- 10. X8 = Banyaknya anggota rumah tangga yang bekerja
- 11. X9 = Pengeluaran sebulan, menurut “Laporan Eksekutif Susenas BPS (2001)”, dibagi lima kategori, yaitu :

- 1 ≤ Rp. 99.999,-
- 2 = Rp. 100.000 – Rp. 149.999,-
- 3 = Rp. 150.000 – Rp. 199.999,-
- 4 = Rp. 200.000 – Rp. 299.999,-
- 5 ≥ Rp. 300.000,-

Tabel 3.1 Ringkasan Variabel Respon dan Prediktor.

Label	Variabel	Skala data	Keterangan
Y	Partisipasi dalam angkatan kerja	Nominal	1 = bekerja 0 = tidak bekerja
X1	Hubungan dengan kepala RT	Nominal	1 = Kepala RT 2 = Istri/suami 3 = Anak 4 = Menantu 5 = Cucu 6 = Orang tua/mertua 7 = Famili lain 8 = lainnya
X2	Jenis kelamin	Nominal	1 = Laki-laki 2 = Perempuan
X3	Umur	Ordinal	1 = 10 – 24 tahun 2 = 25 – 44 tahun 3 = 45 – 64 tahun
X4	Status perkawinan	Nominal	1 = Belum kawin 2 = Kawin 3 = Cerai
X5	Status kesehatan	Nominal	1 = Sehat 2 = Sakit
X6	Daerah tempat tinggal	Nominal	1 = Kota 2 = Desa
X7	Pendidikan yang ditamatkan	Ordinal	1= tidak tamat SD / tidak sekolah 2 = SD/sederajat 3 = SLTP/sederajat 4 ≥ SLTA/sederajat
X8	Banyaknya anggota RT	Ratio	
X9	Banyaknya ART yang bekerja	Ratio	
X10	Jumlah pengeluaran perkapita sebulan	Ordinal	1 ≤ Rp. 99.999,- 2 = Rp.100.000 - Rp.149.999,- 3 = Rp.150.000 - Rp.199.999,- 4 = Rp. 200.000 - Rp.299.999,- 5 ≥ Rp. 300.000,-

3.3. Pengolahan Data

Sebelum melakukan analisis, data kasus dibagi menjadi dua bagian yaitu training (70%) dan testing (30%). Hal ini dilakukan untuk menguji ketepatan model yang dihasilkan dalam melakukan pengklasifikasian data. Berikut adalah langkah-langkah yang dilakukan untuk analisa :

1. *Deskriptif data*

2. *Tahap metode regresi logistik*

- Identifikasi variabel prediktor
- Penaksiran secara univariat
- Uji signifikansi univariat
- Penaksiran secara multivariate
- Uji signifikansi multivariate
- Pemilihan model terbaik
- Melakukan prediksi dari model terbaik dengan data yang dimodelkan (*training sample*)
- Melakukan prediksi data testing
- Menggunakan kriteria MCC dan PCC dalam pengelompokan.

3. *Tahap metode kernel*

- Pemilihan fungsi kernel
- Pemilihan bandwith optimal
- Ketepatan klasifikasi data training
- Prediksi data testing

- Menggunakan kriteria MCC dan PCC dalam pengelompokan.

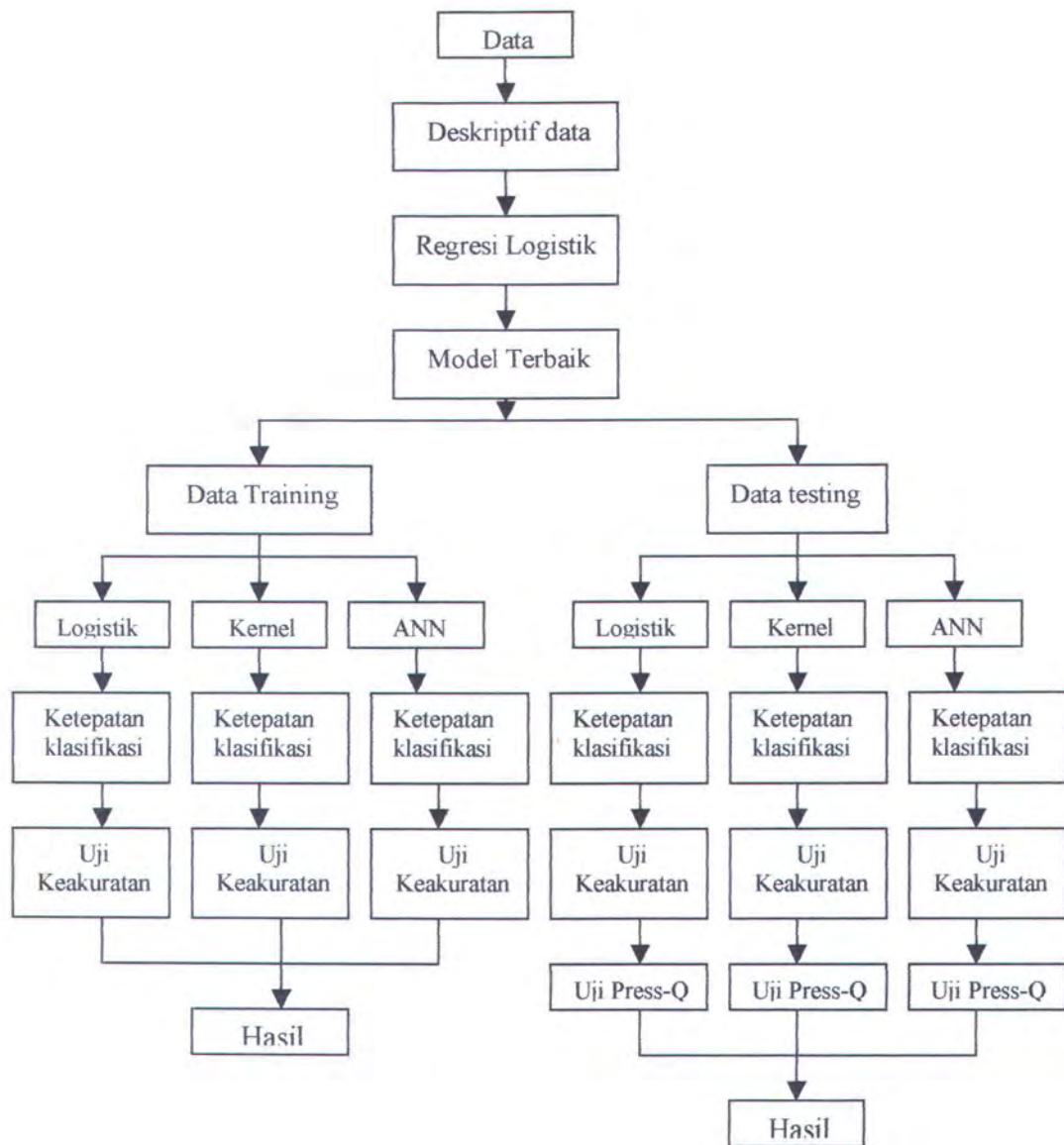
4. Tahap metode ANN

- Penentuan vektor input
- Penentuan jumlah hidden Unit
- Penentuan fungsi aktifasi
- Perhitungan dengan menggunakan pembobot
- Menentukan arsitektur *weight* yang digunakan berdasarkan kriteria BIC
- Tahap *Feed Forward*, tahap *Backpropagation*, dan tahap *Update* nilai bobot
- Hasil pembobotan
- Prediksi kelas dari masing-masing grup (ketepatan pengklasifikasian data training dan testing).
- Menggunakan kriteria MCC dan PCC dalam pengelompokan

5. Tahap Pembandingan metode regresi logistik, kernel dan ANN

- Sebagai pembanding untuk ketiga metode yang digunakan digunakan prosentase kesalahan berdasarkan kriteria MCC dan PCC.
- Melakukan uji kestabilan untuk ketiga metode dengan menggunakan *Press-Q*.

Untuk lebih mempermudah tahapan pembentukan model pada data training dan data testing sehingga diperoleh ketepatan pengelompokan, berikut akan dibuat bagan pengolahan data secara umum.



Gambar 3.1 Bagan Pengolahan Data secara Umum

BAB IV
ANALISA DAN PEMBAHASAN

BAB IV

ANALISA DAN PEMBAHASAN

Berdasarkan data Susenas 2000 untuk kabupaten Kediri, berikut ini adalah hasil pengolahan data penduduk usia kerja dalam angkatan kerja untuk kabupaten Kediri.

4.1. DESKRIPTIF STATISTIK

Selanjutnya akan dilihat berbagai karakteristik para penduduk usia kerja dalam angkatan kerja. Sedangkan batasan bekerja sendiri adalah kegiatan melakukan pekerjaan dengan tujuan memperoleh atau membantu memperoleh nafkah paling sedikit selama satu jam secara terus-menerus selama seminggu yang lalu. Kegiatan bekerja ini mencangkup baik yang sedang bekerja maupun yang punya pekerjaan tetapi dalam seminggu yang lalu sementara tidak aktif bekerja, misal karena cuti, sakit dan sejenisnya (BPS, 2001).

Tabel 4.1 Status Usia Kerja menurut Hubungan dengan Kepala Rumah Tangga

Hubungan dengan KRT	Tidak bekerja	Bekerja	Total
Kepala rumah tangga	72 (9.4%)	692 (90.6 %)	764 (28.3%)
Istri/suami	350 (50.9%)	337 (49.1%)	687 (25.5%)
Anak	725 (70.8%)	299 (29.2%)	1024 (38.0%)
Menantu	23 (30.7%)	52 (69.3 %)	75 (2.8%)
Cucu	51 (73.9%)	18 (26.1%)	69 (2.6%)
Orang tua/mertua	17 (68.0%)	8 (32.0%)	25 (0.9%)
Famili lain	23 (56.1%)	18 (43.9%)	41 (1.5%)
Lainnya	3 (27.3%)	8 (72.7 %)	11 (0.4%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Berdasarkan tabel 4.1 diketahui bahwa dari 2696 penduduk angkatan kerja, yang berstatus kerja adalah sebesar 53,1 %. Sedangkan persentase terbesar yang bekerja adalah berstatus sebagai kepala rumah tangga sebesar 90,6 %. Hal ini disebabkan oleh adanya rasa tanggung jawab kepala rumah tangga untuk menghidupi keluarganya.

Selain itu terlihat pula bahwa pekerja laki-laki lebih banyak daripada perempuan, yaitu sebesar 67,4 % dari jumlah angkatan kerja laki-laki dibandingkan dengan angkatan kerja perempuan yang hanya berkisar 38,3 %, seperti yang ditunjukkan pada tabel 4.2.

Tabel 4.2 Status Usia Kerja menurut Jenis Kelamin

Jenis kelamin	Tidak bekerja	Bekerja	Total
Laki-laki	447 (32.6%)	924 (67.4%)	1371 (50.9%)
Perempuan	817 (61.7%)	508 (38.3%)	1325 (49.1%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Jika dilihat dari kelompok umur seperti pada tabel 4.3, maka dapat diketahui bahwa jumlah penduduk usia kerja yang terbesar persentasenya adalah kelompok umur 25-44 tahun yaitu sebesar 39,9 %. Akan tetapi jika dilihat berdasarkan status kerja, maka yang terbesar adalah pada kelompok umur 45-64 tahun yaitu sebesar 71,7 %.

Status perkawinan sangat mempengaruhi seseorang untuk melakukan aktivitas ekonomi, karena selain bertujuan untuk menghidupi dirinya, orang tersebut harus menghidupi keluarga yang menjadi tanggung jawabnya. Seperti pada tabel 4.4 terlihat bahwa persentase terbesar yang berstatus kerja adalah mereka yang masih terikat perkawinan, yaitu sebesar 69,1 %.

Tabel 4.3 Status Usia Kerja menurut Kelompok Umur.

Kelompok umur	Tidak bekerja	Bekerja	Total
10 – 24 tahun	759 (77.8%)	217 (22.2%)	976 (36.2%)
25 – 44 tahun	323 (30.0%)	753 (70.0%)	1076 (39.9%)
45 – 64 tahun	182 (28.3%)	462 (71.7%)	644 (23.9%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Tabel 4.4 Status Usia Kerja menurut Status Perkawinan.

Status Kawin	Tidak bekerja	Bekerja	Total
Belum kawin	731 (73.5%)	263 (26.5%)	994 (36.9%)
Kawin	478 (30.9%)	1070 (69.1%)	1548 (57.4%)
Cerai	55 (35.7%)	99 (64.3%)	154 (5.7%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Tabel 4.5 Status Usia Kerja menurut Status Kesehatan.

Status kesehatan	Tidak bekerja	Bekerja	Total
Sehat	982 (47.8%)	1071 (52.2%)	2053 (76.1%)
Sakit	282 (43.9%)	361 (56.1%)	643 (23.9%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Pekerja umumnya ditunjang dengan kondisi kesehatan. Namun kenyataannya tidak tertutup kemungkinan ditemukan pekerja yang tidak sehat tapi tetap bekerja seperti pada tabel 4.5, terdapat 56,1 % penduduk usia kerja yang sakit tetapi masih tetap bekerja.

Tabel 4.6 Status Usia Kerja menurut Tempat Tinggal.

Tempat tinggal	Tidak bekerja	Bekerja	Total
Kota	438 (50.7%)	426 (49.3%)	864 (32.0%)
Desa	826 (45.1%)	1006 (54.9%)	1832 (68.0%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Jumlah terbesar penduduk angkatan kerja yang berstatus kerja ternyata tinggal di daerah pedesaan, seperti pada tabel 4.6 sebesar 54,9 %.

Berdasarkan pendidikan yang ditamatkan, semakin tinggi tingkat pendidikan maka persentase yang bekerja cenderung semakin tinggi. Hal ini mungkin berkenaan dengan keahlian yang dimilikinya, seperti pada tabel 4.7 terdapat 62,0 % penduduk angkatan kerja yang berstatus kerja dengan latar belakang pendidikan minimal SLTA.

Tabel 4.7 Status Usia Kerja menurut Tingkat Pendidikan.

Tingkat pendidikan	Tidak bekerja	Bekerja	Total
Tdk sekolah/tdk tamat SD	386 (48.5%)	410 (51.5%)	796 (29.5%)
Tamat SD/sederajat	472 (45.7%)	561 (54.3%)	1033 (38.3%)
Tamat SLTP/sederajat	237 (56.2%)	185 (43.8%)	422 (15.7%)
≥ SLTA/sederajat	169 (38.0%)	276 (62.0%)	445 (16.5%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Tabel 4.8 Status Usia Kerja menurut Jumlah Anggota Rumah Tangga.

Jumlah ART	Tidak bekerja	Bekerja	Total
1	4 (19.0%)	17 (81.0%)	21 (0.8%)
2	55 (27.4%)	146 (72.6%)	201 (7.5%)
3	226 (42.2%)	309 (57.8%)	535 (19.8%)
4	330 (45.4%)	397 (54.6%)	727 (27.0%)
5	285 (50.4%)	280 (49.6%)	565 (21.0%)
6	198 (55.0%)	162 (45.0%)	360 (13.4%)
7	88 (57.9%)	64 (42.1%)	152 (5.6%)
8	43 (50.6%)	42 (49.4%)	85 (3.2%)
9	20 (69.0%)	9 (31.0%)	29 (1.1%)
10	15 (71.4%)	6 (28.6%)	21 (0.8%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Dari tabel 4.8 terlihat adanya kecenderungan semakin banyak anggota rumah tangga maka yang berstatus kerja semakin sedikit. Hal ini ditunjang dengan nilai persentase terbesar yaitu 81% bagi mereka yang memiliki 1 orang anggota

rumah tangga. Hal ini dimungkinkan karena ia adalah suami/istri yang baru saja menikah, sehingga untuk menghidupi keluarga maka harus bekerja.

Tuntutan ekonomi dan keinginan untuk hidup lebih baik mendorong seseorang untuk bekerja. Hal ini terlihat dari tabel 4.9, terjadi kecenderungan kenaikan persentase penduduk yang bekerja sesuai dengan kenaikan jumlah anggota rumah tangga yang bekerja dan persentase terbesar untuk jumlah anggota rumah tangga yang bekerja dalam satu keluarga adalah 5 orang, yaitu sebesar 88,6%. Selain itu jumlah anggota rumah tangga yang bekerja paling dominan berjumlah dua orang dalam satu keluarga, yaitu sebesar 38 % dari total penduduk yang disurvei.

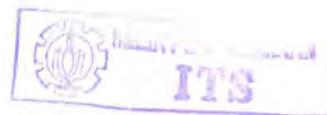
Tabel 4.9 Status Usia Kerja menurut Jumlah Anggota Rumah Tangga yang Bekerja.

Jml ART bekerja	Tidak bekerja	Bekerja	Total
0	104 (100%)	0	104 (3.9%)
1	608 (64.5%)	334 (35.5%)	942 (34.9%)
2	392 (38.2%)	633 (61.8%)	1025 (38.0%)
3	111 (27.9%)	287 (72.1%)	398 (14.8%)
4	45 (23.4%)	147 (76.6%)	192 (7.1%)
5	4 (11.4%)	31 (88.6%)	35 (1.3%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Tabel 4.10 Status Usia Kerja menurut Pengeluaran Perkapita Sebulan.

Pengeluaran perkapita/sebulan	Tidak bekerja	Bekerja	Total
≤ Rp. 99.999,-	873 (49.6%)	888 (50.4%)	1761 (65.3%)
Rp.100.000-Rp.149.999,-	266 (40.3%)	394 (59.7%)	660 (24.5%)
Rp.150.000-Rp.199.999,-	89 (47.8%)	97 (52.2%)	186 (6.9%)
Rp.200.000-Rp.299.999,-	19 (38.8%)	30 (61.2%)	49 (1.8%)
≥ Rp. 300.000,-	17 (42.5%)	23 (57.5%)	40 (1.5%)
Total	1264 (46.9%)	1432 (53.1%)	2696 (100%)

Untuk wilayah kabupaten Kediri ternyata persentase terbesar pengeluaran perkapita/sebulan berada pada kisaran Rp. 99.999,- kebawah, yaitu sebesar 65,3%. Sedangkan bagi mereka yang bekerja maka persentase terbesar pengeluaran adalah berkisar Rp.200.000-Rp.299.999,- yaitu sebanyak 61,2 % seperti ditunjukkan pada tabel 4.10.



4.2. METODE REGRESI LOGISTIK

4.2.1. Metode Regresi Logistik Data Training

4.2.1.1. Pemodelan Univariat

Pada tahap univariat ini penaksiran parameter dilakukan untuk tiap-tiap variabel prediktor. Kemudian hasil estimasi tiap-tiap parameter β_1 akan diuji secara individual. Pada pengujian univariat dan multivariate penelitian ini α yang digunakan adalah 0.25 berdasarkan dari penelitian yang telah dilakukan oleh Mickey dan Greenland pada regresi logistik, dan oleh Bendel dan Afifi di regresi linier (Hosmer, 1989). Para peneliti tersebut memperlihatkan bahwa penggunaan $\alpha = 0.05$ yang biasa dipakai sering kali gagal untuk mengidentifikasi variabel yang dianggap penting untuk data sosial. Karena itu nilai α yang digunakan ditoleransi hingga 25%.

Dari model regresi logistik secara univariat dapat diketahui bahwa semua variabel prediktor signifikan dalam model dengan menggunakan metode *Backward Wald* seperti pada lampiran 2, akan tetapi jika diuji secara parameter dengan nilai p-value sebesar 0.25 seperti pada tabel 4.11, maka didapatkan semua variabel juga signifikan.

Tabel 4.11 Hasil Metode Regresi Logistik Univariat

Variabel	B	SE	Wald	df	Sig	Exp (B)
X1			390.7825	7	0.0000	
X1 (1)	1.1433	1.1641	0.9647	1	0.3260 *	3.1372
X1 (2)	-1.0747	1.1581	0.8611	1	0.3534 *	0.3414
X1 (3)	-2.0336	1.1577	3.0857	1	0.0790	0.1309
X1 (4)	0.5108	1.2267	0.1734	1	0.6771 *	1.6667
X1 (5)	-2.4849	1.2134	4.1942	1	0.0406	0.0833
X1 (6)	-2.6391	1.3184	4.0070	1	0.0453	0.0714
X1 (7)	-1.0380	1.2061	0.7407	1	0.3894 *	0.3542
Constant	1.0987	1.5147	0.9052	1	0.3414 *	
X2 (1)	1.0653	0.0957	124.0320	1	0.000	2.9017
constant	-0.3915	0.0659	35.3090	1	0.000	
X3			395.0506	2	0.000	
X3 (1)	-2.3548	0.1405	281.0976	1	0.000	0.0949
X3 (2)	-0.1262	0.1310	0.9272	1	0.3356 *	0.8815
constant	0.9924	0.1031	92.6862	1	0.000	
X4			319.1660	2	0.000	
X4 (1)	-1.6190	0.2013	64.0261	1	0.000	0.1997
X4 (2)	0.3554	0.1930	3.3924	1	0.0655	1.4267
constant	0.5026	0.1810	7.7132	1	0.0055	
X5 (1)	-0.2280	0.1101	4.2864	1	0.0384	0.7962
constant	0.2997	0.0967	9.5961	1	0.0019	
X6 (1)	-0.2196	0.0980	5.0204	1	0.0251	0.8029
constant	0.1971	0.0566	12.144	1	0.0005	
X7			32.1991	3	0.000	
X7 (1)	-0.2871	0.1420	4.0879	1	0.0432	0.7504
X7 (2)	-0.1555	0.1374	1.2801	1	0.2579 *	0.8560
X7 (3)	-0.8628	0.1661	26.9996	1	0.000	0.4220
constant	0.4028	0.1146	12.3524	1	0.000	
X8	-0.1619	0.0290	31.1180	1	0.000	0.8505
constant	0.8462	0.1376	37.8342	1	0.000	
X9	0.7856	0.0571	189.5228	1	0.000	2.1937
constant	-1.3273	0.1133	137.1204	1	0.000	
X10			23.4293	4	0.0001	
X10 (1)	-0.5942	0.3985	2.2237	1	0.1359	0.5520
X10 (2)	-0.1408	0.4057	0.1204	1	0.7286 *	0.8687
X10 (3)	0.6533	0.4340	2.2660	1	0.1322	0.5203
X10 (4)	0.2720	0.5338	0.2596	1	0.6104 *	1.3126
constant	0.5877	0.3944	2.2208	1	0.1362	

Ket : * Tidak signifikan pada p-value 25%

4.2.1.2. Pemodelan Multivariat

Selanjutnya pada pengujian multivariat variabel prediktor yang digunakan adalah variabel yang signifikan pada pengujian individu Untuk mencari model terbaik. Berdasarkan hasil pengolahan data (lampiran 3) diketahui nilai chi-squarenya sebesar 1325.783 dengan p-value=0.000 dan nilai ini kurang dari level 0.25. Sehingga dapat dinyatakan bahwa ada satu atau lebih variabel prediktor yang berpengaruh nyata terhadap variabel respon. Kemudian dari hasil pengolahan data yang ada pada tabel 4.12 akan diteliti kembali variabel prediktor mana saja yang signifikan pada level 25%.

Berdasarkan tabel 4.12 diatas diketahui ada 3 varibel prediktor yang tidak signifikan pada level 0.25 yaitu variabel status kesehatan (X_5), daerah tempat tinggal (X_6) dan pengeluaran perkapita/sebulan (X_{10}). Untuk mengetahui apakah model tanpa ketiga variabel tersebut adalah model terbaik maka akan dilakukan uji kesesuaian model. Hipotesa yang digunakan adalah :

H_0 : Model tanpa variable tertentu adalah model terbaik

H_1 : Model dengan variable tertentu adalah model terbaik

Diketahui bahwa dari tabel 4.13 nilai dari p-value sebesar 0.6116 yang berarti tidak signifikan pada $\alpha = 25\%$ yang berarti model tanpa variabel X_{10} adalah model terbaik, tapi masih ada dua lagi variabel yang tidak signifikan yaitu variabel X_5 dan X_6 sehingga harus dikeluarkan.

Tabel 4.12 Hasil Metode Regresi Logistik Multivariat.

Variabel	B	SE	Wald	df	Sig	Exp (B)
X1			72.5862	7		
X1 (1)	0.25896	1.7780	2.1213	1	0.000	13.3243
X1 (2)	0.4481	1.7687	0.0645	1	0.1453	1.5669
X1 (3)	0.4707	1.7411	0.0731	1	0.7996 *	1.6011
X1 (4)	2.1447	1.8426	1.3548	1	0.7869 *	8.5397
X1 (5)	0.0926	1.8151	0.0026	1	0.2444	1.0971
X1 (6)	-0.4365	1.9660	0.0493	1	0.8243 *	0.6463
X1 (7)	0.2646	1.8037	0.0215	1	0.8834 *	1.3029
X2 (1)	1.5228	0.2145	50.4051	1	0.000	4.5853
X3			84.2648	2	0.000	
X3 (1)	-1.7193	0.2971	33.4880	1	0.000	0.1792
X3 (2)	0.5227	0.2006	6.7887	1	0.092	1.6866
X4			19.3715	2	0.0001	
X4 (1)	-0.5372	0.4322	1.5450	1	0.2139	0.5844
X4 (2)	0.7711	0.3438	5.0305	1	0.0249	2.1611
X5 (1)	0.1591	0.1743	0.8334	1	0.3613 *	1.1724
X6 (1)	0.1126	0.1563	0.5193	1	0.4711 *	1.1192
X7			7.4293	3	0.0594	
X7 (1)	-0.6099	0.2352	6.7251	1	0.0095	0.5434
X7 (2)	-0.2713	0.2153	1.5870	1	0.2078	0.7624
X7 (3)	-0.4223	0.2522	2.8029	1	0.0941	0.6555
X8	-0.5398	0.0565	91.3214	1	0.000	0.5829
X9	1.9265	0.1120	295.8986	1	0.000	6.8655
X10			2.6948	4	0.6101 *	
X10 (1)	0.5521	0.5485	1.0131	1	0.3142 *	1.7369
X10 (2)	0.4946	0.5555	0.7927	1	0.3733 *	1.6398
X10 (3)	0.8481	0.5990	2.0051	1	0.1568	2.3353
X10 (4)	0.7947	0.7272	1.1941	1	0.2745 *	2.2137
constant	-2.9325	1.9055	2.3685	1	0.1238	

Ket : * Tidak signifikan pada p-value 25%

Tabel 4.13 Hasil Pemodelan Metode Regresi Logistik Multivariat.

Step	Variabel	-2log likelihood	df	G ²	P-value
1	Semua variabel	1282.895	23		0.000
2	Tanpa variable X10	1285.582	4	2.686	0.6116
3	Tanpa variable X6	1286.040	1	0.458	0.4986
4	Tanpa variable X5	1286.951	1	0.911	0.3398

Dengan mengeluarkan variabel X6 dan didapatkan p-value = 0.4986 yang berarti tidak signifikan pada $\alpha = 25\%$ yang berarti model tanpa variabel X6 adalah model terbaik, demikian pula pada variabel X10 didapatkan nilai P-value = 0.3898 yang berarti tidak signifikan pada $\alpha = 25\%$ yang berarti model tanpa variabel X10 adalah model terbaik selanjutnya model tersebut diebut model efek utama (*Main Effect Model*) seperti pada tabel 4.14.

Tabel 4.14 Hasil Akhir Metode Regresi Logistik Multivariat.

Variabel	B	SE	Wald	df	Sig	Exp (B)
X1			72.0051	7	0.000	
X1 (1)	2.5363	1.7571	2.0835	1	0.1489	12.6329
X1 (2)	0.4183	1.7478	0.0573	1	0.8109	1.5193
X1 (3)	0.4583	1.7222	0.0708	1	0.7902	1.5813
X1 (4)	2.1266	1.8244	1.3587	1	0.2438	8.3865
X1 (5)	0.1038	1.7965	0.0033	1	0.9539	1.1093
X1 (6)	-0.5351	1.9531	0.0751	1	0.7841	0.5856
X1 (7)	0.2374	1.7843	0.0177	1	0.8942	1.2679
X2 (1)	1.5159	0.2131	50.6135	1	0.000	4.5534
X3			83.3427	2	0.000	
X3 (1)	-1.7154	0.2960	33.5968	1	0.000	0.1799
X3 (2)	0.4900	0.1993	6.0455	1	0.0139	1.6324
X4			19.4389	2	0.0001	
X4 (1)	-0.5728	0.4313	1.7637	1	0.1842	0.5640
X4 (2)	0.7381	0.3413	4.6768	1	0.0306	2.0919
X7			9.3621	3	0.0248	
X7 (1)	-0.6690	0.2235	8.9609	1	0.0028	0.5122
X7 (2)	-0.3320	0.2061	2.5953	1	0.1072	0.7175
X7 (3)	-0.4625	0.2478	3.4834	1	0.0620	0.6297
X8	-0.5285	0.0536	97.2195	1	0.000	0.5895
X9	1.9059	0.1098	301.1927	1	0.000	6.7252
constant	-2.1068	1.7968	1.3748	1	0.2410	

Interpretasi model dilihat dari nilai *odds ratio*. Misal untuk variabel jenis kelamin, rasio kecenderungan penduduk usia kerja perempuan adalah 4.5534 kali.

Artinya kecenderungan untuk bekerja bagi perempuan adalah 4.5534 kali lebih kecil dari penduduk usia kerja laki-laki. Demikian berlaku untuk variabel prediktor kategorikal lainnya. Sedangkan untuk variabel prediktor yang kontinyu, interpretasi koefisien estimasi tergantung bagaimana variabel ini diperlakukan dalam model. Pada variabel banyaknya anggota rumah tangga yang bekerja, misal terdapat satu orang anggota rumah tangga yang bekerja, maka *odds rasio* = 6.7252. Hal ini berarti untuk setiap peningkatan satu anggota rumah tangga yang bekerja, maka orang akan cenderung bekerja 6.7252 kali.

Berdasarkan tabel 4.14 terlihat semua variabel yang ada masuk dalam model. Maka model terbaik untuk menggambarkan hubungan antara variabel prediktor dengan variabel respon dalam bentuk logit sehingga menjadi persamaan yang linier yaitu :

$$\begin{aligned}\hat{g(x)} = & -2.1068 + 2.5363X1(1) + 0.4183X1(2) + 0.4583X1(3) + 2.1266X1(4) + \\& + 0.1038X1(5) - 0.5351X1(6) + 0.2374X1(7) - 1.5159X2(1) - 1.7154X3(1) \\& + 0.4900X3(2) + 0.5728X4(1) + 0.7381X4(2) + - 0.6690X7(1) + \\& - 0.3320X7(2) - 0.4625X7(3) - 0.5285X8 + 1.90593X9.\end{aligned}$$

4.2.1.3 Ketepatan Klasifikasi

Analisa selanjutnya adalah melihat ketepatan klasifikasi pada data training untuk regresi logistik dengan menggunakan 7 variabel terpilih. Didapatkan nilai ketepatan klasifikasi sebesar 84.68 % dengan nilai misklasifikasi sebesar 15.32%. Pada model regresi lengkap ada 143 observasi nilai $y = '0'$ yang salah

diprediksikan pada $y = '1'$ dan ada 146 observasi nilai $y = '1'$ yang salah diprediksikan pada $y = '0'$. Seperti yang terlihat pada tabel dibawah ini.

Tabel 4.15 Hasil Klasifikasi Metode Regresi Logistik Data Training.

Observasi	Prediksi		Percentase benar
	0	1	
0	742	143	83.84 %
1	146	856	85.43 %
Ketepatan klasifikasi keseluruhan			84.68 %

4.2.1.4 Uji Keakuratan

Selanjutnya untuk memperoleh konsistensi hasil pengujian, pada analisa logistik dilakukan dengan menguji model logistik pada sampel validasi (*Holdout Sampel*) yang terdiri dari 1887 data berdasarkan tabel 4.15. Hal ini guna menentukan apakah pengelompokan usia kerja dalam angkatan kerja mempunyai tingkat akurasi yang tinggi. Dalam penelitian ini digunakan metode *Maximum Chance Criterion (Cmax)* dan *Proportional Chance Criterion (Cpro)* yang diformulasikan sebagai berikut:

$$\text{Bekerja} \quad = p = 856/1887 = 0.4536$$

$$\text{Tidak bekerja} \quad = q = 742/1887 = 0.3932$$

$$C_{pro} = p^2 + q^2 = (0.4536)^2 + (0.3932)^2 = 0.3787 = 37.87 \%$$

$$C_{max} = (n_{\max} / N) \times 100\% = 856/1887 \times 100\% = 45.36 \%$$

Dihasilkan nilai C_{pro} sebesar 37.87 % dan C_{max} sebesar 45.36%, oleh karena nilai *Hit Ratio* sebesar 84.68 % yang berarti lebih besar dibandingkan C_{pro}

maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode regresi logistik adalah akurat.

4.2.2 Metode Regresi Logistik Data Testing

4.2.2.1 Ketepatan Klasifikasi

Data yang digunakan pada analisa testing adalah data yang berdasarkan variabel terpilih saja pada metode regresi logistik seperti halnya juga pada data training, dimana terdapat 7 variabel prediktor yang signifikan terhadap model.

Hasil ketepatan klasifikasi data testing dapat dilihat pada tabel 4.16 Dimana terdapat 65 observasi nilai $y = '0'$ yang salah diprediksikan pada $y = '1'$ dan ada 53 observasi nilai $y = '1'$ yang salah diprediksikan pada $y = '0'$ sehingga didapatkan nilai ketepatan klasifikasi sebesar 85.41 % dengan nilai misklasifikasi sebesar 14.59 %.

Tabel 4.16 Hasil Klasifikasi Metode Regresi Logistik Data Testing.

Observasi	Prediksi		Percentase benar
	0	1	
0	314	65	82.85 %
1	53	377	87.67 %
Ketepatan klasifikasi keseluruhan			85.41 %

4.2.2.2 Uji Keakuratan dan Kestabilan.

Untuk memperoleh konsistensi hasil pengujian, pada analisa logistik dilakukan dengan menguji model logistik pada sampel validasi (*Holdout Sampel*) yang terdiri dari 809 data berdasarkan tabel 4.16. Hal ini guna menentukan

apakah pengelompokan usia kerja dalam angkatan kerja mempunyai tingkat akurasi yang tinggi. Dalam penelitian ini digunakan metode *Maximum Chance Criterion (Cmax)* dan *Proportional Chance Criterion (Cpro)* yang diformulasikan sebagai berikut:

$$\text{Bekerja} = p = 377/809 = 0.4660$$

$$\text{Tidak bekerja} = q = 314/809 = 0.3881$$

$$C_{pro} = p^2 + q^2 = (0.4660)^2 + (0.3881)^2 = 0.3678 = 36.78 \%$$

$$C_{max} = (n_{max} / N) \times 100\% = 377/809 \times 100\% = 46.60 \%$$

Dihasilkan nilai Cpro sebesar 36.78 % dan Cmax sebesar 46.60 %, oleh karena nilai *Hit Ratio* sebesar 85.41 % yang berarti lebih besar dibandingkan Cpro maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode regresi logistik adalah akurat.

Pengujian selanjutnya yaitu menguji tingkat kestabilan dari fungsi regresi logistik, yaitu dengan menghitung tingkat kestabilan pengalokasian, hal ini bertujuan untuk mengkaji apakah ada kemungkinan pengalokasian dari tiap sampel dalam kelompok relatif stabil atau tidak sebagai akibat adanya perubahan perbedaan jumlah sampel yang diteliti.. Dengan menghitung nilai *Press's-Q*.

$$press's-Q = \frac{[N - (nxk)]^2}{N(k-1)} = \frac{[809 - (691x2)]^2}{809(2-1)}$$

$$= 405.85$$

Nilai *Press's-Q* = 405.85 dibandingkan ($\chi^2_{0.05,1}$) = 3.841 yang berarti bahwa nilai *Press's-Q* > ($\chi^2_{0.05,1}$) sehingga kesimpulannya adalah fungsi logistik tersebut memiliki kestabilan dalam melakukan pengelompokan.

4.3. METODE KERNEL

4.3.1 Pemilihan Fungsi Kernel dan Penentuan Bandwith Optimal

Metode Kernel merupakan metode yang tergantung pada dua hal, yaitu pemilihan fungsi kernel (K) dan penentuan besarnya *bandwidth* (h). Fungsi Kernel terdiri dari tujuh macam diantaranya *Uniform*, *Triangle*, *Epannechnikov*, *Quartic/Biweight*, *Triweight*, *Gaussian/Normal*, dan *Cosinus* (Härdle, 1990). Fungsi kernel *Uniform* berbentuk histogram, diskontinu pada -1 dan 1 serta tidak dapat diturunkan pada x_i-h dan x_i+h . Fungsi kernel *Triangle* berbentuk segitiga, kontinu tetapi tidak dapat diturunkan pada x_i-h dan x_i+h . Fungsi kernel *Epannechnikov*, *Quartic/Biweight*, *Triweight*, *Gaussian/Normal*, dan *Cosinus* lebih halus dibandingkan dengan *Uniform* dan *Triangle*, tetapi yang menghasilkan taksiran paling bagus adalah fungsi kernel *Gaussian/Normal*. Perhitungan kernel *Epannechnikov* lebih mudah, tetapi tidak dapat diturunkan beberapa kali.

Berdasarkan penjelasan tentang berbagai macam fungsi kernel tersebut, maka untuk analisis dengan regresi kernel pada kasus pengelompokan usia kerja dalam angkatan kerja digunakan fungsi kernel Normal dimana bentuknya adalah sebagai berikut :

$$K_t(z) = \frac{1}{C_0(t)} \exp(-0.5 z_t^{-1} z/h^2)$$

$$\text{dimana } C_0(t) = (2\pi)^{p/2} h^p |V_t|^{1/2}$$

Setelah fungsi kernel diketahui, selanjutnya ditentukan nilai parameter penghalusan yang optimal dimana pada penelitian ini menggunakan metode *cross-validation* untuk penentuan nilai *bandwidth*. Berdasarkan persamaan (2.3.10),

dimana konstanta optimal $A(K_t)$ tergantung pada kernel K_t . Dengan kernel Normal ditentukan nilai $A(K_t)$ sebesar :

$$A(K_t) = \frac{4}{2p+1} = \frac{4}{2.2+1} = 0.8$$

Nilai itu disubtitusikan ke dalam persamaan (2.3.9), sehingga diperoleh nilai h yang optimum adalah:

$$\begin{aligned} h &= A(K_t)/n_t^{1/(p+4)} \\ &= \left(\frac{0.8}{1432} \right)^{\left(\frac{1}{2+4} \right)} \\ &= 0.287 \end{aligned}$$

Untuk penelitian dengan metode kernel ini menggunakan 2 macam data yaitu data training dan data testing (7 variabel terpilih berdasarkan metode regresi logistik).

4.3.2 Analisa Data Training Metode Kernel

4.3.2.1 Ketepatan Klasifikasi Asumsi Varians Sama

Dengan asumsi menggunakan nilai varians sama, maka untuk data training didapatkan bahwa untuk kelompok usia kerja yang tidak bekerja tepat diklasifikasikan sebanyak 799 orang dan yang salah diklasifikasikan sebanyak 86 orang. Untuk kelompok Usia kerja yang bekerja tepat diklasifikasikan sebanyak 904 orang dan yang salah diklasifikasikan sebanyak 98 orang, seperti pada tampilan tabel 4.17.

Tabel 4.17 Hasil Klasifikasi Metode Kernel Asumsi Varians Sama Data Training.

Observasi	Prediksi		Percentase benar
	0	1	
0	799	86	90.28 %
1	98	904	90.22 %
Ketepatan klasifikasi keseluruhan			90.25 %

Hasil klasifikasi tersebut memberikan persentase kesalahan pengklasifikasian sebesar 9.75 % dan ketepatan pengklasifikasian sebesar 90.25 %.

4.3.2.2 Uji Keakuratan Asumsi Varians sama

Selanjutnya untuk untuk memperoleh konsistensi hasil pengujian, pada analisa kernel data training asumsi varians sama dilakukan dengan menguji metode kernel yang terdiri dari 1887 data berdasarkan tabel 4.17. Hal ini guna menentukan apakah pengelompokan usia kerja dalam angkatan kerja mempunyai tingkat akurasi yang tinggi. Dalam penelitian ini digunakan metode *Maximum Chance Criterion (Cmax)* dan *Proportional Chance Criterion (Cpro)* yang diformulasikan sebagai berikut:

$$\text{Bekerja} = p = 904/1887 = 0.4791$$

$$\text{Tidak bekerja} = q = 799/1887 = 0.4234$$

$$Cpro = p^2 + q^2 = (0.4791)^2 + (0.4234)^2 = 0.4088 = 40.88\%$$

$$Cmax = (n_{\max} / N) \times 100\% = 904/1887 \times 100\% = 47.91\%$$

Dihasilkan nilai Cpro sebesar 40.88% dan Cmax sebesar 47.91%, oleh karena nilai *Hit Ratio* sebesar 90.25% yang berarti lebih besar dibandingkan Cpro

maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode kernel adalah akurat.

4.3.2.3 Ketepatan Klasifikasi Asumsi Varians Tidak Sama

Untuk asumsi menggunakan nilai varians tidak sama, maka untuk data training didapatkan bahwa untuk kelompok usia kerja yang tidak bekerja tepat diklasifikasikan sebanyak 782 orang dan yang salah diklasifikasikan sebanyak 103 orang. Untuk kelompok Usia kerja yang bekerja tepat diklasifikasikan sebanyak 888 orang dan yang salah diklasifikasikan sebanyak 114 orang, seperti pada tampilan tabel 4.18.

Tabel 4.18 Hasil Klasifikasi Metode Kernel Asumsi Varians Tidak Sama Data Training.

Observasi	Prediksi		Percentase benar
	0	1	
0	782	103	90.06 %
1	114	888	87.82 %
Ketepatan klasifikasi keseluruhan			88.87 %

Hasil klasifikasi tersebut memperoleh persentase kesalahan pengklasifikasian sebesar 11.13% dan ketepatan pengklasifikasian sebesar 88.87%.

4.3.2.4 Uji Keakuratan Asumsi Varians Tidak sama

Selanjutnya untuk untuk memperoleh konsistensi hasil pengujian, pada metode kernel data training asumsi nilai varians tidak sama dilakukan dengan menguji data training yang terdiri dari 1887 data berdasarkan tabel 4.18. Hal ini

untuk menentukan apakah pengelompokan usia kerja dalam angkatan kerja mempunyai tingkat akurasi yang tinggi. Dalam penelitian ini digunakan metode *Maximum Chance Criterion (Cmax)* dan *Proportional Chance Criterion (Cpro)* yang diformulasikan sebagai berikut:

$$\text{Bekerja} = p = 888/1887 = 0.4706$$

$$\text{Tidak bekerja} = q = 782/1887 = 0.4144$$

$$Cpro = p^2 + q^2 = (0.4706)^2 + (0.4144)^2 = 0.3932 = 39.32 \%$$

$$Cmax = (n_{\max} / N) \times 100\% = 888/1887 \times 100\% = 47.06 \%$$

Dihasilkan nilai Cpro sebesar 39.32% dan Cmax sebesar 47.06%, oleh karena nilai *Hit Ratio* sebesar 88.87% yang berarti lebih besar dibandingkan Cpro maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode kernel data training adalah akurat.

4.3.3 Metode Kernel Data Testing

4.3.3.1 Ketepatan Klasifikasi Asumsi Varians Sama

Dengan asumsi menggunakan nilai variance sama, maka untuk data testing didapatkan bahwa untuk kelompok usia kerja yang tidak bekerja tepat diklasifikasikan sebanyak 330 orang dan yang salah diklasifikasikan sebanyak 49 orang. Untuk kelompok Usia kerja yang bekerja tepat diklasifikasikan sebanyak 381 orang dan yang salah diklasifikasikan sebanyak 48 orang, seperti pada tampilan tabel 4.19.

Tabel 4.19 Hasil Klasifikasi Metode Kernel Asumsi Varians Sama Data Testing.

Observasi	Prediksi		Percentase benar
	0	1	
0	330	49	87.07 %
1	48	381	88.81 %
Ketepatan klasifikasi keseluruhan			88.00 %

Hasil klasifikasi tersebut memberikan persentase kesalahan pengklasifikasian sebesar 12.00 % dan ketepatan pengklasifikasian sebesar 88.00 %.

4.3.3.2 Uji Keakuratan dan Kestabilan Asumsi Varians Sama

Untuk memperoleh konsistensi hasil pengujian, pada metode kernel dilakukan dengan menguji pada sampel validasi (data testing) yang terdiri dari 808 data berdasarkan tabel 4.19 (asumsi nilai varians sama). Hal ini guna menentukan apakah pengelompokan usia kerja dalam angkatan kerja mempunyai tingkat akurasi yang tinggi. Dalam penelitian ini digunakan metode *Maximum Chance Criterion (Cmax)* dan *Proportional Chance Criterion (Cpro)* yang diformulasikan sebagai berikut:

$$\text{Bekerja} = p = 381/808 = 0.4715$$

$$\text{Tidak bekerja} = q = 330/808 = 0.4048$$

$$C_{\text{pro}} = p^2 + q^2 = (0.4715)^2 + (0.4048)^2 = 0.3891 = 38.91\%$$

$$C_{\text{max}} = (n_{\text{max}} / N) \times 100\% = 381/808 \times 100\% = 47.15\%$$

Dihasilkan nilai Cpro sebesar 38.91% dan Cmax sebesar 47.15%, oleh karena nilai *Hit Ratio* sebesar 88.00 % yang berarti lebih besar dibandingkan Cpro

maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode kernel adalah akurat.

Pengujian selanjutnya yaitu menguji tingkat kestabilan dari fungsi kernel yang dihasilkan, yaitu dengan menghitung tingkat kestabilan pengalokasian, hal ini bertujuan untuk mengkaji apakah ada kemungkinan pengalokasian dari tiap sampel dalam kelompok relatif stabil atau tidak sebagai akibat adanya perubahan perbedaan jumlah sampel yang diteliti.. Dengan menghitung nilai *Press's-Q*.

$$press's-Q = \frac{[N - (nxk)]^2}{N(k-1)} = \frac{[808 - (711x2)]^2}{808(2-1)} = 466.58$$

Nilai *Press's-Q* = 466.58 dibandingkan ($\chi^2_{0.05,1}$) = 3.841 yang berarti bahwa nilai *Press's-Q* > ($\chi^2_{0.05,1}$) sehingga kesimpulannya adalah ketepatan pengelompokan metode kernel tersebut memiliki kestabilan dalam melakukan pengklasifikasian.

4.3.3.3 Ketepatan Klasifikasi Asumsi Varians Tidak Sama

Untuk asumsi menggunakan nilai varians tidak sama, maka untuk data testing didapatkan bahwa untuk kelompok usia kerja yang tidak bekerja tepat diklasifikasikan sebanyak 331 orang dan yang salah diklasifikasikan sebanyak 48 orang. Untuk kelompok Usia kerja yang bekerja tepat diklasifikasikan sebanyak 379 orang dan yang salah diklasifikasikan sebanyak 50 orang, seperti pada tampilan tabel 4.20.

Tabel 4.20 Hasil Klasifikasi Metode Kernel Asumsi Varians
Tidak Sama Data Testing.

Observasi	Prediksi		Percentase benar
	0	1	
0	331	48	87.34 %
1	50	379	88.34 %
Ketepatan klasifikasi keseluruhan			87.87 %

Hasil klasifikasi tersebut memberikan prosentase kesalahan pengklasifikasian sebesar 12.13 % dan ketepatan pengklasifikasian sebesar 87.87 %.

4.3.3.4 Uji Keakuratan dan Kestabilan Asumsi Varians Tidak Sama

Seperti juga halnya analisa pada nilai varians yang sama, maka uji keakuratan untuk mengetahui apakah fungsi pada metode kernel memiliki tingkat akurasi tinggi dilakukan dengan membandingkan nilai dari *Hit Ratio* dengan nilai Cmax dan Cpro. Dimana perhitungannya didasarkan pada keterangan pada tabel 4.20 yang diformulasikan sebagai berikut:

$$\text{Kriteria bekerja} = p = 379/808 = 0.4690$$

$$\text{Kriteria tidak bekerja} = q = 331/808 = 0.4069$$

$$C_{\text{pro}} = p^2 + q^2 = (0.4690)^2 + (0.4069)^2 = 0.3878 = 38.78\%$$

$$C_{\text{max}} = (n_{\text{max}} / N) \times 100\% = 379/808 \times 100\% = 46.90 \%$$

Dihasilkan nilai Cpro sebesar 38.78% dan Cmax sebesar 46.90 %, oleh karena nilai *Hit Ratio* sebesar 87.87 % yang berarti lebih besar dibandingkan Cpro

maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode kernel adalah akurat.

Pengujian selanjutnya yaitu menguji tingkat kestabilan dari fungsi kernel yang dihasilkan, yaitu dengan menghitung tingkat kestabilan pengalokasian, hal ini bertujuan untuk mengkaji apakah ada kemungkinan pengalokasian dari tiap sampel dalam kelompok relatif stabil atau tidak sebagai akibat adanya perubahan perbedaan jumlah sampel yang diteliti.. Dengan menghitung nilai *Press's-Q*.

$$press's-Q = \frac{[N - (nxk)]^2}{N(k-1)} = \frac{[808 - (710x2)]^2}{808(2-1)}$$

$$= 463.54$$

Nilai *Press's-Q* = 463.54 dibandingkan ($\chi^2_{0.05,1}$) = 3.841 yang berarti bahwa nilai *Press's-Q* > ($\chi^2_{0.05,1}$) sehingga kesimpulannya adalah ketepatan pengelompokan metode kernel tersebut memiliki kestabilan dalam melakukan pengklasifikasian.

4.4 METODE ARTIFICIAL NEURAL NETWORK (ANN)

Pada tahap ini dilakukan pemodelan dengan metode ANN. Pemodelan metode ini juga bertujuan untuk mencari ketepatan klasifikasi seperti halnya pada dua metode sebelumnya.

Tidak ada prosedur umum yang dapat digunakan untuk menentukan jumlah *input*, jumlah *hidden layer* dan jumlah *node* pada masing-masing *hidden layer*. Dalam penelitian ini digunakan satu *hidden layer* agar jumlah *weight* yang ditaksir tidak terlalu banyak serta MSE yang dihasilkan sudah berpengaruh

sebagai kriteria pembanding. Sedangkan fungsi aktifasi yang digunakan adalah *Logistik Sigmoid* untuk *hidden layer* dan *output layer*.

4.4.1 Metode ANN Data Training

4.4.1.1 Ketepatan Klasifikasi

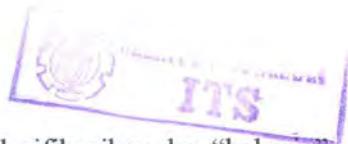
Pada analisa data training data yang digunakan berjumlah 1887 data dengan menggunakan 7 variabel terpilih. Penentuan arsitektur optimal dilakukan dengan *trial and error* sampai dengan jumlah hidden node 5. Berdasarkan prinsipnya maka akan terpilih jumlah *hidden node* dimana nilai BIC terkecil karena kriteria tersebut sangat peka terhadap besarnya observasi efektif.

Nilai MSE dianggap kurang mewaliki karena nilai MSE cenderung akan semakin kecil dengan semakin bertambahnya jumlah parameter. Karena nilai BIC terkecil terdapat pada hidden node 5 dengan menghasilkan nilai BIC minimum seperti yang terdapat pada tabel 4.21. Maka ketepatan klasifikasi menggunakan arsitektur 7-5-1.

Tabel 4.21 Hasil Metode ANN dengan Hidden Node 1-5

No	Jumlah Hidden Node	Jumlah Bobot	MSE	BIC
1	1	10	0.3596488	-1854.297
2	2	19	0.3748536	-1708.346
3	3	28	0.3139243	-1975.298
4	4	37	0.3178348	-1884.210
5	5	46	0.2820031	-2042.239

Dari jumlah *hidden node* yang dihasil *trial and error* maka didapat nilai pengklasifikasian seperti yang terlihat pada tabel 4.22. Terlihat bahwa terdapat



121 data klasifikasi “tidak bekerja” yang salah diklasifikasikan ke “bekerja” dengan jumlah ketepatan klasifikasi yang tidak bekerja sebanyak 764 data. Demikian pula untuk klasifikasi bekerja ketepatannya adalah sebanyak 910 data dengan misklasifikasi data sebesar 92 data dengan nilai MSE yang didapatkan sebesar -0.2820031.

Tabel 4.22 Hasil Klasifikasi Metode ANN Data Training.

Observasi	Prediksi		Percentase benar
	0	1	
0	764	121	86.33 %
1	92	910	90.82 %
Ketepatan klasifikasi keseluruhan			88.71 %

Dari tabel diatas diketahui bahwa ketepatan klasifikasi pengelompokan sebesar 88.77 % dan misklasifikasinya sebesar 11.23 %.

4.4.1.2 Uji Keakuratan

Seperti juga halnya analisa pada metode regresi logistik dan metode kernel, maka uji keakuratan untuk mengetahui apakah model memiliki tingkat akurasi tinggi dilakukan dengan membandingkan nilai dari *Hit Ratio* dengan nilai Cmax dan Cpro. Dimana perhitungannya didasarkan pada keterangan pada tabel 4.22 yang diformulasikan sebagai berikut:

$$\text{Kriteria bekerja} = p = 913/1887 = 0.4838$$

$$\text{Kriteria tidak bekerja} = q = 760/1887 = 0.4027$$

$$C_{pro} = p^2 + q^2 = (0.4838)^2 + (0.4027)^2 = 0.3963 = 39.63\%$$

$$C_{max} = (n_{max} / N) \times 100\% = 913/1887 \times 100\% = 48.38\%$$

Dihasilkan nilai Cpro sebesar 39.63% dan Cmax sebesar 48.38%, oleh karena nilai *Hit Ratio* sebesar 88.71 % yang berarti lebih besar dibandingkan Cpro maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan analisis metode ANN untuk data training adalah akurat.

4.4.2 Metode ANN Data Testing

4.4.2.1 Ketepatan Klasifikasi

Pada analisa data testing, data yang digunakan berjumlah 809 data dengan menggunakan 7 variabel yang dilakukan dengan jumlah *hidden node* 3 yang sesuai dengan kriteria BIC yang terkecil (tabel 4.21) dari hasil *trial and error* yang dicobakan dari 1 sampai 5 *hidden node* dengan nilai MSE sebesar -0.282003 dengan ketepatan klasifikasi seperti yang terdapat pada tabel 4.23.

Terlihat bahwa terdapat 50 data klasifikasi “tidak bekerja” yang salah diklasifikasikan ke “bekerja” dengan jumlah ketepatan klasifikasi yang tidak bekerja sebanyak 329 data. Demikian pula untuk klasifikasi bekerja ketepatannya adalah sebanyak 383 data dengan misklasifikasi data sebesar 47 data.

Tabel 4.23 Hasil Klasifikasi Metode ANN Data Testing.

Observasi	Prediksi		Percentase benar
	0	1	
0	329	50	86.81 %
1	47	383	89.07 %
Ketepatan klasifikasi keseluruhan			88.01 %

Dari tabel diatas diketahui bahwa ketepatan klasifikasi pengelompokan sebesar 88.01 % dan misklasifikasinya sebesar 11.99 %.

4.4.1.2 Uji Keakuratan dan Kestabilan

Seperti juga halnya analisa pada metode regresi logistik dan metode kernel, maka uji keakuratan untuk mengetahui apakah model memiliki tingkat akurasi tinggi dilakukan dengan membandingkan nilai dari *Hit Ratio* dengan nilai Cmax dan Cpro. Dimana perhitungannya didasarkan pada keterangan pada tabel 4.23 yang diformulasikan sebagai berikut:

$$\text{Kriteria tidak bekerja} = q = 329/809 = 0.4067$$

$$\text{Kriteria bekerja} = p = 383/809 = 0.4734$$

$$C_{\text{pro}} = p^2 + q^2 = (0.4734)^2 + (0.4067)^2 = 0.3895 = 38.95 \%$$

$$C_{\text{max}} = (n_{\text{max}} / N) \times 100\% = 383/809 \times 100\% = 47.34 \%$$

Dihasilkan nilai Cpro sebesar 40.67 % dan Cmax sebesar 47.34 %, oleh karena nilai *Hit Ratio* sebesar 88.01 % yang berarti lebih besar dibandingkan Cpro maupun Cmax. Dengan demikian dapat disimpulkan bahwa tingkat pengelompokan dari perhitungan metode ANN untuk data testing adalah akurat.

Pengujian selanjutnya yaitu menguji tingkat kestabilan dari fungsi ANN yang dihasilkan, yaitu dengan menghitung tingkat kestabilan pengalokasian, hal ini bertujuan untuk mengkaji apakah ada kemungkinan pengalokasian dari tiap sampel dalam kelompok relatif stabil atau tidak sebagai akibat adanya perubahan perbedaan jumlah sampel yang diteliti. Dengan menghitung nilai *Press's-Q*.

$$press's-Q = \frac{[N - (nxk)]^2}{N(k-1)} = \frac{[809 - (712 \times 2)]^2}{809(2-1)}$$

$$= 467.52$$

Nilai $Press's-Q = 467.52$ dibandingkan $(\chi^2_{0.05,1}) = 3.841$ yang berarti bahwa nilai $Press's-Q > (\chi^2_{0.05,1})$ sehingga kesimpulannya adalah ketepatan pengelompokan metode ANN tersebut memiliki kestabilan dalam melakukan pengklasifikasian.

4.5 PERBANDINGAN HASIL KLASIFIKASI METODE REGRESI LOGISTIK, KERNEL DAN ANN

4.5.1. Data Training

Dari hasil analisa yang dihasilkan, maka dapat diketahui bahwa metode regresi logistik, kernel, maupun ANN memiliki tingkat ketepatan pengklasifikasian yang berbeda-beda. Seperti yang tertera pada tabel 4.24 berikut ini.

Tabel 4.24 Perbandingan Ketepatan Klasifikasi Tiga Metode Data Training

Klasifikasi	Logistik	Kernel		ANN
		Varians sama	Varians tidak sama	
Data training	84.68 %	90.25 %	88.87 %	88.71%

Dari hasil tabel 4.24 dapat diketahui bahwa untuk data training studi kasus pengelompokan usia kerja dalam angkatan kerja untuk kabupaten Kediri yang berjumlah 1887 data ternyata metode yang memiliki tingkat ketepatan klasifikasi

yang paling baik berdasarkan kriteria keakuratan model dengan MCC dan PCC adalah metode Kernel asumsi varians sama dengan nilai ketepatan sebesar 90.25% jika dibandingkan dengan metode regresi logistik sebesar 84.68 % dan metode ANN (88.71 %) maupun metode kernel asumsi varians tidak sama (88.87 %).

4.5.2. Data Testing

Pada data testing yang berjumlah 809 data studi kasus pengelompokan usia kerja dalam angkatan kerja kabupaten Kediri, dapat diketahui bahwa ketepatan klasifikasi yang paling baik adalah pada metode ANN, dimana didapatkan nilai ketepatan klasifikasi sebesar 88.01 %. Hal ini juga didukung oleh nilai dari uji kestabilan model yang menghasilkan nilai *Press-Q* yang paling besar yaitu sebesar 467.52, seperti yang terlihat pada tabel 4.25.

Tabel 4.25 Perbandingan Ketepatan Klasifikasi Tiga Metode Data Testing

Klasifikasi	Logistik	Kernel		ANN
		Varians sama	Varians tidak sama	
Data testing	85.41 %	88.00%	87.87 %	88.01%
Press-Q	405.85	466.58	463.54	467.52

BAB V

KESIMPULAN DAN SARAN

BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Dari hasil analisa data Susenas 2000 untuk kabupaten Kediri dalam klasifikasi usia kerja dalam angkatan kerja, maka dapat disimpulkan sebagai berikut :

1. Hasil perhitungan dengan metode regresi logistik menyatakan bahwa faktor-faktor yang mempengaruhi seseorang dalam usia kerja itu masuk dalam angkatan kerja adalah variabel hubungan dengan kepala rumah tangga, jenis kelamin, umur, status perkawinan, pendidikan yang ditamatkan, banyaknya anggota rumah tangga dan variabel banyaknya anggota rumah tangga yang bekerja.
2. Jika data tidak membutuhkan asumsi kenormalan, maka metode regresi logistik lebih baik untuk menentukan variabel bebas yang berpengaruh terhadap variabel respon, karena pada regresi logistik terdapat seleksi model yang menghasilkan *Best Model* yang pada metode kernel dan ANN tidak ada.
3. Pada data training studi kasus klasifikasi usia kerja dalam angkatan kerja ternyata metode kernel asumsi varians sama lebih memberikan ketepatan klasifikasi yang lebih baik sebesar 90.25% dibandingkan dengan regresi logistik (84.68 %) ataupun ANN (88.71%), berdasarkan kriteria MCC dan PCC.

4. Untuk data testing dengan studi kasus yang sama pada data training, dapat diketahui bahwa metode ANN lebih menghasilkan ketepatan klasifikasi yang baik dibandingkan dengan metode regresi logistik dan kernel. Hal tersebut didukung pula oleh nilai uji $press-Q$ yang paling besar nilainya, yaitu untuk metode ANN sebesar 88.01 % dan nilai $press-Q = 467.52$, metode kernel asumsi varians sama sebesar 88.00% dan nilai $press-Q = 466.58$, sedangkan metode kernel asumsi varians tidak sama sebesar 87.87 % dan $press-Q = 463.54$. Untuk metode regresi logistik ketepatannya sebesar 85.41% dan nilai $press-Q= 405.85$.

5.2 SARAN

Sebagai tindak lanjut dari penelitian ini, maka penulis memberikan beberapa saran sebagai berikut :

1. Untuk kasus-kasus klasifikasi ini dapat dikembangkan dengan kasus-kasus lain dengan variabel bebas diskrit ataupun kontinu.
2. Untuk metode regresi logistik dapat dikembangkan dengan menggunakan variabel respon yang polikotomus.
3. Studi lanjut dapat dilakukan untuk meneliti efektifitas dan efisiensi ANN dengan struktur hidden layer lebih dari satu dengan hidden node yang lebih sedikit.

DAFTAR PUSTAKA

DAFTAR PUSTAKA

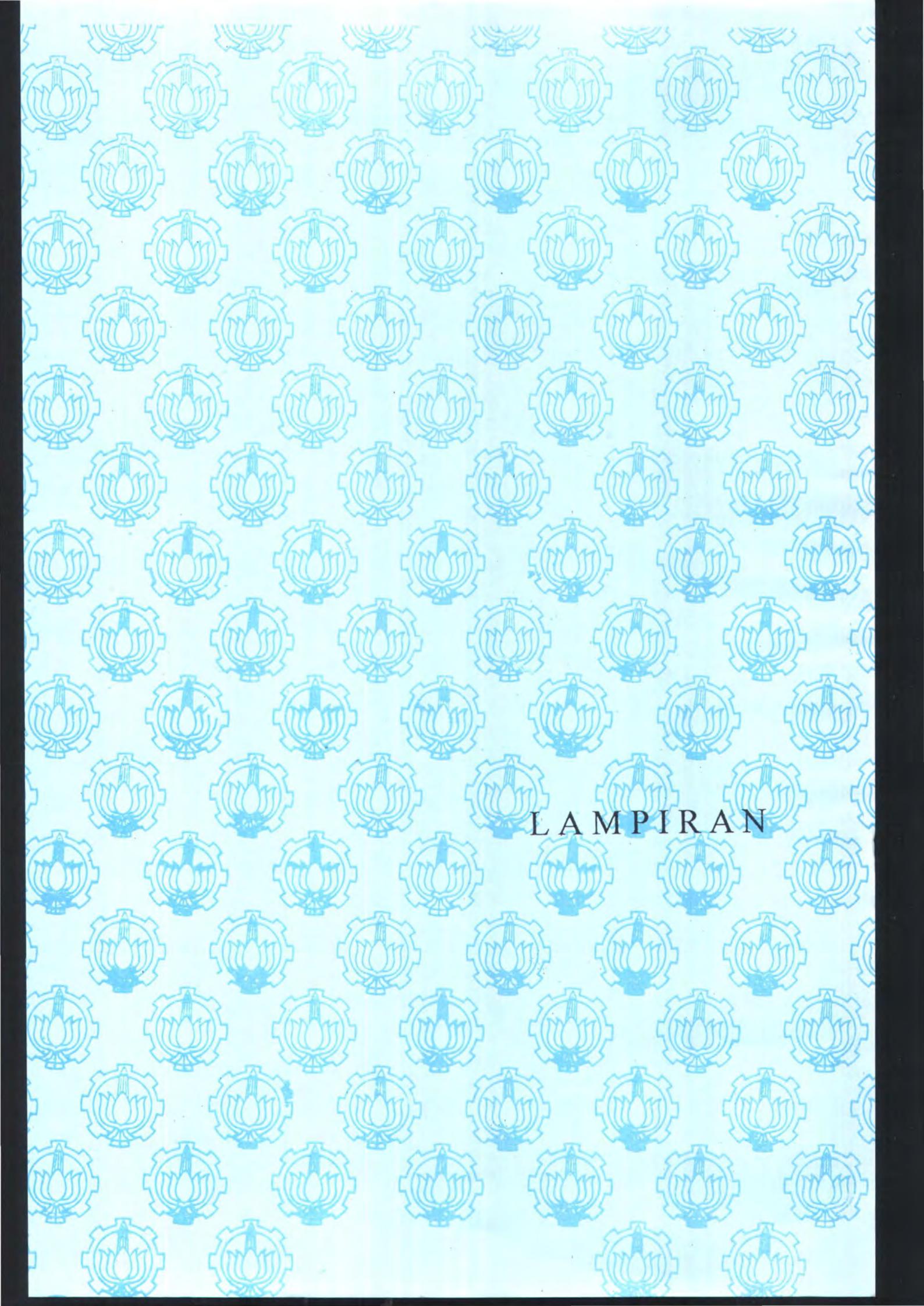
- Agresti, A., 1990, *Categorical Data Analysis*, John Wiley and Sons, New York.
- Anik Hidayat, 2002, *Formulasi Model Prediksi Kinerja Bisnis Perbankan di Indonesia Menggunakan Analisis Diskriminan dan Pengaruh Variabel Independen dalam Model terhadap Perubahan Relatif Laba Bank*, Tesis Magister Manajemen Universitas Airlangga, Surabaya.
- BPS, 1998, *Panduan Pelatihan Pemantauan Perkembangan Kesejahteraan Rakyat*, BPS, Jakarta.
- BPS, 2002, *Penjelasan Singkat Survei Angkatan Kerja Nasional*, BPS, Jakarta.
- BPS, 2001, *Laporan Eksekutif Susenas di Propinsi Jawa Timur*, BPS, Jakarta.
- Dillon, W.R. and M. Goldstein, 1978, On the Performance of Some Multinomial Classification Rules, *Journal of American Statistical Association*, 73, pp.305-313.
- Dwi Budi, *Modul Pelatihan Metode Kuantitatif untuk Ekonomi dan Bisnis*, 1998 Fakultas Ekonomi, Universitas Brawijaya, Malang.
- Fauset. L , 1994, *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice Hall International, Inc : New Jersey.
- Härdle, W, 1990, *Applied Nonparametrics Regression*, Cambridge University Press, Cambridge.
- Holmstrom, L. and Sain, S.R, 1997, *Multivariate Diskriminan Methods for Top Quark Analysis*, Technometrics, 39, 1, 91-99.
- Hosmer, D.W. and S. Lemeshow, 1989, *Applied Logistic Regression*, John Wiley & Sons, New York.
- Johnson, R.A. and D.W. Wichern, 1988, *Applied Multivariate Statistical Analysis*, Prentice Hall Inc, New Jersey.
- Nareshi K. Maholtra, 1993, *Marketing Research An Applied Orientation*, Prentice Hall Inc, New Jersey.
- Ripley, B.D, 1996, *Pattern Recognition and Neural Network*, Cambridge University Press, Cambridge.

Sakerti, 2001, *Aspek Kehidupan Rumah Tangga Indonesia*, Majalah Warta Demografi, Padjajaran, Edisi XII, hal 25-28.

Seber, G.A.F, 1984, *Multivariat Observations*, John Wiley and Son, New York.

Suwartapradja, Opan Suhendi, 2000, *Karakteristik Sosial Demografi Penduduk Lansia*, Jurnal kependudukan Padjajaran, Vol.2, No.1.

Yuli Arifiani, R, 2000, *Analisis Diskriminan Multi dengan Metode Linier dan Kernel “Studi Kasus Identifikasi Warna Objek Penggunaan Kamera”*, Tugas Akhir jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, ITS.



LAMPIRAN

LAMPIRAN 1

Deskriptif Statistik Usia Kerja Kabupaten Kediri

Hubungan dengan Kepala Rumah Tangga * Status Pekerjaan
Crosstabulation

	Kepala Rumah Tangga	Status Pekerjaan		Total
		Tidak Bekerja	Bekerja	
Hubungan dengan Kepala Rumah Tangga	Count % within Hubungan dengan Kepala Rumah Tangga	72 9.4%	892 90.6%	764 100.0%
	% within Status Pekerjaan % of Total	5.7% 2.7%	48.3% 25.7%	28.3% 28.3%
Istri/Suami	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	350 50.9%	337 49.1%	687 100.0%
		27.7% 13.0%	23.5% 12.5%	25.5% 25.5%
Anak	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	725 70.8%	299 29.2%	1024 100.0%
		57.4% 26.9%	20.9% 11.1%	38.0% 38.0%
Menantu	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	23 30.7%	52 69.3%	75 100.0%
		1.9% 1.9%	3.6% 1.9%	2.8% 2.8%
Cucu	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	51 73.9%	16 26.1%	69 100.0%
		4.0% 1.9%	1.3% .7%	2.6% 2.6%
Orang Tua/Mertua	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	17 68.0%	8 32.0%	25 100.0%
		1.3% 8%	.8% 3%	.9% 9%
Famili lain	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	23 56.1%	18 43.9%	41 100.0%
		1.8% 9%	1.3% .7%	1.5% 1.5%
Lainnya	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	3 27.3%	8 72.7%	11 100.0%
		.2% 1%	.6% 3%	.4% 4%
Total	Count % within Hubungan dengan Kepala Rumah Tangga % within Status Pekerjaan % of Total	1264 46.9%	1432 53.1%	2696 100.0%
		100.0% 46.9%	100.0% 53.1%	100.0% 100.0%

Jenis Kelamin * Status Pekerjaan Crosstabulation

		Status Pekerjaan		Total
	Laki-laki	Tidak Bekerja	Bekerja	
Jenis Kelamin	Laki-laki	Count	447	1371
		% within Jenis Kelamin	32.6%	67.4%
		% within Status Pekerjaan	35.4%	64.5%
	Perempuan	% of Total	16.6%	34.3%
		Count	817	1325
		% within Jenis Kelamin	61.7%	38.3%
	Total	% within Status Pekerjaan	64.6%	35.5%
		% of Total	30.3%	18.8%
		Count	1264	1432
		% within Jenis Kelamin	46.9%	53.1%
		% within Status Pekerjaan	100.0%	100.0%
		% of Total	46.9%	53.1%
				100.0%

Kelompok Umur * Status Pekerjaan Crosstabulation

		Status Pekerjaan		Total
	Kelompok Umur	Tidak Bekerja	Bekerja	
Kelompok Umur	10 - 24	Count	759	976
		% within Kelompok Umur	77.8%	22.2%
		% within Status Pekerjaan	60.0%	36.2%
	25 - 44	% of Total	28.2%	8.0%
		Count	323	1076
		% within Kelompok Umur	30.0%	70.0%
	45 - 64	% within Status Pekerjaan	25.6%	39.9%
		% of Total	12.0%	27.9%
		Count	182	644
		% within Kelompok Umur	28.3%	71.7%
		% within Status Pekerjaan	14.4%	32.3%
		% of Total	6.8%	23.9%
	Total	Count	1264	1432
		% within Kelompok Umur	46.9%	53.1%
		% within Status Pekerjaan	100.0%	100.0%
		% of Total	46.9%	53.1%
				100.0%

Status Kesehatan * Status Pekerjaan Crosstabulation

			Status Pekerjaan		Total
Status Kesehatan	Sehat	Count % within Status Kesehatan % within Status Pekerjaan % of Total	Tidak Bekerja	Bekerja	
Sehat	Sehat	982 47.8% 77.7% 36.4%	1071 52.2% 74.8% 39.7%	2053 100.0% 76.1% 76.1%	
Sakit	Sakit	282 43.9% 22.3% 10.5%	361 56.1% 25.2% 13.4%	643 100.0% 23.9% 23.9%	
Total		1264 46.9% 100.0% 46.9%	1432 53.1% 100.0% 53.1%	2696 100.0% 100.0% 100.0%	

Status Kota/Desa * Status Pekerjaan Crosstabulation

			Status Pekerjaan		Total
Status Kota/Desa	Perkotaan	Count % within Status Kota/Desa % within Status Pekerjaan % of Total	Tidak Bekerja	Bekerja	
Perkotaan	Perkotaan	438 50.7% 34.7% 16.2%	426 49.3% 29.7% 15.8%	864 100.0% 32.0% 32.0%	
Pedesaan	Pedesaan	826 45.1% 65.3% 30.6%	1006 54.9% 70.3% 37.3%	1832 100.0% 68.0% 68.0%	
Total		1264 46.9% 100.0% 46.9%	1432 53.1% 100.0% 53.1%	2696 100.0% 100.0% 100.0%	

Pendidikan yang Ditamatkan * Status Pekerjaan Crosstabulation

		Status Pekerjaan		Total
		Tidak Bekerja	Bekerja	
Pendidikan yang Ditamatkan	Tidak Sekolah/Tidak Tamat SD	Count % within Pendidikan yang Ditamatkan % within Status Pekerjaan % of Total	386 48.5% 30.5% 14.3%	410 51.5% 28.6% 15.2%
	Tamat SD/Sederajat	Count % within Pendidikan yang Ditamatkan % within Status Pekerjaan % of Total	472 45.7% 37.3% 17.5%	561 54.3% 39.2% 20.8%
	Tamat SLTP/Sederajat	Count % within Pendidikan yang Ditamatkan % within Status Pekerjaan % of Total	237 56.2% 18.8% 8.8%	185 43.8% 12.9% 6.9%
	Tamat SLTA/Sederajat ke Atas	Count % within Pendidikan yang Ditamatkan % within Status Pekerjaan % of Total	169 38.0% 13.4% 6.3%	276 62.0% 19.3% 10.2%
Total		Count % within Pendidikan yang Ditamatkan % within Status Pekerjaan % of Total	1264 46.9% 100.0% 46.9%	1432 53.1% 100.0% 53.1%
				2696 100.0% 100.0% 100.0%



Banyaknya Anggota Rumah Tangga * Status Pekerjaan Crosstabulation

		Status Pekerjaan		Total
		Tidak Bekerja	Bekerja	
1	Count	4	17	21
	% within Banyaknya Anggota Rumah Tangga	19.0%	81.0%	100.0%
	% within Status Pekerjaan	3%	1.2%	8%
	% of Total	1%	8%	8%
2	Count	55	146	201
	% within Banyaknya Anggota Rumah Tangga	27.4%	72.6%	100.0%
	% within Status Pekerjaan	4.4%	10.2%	7.5%
	% of Total	2.0%	5.4%	7.5%
3	Count	226	309	535
	% within Banyaknya Anggota Rumah Tangga	42.2%	57.8%	100.0%
	% within Status Pekerjaan	17.9%	21.6%	19.8%
	% of Total	8.4%	11.5%	19.8%
4	Count	330	397	727
	% within Banyaknya Anggota Rumah Tangga	45.4%	54.6%	100.0%
	% within Status Pekerjaan	26.1%	27.7%	27.0%
	% of Total	12.2%	14.7%	27.0%
5	Count	285	280	565
	% within Banyaknya Anggota Rumah Tangga	50.4%	49.6%	100.0%
	% within Status Pekerjaan	22.5%	19.6%	21.0%
	% of Total	10.6%	10.4%	21.0%
6	Count	198	162	360
	% within Banyaknya Anggota Rumah Tangga	55.0%	45.0%	100.0%
	% within Status Pekerjaan	15.7%	11.3%	13.4%
	% of Total	7.3%	6.0%	13.4%
7	Count	88	64	152
	% within Banyaknya Anggota Rumah Tangga	57.9%	42.1%	100.0%
	% within Status Pekerjaan	7.0%	4.5%	5.6%
	% of Total	3.3%	2.4%	5.6%
8	Count	43	42	85
	% within Banyaknya Anggota Rumah Tangga	50.6%	49.4%	100.0%
	% within Status Pekerjaan	3.4%	2.9%	3.2%
	% of Total	1.6%	1.6%	3.2%
9	Count	20	9	29
	% within Banyaknya Anggota Rumah Tangga	69.0%	31.0%	100.0%
	% within Status Pekerjaan	1.6%	.8%	1.1%
	% of Total	7%	3%	1.1%
10	Count	15	8	23
	% within Banyaknya Anggota Rumah Tangga	71.4%	28.6%	100.0%
	% within Status Pekerjaan	1.2%	.4%	.8%
	% of Total	8%	2%	8%
Total	Count	1264	1432	2696
	% within Banyaknya Anggota Rumah Tangga	46.9%	53.1%	100.0%
	% within Status Pekerjaan	100.0%	100.0%	100.0%
	% of Total	46.9%	53.1%	100.0%

Banyaknya ART yang Bekerja * Status Pekerjaan Crosstabulation

		Status Pekerjaan		Total
		Tidak Bekerja	Bekerja	
Banyaknya ART yang Bekerja	0	Count	104	104
		% within Banyaknya ART yang Bekerja	100.0%	100.0%
		% within Status Pekerjaan	8.2%	3.9%
		% of Total	3.9%	3.9%
	1	Count	608	942
		% within Banyaknya ART yang Bekerja	64.5%	35.5%
		% within Status Pekerjaan	48.1%	23.3%
		% of Total	22.6%	34.9%
-	2	Count	392	1025
		% within Banyaknya ART yang Bekerja	38.2%	61.8%
		% within Status Pekerjaan	31.0%	44.2%
		% of Total	14.5%	38.0%
	3	Count	111	398
		% within Banyaknya ART yang Bekerja	27.9%	72.1%
		% within Status Pekerjaan	8.8%	20.0%
		% of Total	4.1%	14.8%
-	4	Count	45	192
		% within Banyaknya ART yang Bekerja	23.4%	76.6%
		% within Status Pekerjaan	3.6%	10.3%
		% of Total	1.7%	7.1%
	5	Count	4	35
		% within Banyaknya ART yang Bekerja	11.4%	88.6%
		% within Status Pekerjaan	.3%	2.2%
		% of Total	.1%	1.3%
Total		Count	1264	2696
		% within Banyaknya ART yang Bekerja	46.9%	53.1%
		% within Status Pekerjaan	100.0%	100.0%
		% of Total	46.9%	53.1%

Pengeluaran Perkapita * Status Pekerjaan Crosstabulation

			Status Pekerjaan		Total
			Tidak Bekerja	Bekerja	
Pengeluaran Perkapita	0 - 99.999	Count	873	888	1761
		% within Pengeluaran Perkapita	49.6%	50.4%	100.0%
		% within Status Pekerjaan	69.1%	62.0%	65.3%
		% of Total	32.4%	32.9%	65.3%
		Count	266	394	660
		% within Pengeluaran Perkapita	40.3%	59.7%	100.0%
	100.000 - 149.999	% within Status Pekerjaan	21.0%	27.5%	24.5%
		% of Total	9.9%	14.6%	24.5%
		Count	89	97	186
		% within Pengeluaran Perkapita	47.8%	52.2%	100.0%
		% within Status Pekerjaan	7.0%	6.8%	6.9%
		% of Total	3.3%	3.6%	6.9%
	150.000 - 199.999	Count	19	30	49
		% within Pengeluaran Perkapita	38.8%	61.2%	100.0%
		% within Status Pekerjaan	1.5%	2.1%	1.8%
		% of Total	.7%	1.1%	1.8%
		Count	17	23	40
		% within Pengeluaran Perkapita	42.5%	57.5%	100.0%
	300.000 +	% within Status Pekerjaan	1.3%	1.6%	1.5%
		% of Total	.6%	.9%	1.5%
		Count	1264	1432	2696
		% within Pengeluaran Perkapita	46.9%	53.1%	100.0%
		% within Status Pekerjaan	100.0%	100.0%	100.0%
		% of Total	46.9%	53.1%	100.0%
Total					

LAMPIRAN 2

Metode Regresi Logistik Univariat

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
1.. X1

Estimation terminated at iteration number 4 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2028.815
Goodness of Fit 1886.994
Cox & Snell - R² .265
Nagelkerke - R² .265

	Chi-Square	df	Significance
Model	579.863	7	.0000
Block	579.863	7	.0000
Step	579.863	7	.0000

Classification Table for Y

The Cut Value is .50

Observed	Predicted			Percent Correct
	0		1	
	0	I	1	
0	I	562	I	63.50%
1	I	213	I	78.74%
		Overall		71.60%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X1			390.7825	7	.0000	.3800	
X1(1)	1.1433	1.1641	.9647	1	.3260	.0000	3.1372
X1(2)	-1.0747	1.1581	.8611	1	.3534	.0000	.3414
X1(3)	-2.0336	1.1577	3.0857	1	.0790	-.0204	.1309
X1(4)	.5108	1.2267	.1734	1	.6771	.0000	1.6667
X1(5)	-2.4849	1.2134	4.1942	1	.0406	-.0290	.0833
X1(6)	-2.6391	1.3184	4.0070	1	.0453	-.0277	.0714
X1(7)	-1.0380	1.2061	.7407	1	.3894	.0000	.3542
Constant	1.0986	1.1547	.9052	1	.3414		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
1.. X2

Estimation terminated at iteration number 2 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2480.079
Goodness of Fit 1886.971
Cox & Snell - R² .066
Nagelkerke - R² .066

	Chi-Square	df	Significance
Model	128.599	1	.0000
Block	128.599	1	.0000
Step	128.599	1	.0000

Classification Table for Y
The Cut Value is .50

Observed	Predicted			Percent Correct
	0	I	1	
0 0	I 571	I 314	I	64.52%
1 1	I 386	I 616	I	61.48%
		Overall		62.90%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X2(1)	1.0653	.0957	124.0320	1	.0000	.2163	2.9017
Constant	-.3915	.0659	35.3090	1	.0000		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
1.. X3

Estimation terminated at iteration number 3 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood	2132.096
Goodness of Fit	1886.991
Cox & Snell - R ²	.223
Nagelkerke - R ²	.223

	Chi-Square	df	Significance
Model	476.583	2	.0000
Block	476.583	2	.0000
Step	476.583	2	.0000

Classification Table for Y
The Cut Value is .50

Observed	Predicted			Percent Correct
	0	I	1	
0 0	I 539	I 346	I	60.90%
1 1	I 138	I 864	I	86.23%
		Overall		74.35%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X3			395.0506	2	.0000	.3872	
X3(1)	-2.3548	.1405	281.0976	1	.0000	-.3271	.0949
X3(2)	-.1262	.1310	.9272	1	.3356	.0000	.8815

Constant .9924 .1031 92.6862 1 .0000

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
1.. X4

Estimation terminated at iteration number 3 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2245.112
Goodness of Fit 1886.999
Cox & Snell - R² .175
Nagelkerke - R² .175

	Chi-Square	df	Significance
Model	363.566	2	.0000
Block	363.566	2	.0000
Step	363.566	2	.0000

Classification Table for Y

The Cut Value is .50

		Predicted			Percent Correct	
		0	1		0	1
Observed	0	I 518	I 367	I	58.53%	
	1	I 171	I 831	I	82.93%	
				Overall	71.49%	

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
X4			319.1660	2	.0000	.3476	
X4(1)	-1.6109	.2013	64.0261	1	.0000	-.1542	.1997
X4(2)	.3554	.1930	3.3924	1	.0655	.0231	1.4267
Constant	.5026	.1810	7.7132	1	.0055		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
1.. X5

Estimation terminated at iteration number 2 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2604.368
Goodness of Fit 1887.000
Cox & Snell - R² .002
Nagelkerke - R² .002

	Chi-Square	df	Significance
Model	4.310	1	.0379
Block	4.310	1	.0379
Step	4.310	1	.0379

Classification Table for Y

The Cut Value is .50

Observed	Predicted		Percent Correct	
	0	1	0	1
0 0	I	0 I	885 I	.00%
1 1	I	0 I	1002 I	100.00%
			Overall	53.10%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X5(1)	-.2280	.1101	4.2864	1	.0384	-.0296	.7962
Constant	.2997	.0967	9.5961	1	.0019		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number

1.. X6

Estimation terminated at iteration number 2 because
parameter estimates changed by less than .001

-2 Log Likelihood	2603.656
Goodness of Fit	1887.000
Cox & Snell - R^2	.003
Nagelkerke - R^2	.003

	Chi-Square	df	Significance
Model	5.023	1	.0250
Block	5.023	1	.0250
Step	5.023	1	.0250

Classification Table for Y

The Cut Value is .50

Observed	Predicted		Percent Correct	
	0	1	0	1
0 0	I	316 I	569 I	35.71%
1 1	I	309 I	693 I	69.16%
		Overall		53.47%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X6(1)	-.2196	.0980	5.0204	1	.0251	-.0340	.8029
Constant	.1971	.0566	12.1444	1	.0005		

No more variables can be deleted or added.

3	292	.000	.000	1.000
4	317	.000	.000	.000

Dependent Variable.. Y
 Beginning Block Number 0. Initial Log Likelihood Function
 -2 Log Likelihood 2608.6784
 * Constant is included in the model.

Estimation terminated at iteration number 2 because parameter estimates changed by less than .001

Classification Table for Y

The Cut Value is .50

Observed	Predicted			Percent Correct	
	0	I	1	0	I
0 0	I	0 I	885 I	.00%	
1 1	I	0 I	1002 I	100.00%	
				Overall	53.10%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
Constant	.1242	.0461	7.2451	1	.0071		

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
 1.. X7

Classification Table for Y

The Cut Value is .50

Observed	Predicted			Percent Correct	
	0	I	1	0	I
0 0	I	179 I	706 I	20.23%	
1 1	I	113 I	889 I	88.72%	
				Overall	56.60%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X7			32.1991	3	.0000	.1002	
X7(1)	-.2871	.1420	4.0879	1	.0432	-.0283	.7504
X7(2)	-.1555	.1374	1.2801	1	.2579	.0000	.8560
X7(3)	-.8628	.1661	26.9996	1	.0000	-.0979	.4220
Constant	.4028	.1146	12.3524	1	.0004		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
 1.. X8

Estimation terminated at iteration number 2 because

Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2576.728
Goodness of Fit 1887.001
Cox & Snell - R² .017
Nagelkerke - R² .017

	Chi-Square	df	Significance
Model	31.951	1	.0000
Block	31.951	1	.0000
Step	31.951	1	.0000

Classification Table for Y

The Cut Value is .50

Observed	Predicted			Percent Correct	
	0	1	0		
0	I 247	I 638	I 27.91%		
1	I 196	I 806	I 80.44%		
			Overall	55.80%	

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X8	-.1619	.0290	31.1180	1	.0000	-.1057	.8505
Constant	.8462	.1376	37.8342	1	.0000		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
1.. X9

Estimation terminated at iteration number 3 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2377.591
Goodness of Fit 1915.195
Cox & Snell - R² .115
Nagelkerke - R² .115

	Chi-Square	df	Significance
Model	231.088	1	.0000
Block	231.088	1	.0000
Step	231.088	1	.0000

Classification Table for Y

The Cut Value is .50

Observed	Predicted			Percent Correct	
	0	1	0		
0	I 489	I 396	I 55.25%		
1	I 238	I 764	I 76.25%		
			Overall	66.40%	

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X9	.7856	.0571	189.5228	1	.0000	.2681	2.1937
Constant	-1.3273	.1133	137.1204	1	.0000		

No more variables can be deleted or added.

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number

1.. X10

Estimation terminated at iteration number 2 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 2584.609
Goodness of Fit 1886.992
Cox & Snell - R^2 .013
Nagelkerke - R^2 .013

Chi-Square df Significance

Model	24.069	4	.0001
Block	24.069	4	.0001
Step	24.069	4	.0001

Classification Table for Y

The Cut Value is .50

		Predicted		Percent Correct	
		0	1		
		0	1		
Observed	0	+	+	77.18%	
	0	I	683	I	202 I
	1	+	+	32.63%	
	1	I	675	I	327 I
		+	+	+	+
		Overall	53.52%		



----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X10			23.4293	4	.0001	.0769	
X10(1)	-.5942	.3985	2.2237	1	.1359	-.0093	.5520
X10(2)	-.1408	.4057	.1204	1	.7286	.0000	.8687
X10(3)	-.6533	.4340	2.2660	1	.1322	-.0101	.5203
X10(4)	.2720	.5338	.2596	1	.6104	.0000	1.3126
Constant	.5877	.3944	2.2208	1	.1362		

No more variables can be deleted or added.

LAMPIRAN 3

Metode Regresi Logistik Multivariat

Beginning Block Number 1. Method: Backward Stepwise (WALD)

Variable(s) Entered on Step Number
 1.. X1
 X10
 X2
 X3
 X4
 X5
 X6
 X7
 X8
 X9

Estimation terminated at iteration number 5 because
 Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 1282.895
 Goodness of Fit 2357.057
 Cox & Snell - R² .505
 Nagelkerke - R² .505

	Chi-Square	df	Significance
Model	1325.783	23	.0000
Block	1325.783	23	.0000
Step	1325.783	23	.0000

Classification Table for Y
 The Cut Value is .50

Observed	Predicted			Percent Correct
	0		1	
	0	I	1	
0 0	I 760	I 125	I	85.88%
1 1	I 115	I 887	I	88.52%
	Overall			87.28%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X1			72.5862	7	.0000	.1499	
X1(1)	2.5896	1.7780	2.1213	1	.1453	.0068	13.3243
X1(2)	.4491	1.7687	.0645	1	.7996	.0000	1.5669
X1(3)	.4707	1.7411	.0731	1	.7869	.0000	1.6011
X1(4)	2.1447	1.8426	1.3548	1	.2444	.0000	8.5397
X1(5)	.0926	1.8151	.0026	1	.9593	.0000	1.0971
X1(6)	-.4365	1.9660	.0493	1	.8243	.0000	.6463
X1(7)	.2646	1.8037	.0215	1	.8834	.0000	1.3029
X10			2.6948	4	.6101	.0000	
X10(1)	.5521	.5485	1.0131	1	.3142	.0000	1.7369
X10(2)	.4946	.5555	.7927	1	.3733	.0000	1.6398
X10(3)	.8481	.5990	2.0051	1	.1568	.0014	2.3353
X10(4)	.7947	.7272	1.1941	1	.2745	.0000	2.2137

X2(1)	1.5228	.2145	50.4051	1	.0000	.1362	4.5853		
X3	□		84.2648	2	.0000	.1754			
X3(1)	-1.7193	.2971	33.4880	1	.0000	-.1099	.1792		
X3(2)	.5227	.2006	6.7887	1	.0092	.0428	1.6866		
X4			19.3715	2	.0001	.0768			
X4(1)	-.5372	.4322	1.5450	1	.2139	.0000	.5844		
X4(2)	.7711	.3438	5.0305	1	.0249	.0341	2.1621		
X5(1)	.1591	.1743	.8334	1	.3613	.0000	1.1724		
X6(1)	.1126	.1563	.5193	1	.4711	.0000	1.1192		
X7			7.4293	3	.0594	.0234			
X7(1)	-.6099	.2352	6.7251	1	.0095	-.0426	.5434		
X7(2)	-.2713	.2153	1.5870	1	.2078	.0000	.7624		
X7(3)	-.4223	.2522	2.8029	1	.0941	-.0175	.6555		
X8			-.5398	.0565	91.3214	1	.0000	-.1850	.5829
X9			1.9265	.1120	295.8986	1	.0000	.3357	6.8655
Constant	-2.9325	1.9055	2.3685	1	.1238				

Variable(s) Removed on Step Number
2.. X10

Estimation terminated at iteration number 5 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood	1285.582
Goodness of Fit	2361.202
Cox & Snell - R^2	.504
Nagelkerke - R^2	.504

Chi-Square df Significance

Model	1323.097	19	.0000
Block	1323.097	19	.0000
Step	-2.686	4	.6116

Note: A negative Chi-Square value indicates that the Chi-Square value has decreased from the previous step.

Classification Table for Y

The Cut Value is .50

Predicted			Percent Correct		
0 I 1					
Observed	0	I	1	I	
0	0	I	760	I	125
1	1	I	119	I	883
			Overall 87.07%		

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X1			72.6018	7	.0000	.1499	
X1(1)	2.5641	1.7651	2.1102	1	.1463	.0065	12.9883
X1(2)	.4385	1.7553	.0624	1	.8027	.0000	1.5504
X1(3)	.4493	1.7287	.0676	1	.7949	.0000	1.5673
X1(4)	2.1286	1.8299	1.3531	1	.2447	.0000	8.4033
X1(5)	.0707	1.8024	.0015	1	.9687	.0000	1.0733
X1(6)	-.5036	1.9534	.0665	1	.7965	.0000	.6043
X1(7)	.2303	1.7924	.0165	1	.8978	.0000	1.2589
X2(1)	1.5222	.2135	50.8186	1	.0000	.1368	4.5822
X3			83.3824	2	.0000	.1744	
X3(1)	-1.7124	.2967	33.3202	1	.0000	-.1096	.1804
X3(2)	.4992	.1997	6.2497	1	.0124	.0404	1.6473
X4			19.0400	2	.0001	.0759	

X4(1)	.5674	.4321	1.7247	1	.1891	.0000	.5670
X4(2)	.7319	.3423	4.5730	1	.0325	.0314	2.0790
X5(1)	.1559	.1738	.8047	1	.3697	.0000	1.1687
X6(1)	.1047	.1547	.4574	1	.4988	.0000	1.1103
X7			8.1845	3	.0423	.0289	
X7(1)	-.6277	.2268	7.6632	1	.0056	-.0466	.5338
X7(2)	-.2975	.2096	2.0141	1	.1558	-.0023	.7427
X7(3)	-.4377	.2490	3.0886	1	.0788	-.0204	.6455
X8	-.5341	.0544	96.3254	1	.0000	-.1902	.5862
X9	1.9114	.1105	299.4642	1	.0000	.3377	6.7625
Constant	-2.2926	1.8136	1.5979	1	.2062		

Variable(s) Removed on Step Number
3.. X6

Estimation terminated at iteration number 5 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood	1286.040
Goodness of Fit	2357.764
Cox & Snell - R^2	.504
Nagelkerke - R^2	.504

	Chi-Square	df	Significance
Model	1322.639	18	.0000
Block	1322.639	18	.0000
Step	-.458	1	.4986

Note: A negative Chi-Square value indicates that the Chi-Square value has decreased from the previous step.

Classification Table for Y
The Cut Value is .50

Observed	Predicted			Percent Correct
	0		1	
	0	I	1	
0	I	761	I	85.99%
1	I	118	I	88.22%
			Overall	87.18%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X1			72.5795	7	.0000	.1499	
X1(1)	2.6038	1.7604	2.1879	1	.1391	.0085	13.5155
X1(2)	.4763	1.7506	.0740	1	.7856	.0000	1.6101
X1(3)	.4841	1.7241	.0788	1	.7789	.0000	1.6227
X1(4)	2.1454	1.8261	1.3803	1	.2401	.0000	8.5455
X1(5)	.1030	1.7982	.0033	1	.9543	.0000	1.1085
X1(6)	-.4830	1.9501	.0613	1	.8044	.0000	.6169
X1(7)	.2827	1.7865	.0250	1	.8743	.0000	1.3267
X2(1)	1.5226	.2135	50.8471	1	.0000	.1368	4.5842
X3		"	83.4223	2	.0000	.1745	
X3(1)	-1.7144	.2963	33.4695	1	.0000	-.1098	.1801
X3(2)	.4951	.1996	6.1538	1	.0131	.0399	1.6407
X4			18.9382	2	.0001	.0757	
X4(1)	-.5617	.4322	1.6892	1	.1937	.0000	.5702
X4(2)	.7326	.3421	4.5862	1	.0322	.0315	2.0804
X5(1)	.1652	.1732	.9101	1	.3401	.0000	1.1797
X7			8.7993	3	.0321	.0328	

X7(1)	-.6503	.2243	8.4042	1	.0037	-.0495	.5219
X7(2)	-.3228	.2063	2.4496	1	.1176	-.0131	.7241
X7(3)	-.4531	.2479	3.3409	1	.0676	-.0227	.6356
X8	-.5283	.0537	96.9452	1	.0000	-.1908	.5896
X9	1.9053	.1100	300.2373	1	.0000	.3381	6.7213
Constant	-2.2965	1.8098	1.6101	1	.2045		

Variable(s) Removed on Step Number
4.. X5

Estimation terminated at iteration number 5 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 1286.951
Goodness of Fit 2357.375
Cox & Snell - R^2 .504
Nagelkerke - R^2 .504

	Chi-Square	df	Significance
Model	1321.728	17	.0000
Block	1321.728	17	.0000
Step	-.911	1	.3398

Note: A negative Chi-Square value indicates that the Chi-Square
value has decreased from the previous step.

Classification Table for Y

The Cut Value is .50

Observed	Predicted			Percent Correct
	0		1	
	0	I	1	
0	I	760	I	85.88%
1	I	115	I	88.52%
			Overall	87.28%

Variables in the Equation							
Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
X1			72.0051	7	.0000	.1491	
X1(1)	2.5363	1.7571	2.0835	1	.1489	.0057	12.6329
X1(2)	.4183	1.7478	.0573	1	.8109	.0000	1.5193
X1(3)	.4583	1.7222	.0708	1	.7902	.0000	1.5813
X1(4)	2.1266	1.8244	1.3587	1	.2438	.0000	8.3865
X1(5)	.1038	1.7965	.0033	1	.9539	.0000	1.1093
X1(6)	-.5351	1.9531	.0751	1	.7841	.0000	.5856
X1(7)	.2374	1.7843	.0177	1	.8942	.0000	1.2679
X2(1)	1.5159	.2131	50.6135	1	.0000	.1365	4.5534
X3			83.3427	2	.0000	.1744	
X3(1)	-1.7154	.2960	33.5968	1	.0000	-.1101	.1799
X3(2)	.4900	.1993	6.0455	1	.0139	.0394	1.6324
X4			19.4389	2	.0001	.0769	
X4(1)	-.5728	.4313	1.7637	1	.1842	.0000	.5640
X4(2)	.7381	.3413	4.6768	1	.0306	.0320	2.0919
X7			9.3621	3	.0248	.0359	
X7(1)	-.6690	.2235	8.9609	1	.0028	-.0517	.5122
X7(2)	-.3320	.2061	2.5953	1	.1072	-.0151	.7175
X7(3)	-.4625	.2478	3.4834	1	.0620	-.0238	.6297
X8	-.5285	.0536	97.2195	1	.0000	-.1911	.5895
X9	1.9059	.1098	301.1927	1	.0000	.3387	6.7252
Constant	-2.1068	1.7968	1.3748	1	.2410		

----- Variables not in the Equation -----
Residual Chi Square 4.074 with 6 df Sig = .6667

Variable	Score	df	Sig	R
X10	2.6165	4	.6239	.0000
X10(1)	.0015	1	.9692	.0000
X10(2)	.1820	1	.6696	.0000
X10(3)	1.1616	1	.2811	.0000
X10(4)	.2147	1	.6431	.0000
X5(1)	.9109	1	.3399	.0000
X6(1)	.5630	1	.4531	.0000

No more variables can be deleted or added.

LAMPIRAN 4

Program ANN dan Regresi Logistik dalam Syntax S-Plus

```
sumfit.nnts <- function(net)
{
  if(length(net) == 10)
    net$softmax <- F
  cat("a ", net$n[1], "-", net$n[2], "-", net$n[3], " network", sep
= "")
  cat(" with", length(net$wts), "weights\n")
  tconn <- diff(net$nconn)
  if(tconn[length(tconn)] > net$n[2] + 1)
    cat(" skip-layer connections ")
  if(net$decay > 0)
    cat(" decay=", net$decay, sep = "")
  cat("\n")
  cat("Unit 0 is constant one input\n")
  cat("Input units: ", paste("Lag",
",net$lags, =", 1:net$n[1], ", ", sep=""), "\n")
  cat("Hidden units are ", (net$n[1]+1):(net$n[1]+net$n[2]), "\n")
  cat("Output unit is", sum(net$n), "\n\n")
  wts <- format(round(net$wts, 4))
  names(wts) <- apply(cbind(net$conn, rep(1:net$nunits - 1,
tconn)), 1,
  function(x)
  paste(x, collapse = "->"))
  print(wts, quote = F)
  cat("Sum of squares is ", format(net$val), "\n")
  n <- length(net$fit)
  p <- length(net$wts)
  ft <- n*log(net$val/n)
  cat("AIC :", format(ft+2*p),
  ", SBC :", format(ft+p*log(n)),
  ", BIC :", format(ft+p+p*log(n)),
  ", MSE :", format(net$val/(n-p)),
  ", residual se :", format(sqrt(net$val/(n-p))), "\n")
  invisible()
}

y1<-ANN[,8,]
x1<-ANN[,1:7,]
data<-data.frame(y1,x1)
datta<-rbind(data)
samp <- c(sample(1:1264,885), sample(1265:2696,1002))
nn1 <- nnet(y1~hub+kel+umur+status+pen+art+artk, size=5,
  data=data,
  subset=samp, rang=0.1, skip=F, trace=T, entropy=T,
  decay=5e-4, maxit=100)

sumfit.nnts(nn1)

# Hasil pada data training dg ANN
train.pred <- format(round(predict(nn1, data[samp,])))
train.real <- y1[samp]
```

```
table(train.pred, train.real)

# Hasil pada data testing dg ANN
test.pred <- format(round(predict(nn1, data[-samp,])))
test.real <- y1[-samp]
table(test.pred, test.real)
# Hasil Logistics regression
log1 <- glm(y1~hub+kel+umur+status+pen+art+artk, family=binomial,
             data=data,
             subset=samp, trace=T)
summary(log1)$coefficients

# Hasil pada data training dg Logistics regression
train.plog <- format(round(predict.glm(log1, type="response")))
table(train.plog, train.real)

# Hasil pada data testing dg Logistics regression
test.plog <- format(round(predict.glm(log1, newdata=data[-samp,],
                                         type="response")))
table(test.plog, test.real)
```

LAMPIRAN 5

Ouput Program ANN dan Regresi Logistik

```
> sumfit.nnts <- function(net)
{
  if(length(net) == 10)
    net$softmax <- F
  cat("a ", net$n[1], "-", net$n[2],
      "-", net$n[3], " network",
      sep = "")
  cat(" with", length(net$wts),
      "weights\n")
  tconn <- diff(net$nconn)
  if(tconn[length(tconn)] > net$n[2] +
      1)
    cat(" skip-layer connections "
        )
  if(net$decay > 0)
    cat(" decay=", net$decay,
        sep = "")
  cat("\n")
  cat("Unit 0 is constant one input\n")
  cat("Input units: ", paste("Lag ",
      net$lags, "=",
      1:net$n[1],
      ",",
      sep = ""),
      "\n")
  cat("Hidden units are ", (net$n[1] +
      1):(net$n[1] + net$n[2]),
      "\n")
  cat("Output unit is", sum(net$n),
      "\n\n")
  wts <- format(round(net$wts, 4))
  names(wts) <- apply(cbind(net$conn,
      rep(1:net$numunits - 1, tconn
          )),
      1, function(x)
  paste(x, collapse = "->"))
  print(wts, quote = F)
  cat("Sum of squares is ", format(
      net$val), "\n")
  n <- length(net$fit)
  p <- length(net$wts)
  ft <- n * log(net$val/n)
  cat("AIC :", format(ft + 2 * p),
      ", SBC :", format(ft + p *
          log(n)), ", BIC :", format(
          ft + p + p * log(n)),
      ", MSE :", format(net$val/(n -
          p)), ", residual se :",
      format(sqrt(net$val/(n - p
          ))), "\n")
  invisible()
}
> y1 <- ANN[, 8, ]
> x1 <- ANN[, 1:7, ]
> data <- data.frame(y1, x1)
> datta <- rbind(data)
> samp <- c(sample(1:1264, 885), sample(1265:
```

```

2696, 1002))
> nn1 <- nnet(y1 ~ hub + kel + umur + status +
  pen + art + artk, size = 5, data =
  data, subset = samp, rang = 0.1,
  skip = F, trace = T, entropy = T,
  decay = 0.0005, maxit = 100)
# weights: 46
initial value 1303.439997
iter 10 value 693.218374
iter 20 value 597.916963
iter 30 value 573.417150
iter 40 value 552.506166
iter 50 value 533.428102
iter 60 value 526.024635
iter 70 value 523.206677
iter 80 value 520.307246
  iter 90 value 519.298638
iter 100 value 519.167645
final value 519.167645
stopped after 100 iterations
> sumfit.nnts(nn1)
  # Hasil pada data training dg ANN
a 7-5-1 network with 46 weights
decay=0.0005
Unit 0 is constant one input
Input units: Lag =1, Lag =2, Lag =3, Lag =4, Lag =5, Lag =6, Lag
=7,
Hidden units are 8 9 10 11 12
Output unit is 13

      0->8    1->8    2->8    3->8
  0.7831 -0.3320 -0.4052  0.3814
  4->8    5->8    6->8    7->8
-0.7297 -0.1877  0.1693  2.8235
  0->9    1->9    2->9    3->9
  4.5491  0.2413 -8.3332  0.6683
  4->9    5->9    6->9    7->9
  0.2232  0.0090  0.0606 -0.0338
  0->10   1->10   2->10   3->10
23.8046  0.0825 -1.9533 -1.9132
  4->10   5->10   6->10   7->10
-1.1716  0.4057 -18.5429 15.7155
  0->11   1->11   2->11   3->11
  2.2637 -1.9558  1.5869 -1.1590
  4->11   5->11   6->11   7->11
-0.4597  0.3099 -0.1208  0.2167
  0->12   1->12   2->12   3->12
  7.7203 -0.5133  3.4692 -3.2893
  4->12   5->12   6->12   7->12
-3.5522  0.0709  0.9416 -2.9833
  0->13   8->13   9->13  10->13
-13.0225 12.7697  9.3500 20.6196
  11->13  12->13
  9.8552 -3.9590

Sum of squares is 519.1676
AIC : -2343.205 , SBC : -2088.239 , BIC : -2042.239 , MSE :
  0.2820031 , residual se : 0.5310396
> train.pred <- format(round(predict(nn1,

```

```

    data[samp, ])))
> train.real <- y1[samp]
> table(train.pred, train.real)
# Hasil pada data testing dg ANN
  0   1
0 764 92
1 121 910
> test.pred <- format(round(predict(nn1, data[
  - samp, ])))
> test.real <- y1[ - samp]
> table(test.pred, test.real)
# Hasil Logistics regression
  0   1
0 329 47
1 50 383
> log1 <- glm(y1 ~ hub + kel + umur + status +
  pen + art + artk, family = binomial,
  data = data, subset = samp, trace
  = T)
GLM      linear loop 1: deviance = 1575.93
GLM      linear loop 2: deviance = 1410.436
GLM      linear loop 3: deviance = 1385.202
GLM      linear loop 4: deviance = 1384.25
GLM      linear loop 5: deviance = 1384.248
> summary(log1)$coefficients
# Hasil pada data training dg Logistics regression
  Value Std. Error
(Intercept) -1.0265582 0.45041580
  hub -0.5061074 0.06676201
  kel -2.4279250 0.16431118
  umur 0.6804175 0.13155719
  status 1.7467939 0.19108129
  pen 0.3232459 0.06652294
  art -0.5043280 0.05159379
  artk 1.7959922 0.10148175
  t value
(Intercept) -2.279134
  hub -7.580769
  kel -14.776384
  umur 5.172028
  status 9.141627
  pen 4.859165
  art -9.774976
  artk 17.697687
> train.plog <- format(round(predict.glm(log1,
  type = "response")))
> table(train.plog, train.real)
# Hasil pada data testing dg Logistics regression
  0   1
0 742 146
1 143 856
> test.plog <- format(round(predict.glm(log1,
  newdata = data[ - samp, ], type =
  "response")))
> table(test.plog, test.real)
  0   1
0 314 53
1 65 377

```

LAMPIRAN 6

Program Kernel dalam Syntax SAS

```
proc format;
  value specname
    1='Bekerja'
    0='Tidak bekerja';
  run;
data bekerja;
  title 'Analisis regresi Kernel untuk pengelompokan usia kerja dalam angkatan kerja';
  input Y X1 X2 X3 X4 X5 X6 X7 ;
  format Y specname.;
  label X1='Hub dengan KRT.'
    X2='Jenis kelamin.'
    X3='Kelompok umur.'
    X4='Status perkawinan.'
    X5='Tingkat pendidikan.'
    X6='Jumlah ART.'
    X7= 'Jumlah ART bekerja。';
  cards;
Data;
proc discrim data= bekerja outd=outd
method=npar kernel=normal r=.287 pool=yes
short noclassify croslisterr;
class Y;
priors prop;
var X1 X2 X3 X4 X5 X6 X7 ;
title2'Using normal dencity estimates with equal variance';
run;
proc discrim data= bekerja outd=outd
method=npar kernel=normal r=.287 pool=no
short noclassify croslisterr;
class Y;
priors prop;
var X1 X2 X3 X4 X5 X6 X7 ;
title2'Using normal dencity estimates with unequal variance';
run;
```

LAMPIRAN 7

Output Metode Kernel Data Training

Analisis regresi Kernel untuk pengelompokan usia kerja dalam angkatan kerja

1

Using normal dencity estimates with equal variance
10:25 Sunday,

October 5, 1997

Discriminant Analysis

1887 Observations 1886 DF Total
7 Variables 1885 DF Within Classes
2 Classes 1 DF Between Classes

Class Level Information

Y	Frequency	Weight	Proportion	Prior Probability
Bekerja	1002	1002	0.531002	0.531002
Tidak bekerja	885	885.0000	0.468998	0.468998

Number of Observations and Percent Classified
into Y:

From Y	Bekerja	Tidak bekerja	Total
Bekerja	904	98	1002
	90.22	9.78	100.00
Tidak bekerja	86	799	885
	9.72	90.28	100.00
Total	990	897	1887
Percent	52.46	47.54	100.00
Priors	0.5310	0.4690	

Error Count Estimates for Y:

	Bekerja	Tidak bekerja	Total
Rate	0.0978	0.0972	0.0975
Priors	0.5310	0.4690	

Analisis regresi Kernel untuk pengelompokan usia kerja dalam angkatan kerja

15

Using normal dencity estimates with unequal variance
10:25 Sunday,

October 5, 1997

Discriminant Analysis Classification Summary for Calibration Data:
WORK.BEKERJA

Cross-validation Summary using Normal Kernel Density

		Number of Observations and Percent Classified		
into Y:		Bekerja	Tidak bekerja	Total
From Y				
Bekerja		880	i 122	1002
		87.82	12.18	100.00
Tidak bekerja		88	797	885
		9.94	90.06	100.00
Total		968	919	1887
Percent		51.30	48.70	100.00
Priors		0.5310	0.4690	

Error Count Estimates for Y:

	Bekerja	Tidak bekerja	Total
Rate	0.1218	0.0994	0.1113
Priors	0.5310	0.4690	

LAMPIRAN 8

Output Metode Kernel Data Testing

Analisis regresi Kernel untuk pengelompokan usia kerja dalam angkatan kerja
1
Using normal dencity estimates with equal variance
10:46 Sunday,

October 5, 1997

Discriminant Analysis

808 Observations 807 DF Total
7 Variables 806 DF Within Classes
2 Classes 1 DF Between Classes

Class Level Information

Y	Frequency	Weight	Proportion	Prior Probability
Bekerja	429	429.0000	0.530941	0.530941
Tidak bekerja	379	379.0000	0.469059	0.469059

Number of Observations and Percent Classified
into Y:

From Y	Bekerja	Tidak bekerja	Total
Bekerja	381 88.81	48 11.19	429 100.00
Tidak bekerja	49 12.93	330 87.07	379 100.00
Total	430	378	808
Percent	53.22	46.78	100.00
Priors	0.5309	0.4691	

Error Count Estimates for Y:

	Bekerja	Tidak bekerja	Total
Rate	0.1119	0.1293	0.1200
Priors	0.5309	0.4691	

Analisis regresi Kernel untuk pengelompokan usia kerja dalam angkatan kerja

6

Using normal dencity estimates with unequal variance
10:46 Sunday,

October 5, 1997

Discriminant Analysis

808 Observations	807 DF Total
7 Variables	806 DF Within Classes
2 Classes	1 DF Between Classes

Class Level Information

Y	Frequency	Weight	Proportion	Prior
				Probability
Bekerja	429	429.0000	0.530941	0.530941
Tidak bekerja	379	379.0000	0.469059	0.469059

Number of Observations and Percent Classified
into Y:

From Y	Bekerja	Tidak bekerja	Total
Bekerja	379	50	429
	88.34	11.66	100.00
Tidak bekerja	48	331	379
	12.66	87.34	100.00
Total	427	381	808
Percent	52.85	47.15	100.00
Priors	0.5309	0.4691	

Error Count Estimates for Y:

	Bekerja	Tidak bekerja	Total
Rate	0.1166	0.1266	0.1213
Priors	0.5309	0.4691	