



Tesis - SS14 2501

**PEMODELAN PEMILIHAN MODA TRANSPORTASI
PADA KOMUTER DI JABODETABEK
MENGUNAKAN *BAYESIAN NETWORK*
DAN REGRESI LOGISTIK *POLYTOMOUS***

Ratih Kusuma Dewi
NRP. 06211650017003

DOSEN PEMBIMBING
Prof. Drs. Nur Iriawan, M.Ikomp., Ph.D.
Irhamah, M.Si., Ph.D.

PROGRAM MAGISTER
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018



Tesis - SS14 2501

**PEMODELAN PEMILIHAN MODA TRANSPORTASI
PADA KOMUTER DI JABODETABEK
MENGUNAKAN *BAYESIAN NETWORK*
DAN REGRESI LOGISTIK *POLYTOMOUS***

Ratih Kusuma Dewi
NRP. 06211650017003

DOSEN PEMBIMBING
Prof. Drs. Nur Iriawan, M.lkomp., Ph.D.
Irhamah, M.Si., Ph.D.

PROGRAM MAGISTER
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018



Thesis - SS14 2501

TRANSPORTATION CHOICE MODELING ON COMMUTERS IN JABODETABEK USING BAYESIAN NETWORK AND POLYTOMOUS LOGISTIC REGRESSION

Ratih Kusuma Dewi
NRP. 06211650017003

SUPERVISOR:
Prof. Drs. Nur Iriawan, M.lkomp., Ph.D.
Irhamah, M.Si., Ph.D.

MAGISTER PROGRAM
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTATION, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018

**PEMODELAN PEMILIHAN MODA TRANSPORTASI
PADA KOMUTER DI JABODETABEK
MENGUNAKAN *BAYESIAN NETWORK*
DAN REGRESI LOGISTIK *POLYTOMOUS***


Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)

di
Institut Teknologi Sepuluh Nopember
Oleh:


**RATIH KUSUMA DEWI
NRP. 06211650017003**


Tanggal Ujian : 16 Januari 2018
Periode Wisuda : Maret 2018

Disetujui Oleh:


1. Prof. Drs. Nur Iriawan, M. Ikomp., Ph.D. (Pembimbing I)
NIP.19621015 198803 1 002

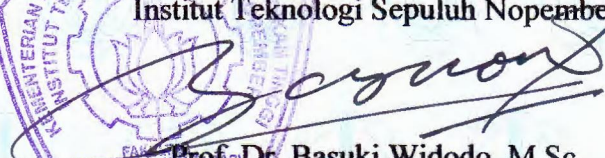

2. Irhamah, M.Si., Ph.D. (Pembimbing II)
NIP.19780406 200112 2 002


3. Dr. Kartika Fithriasari, M.Si. (Penguji)
NIP.19691212 199303 2 002


4. Dr. Erni Tri Astuti, M. Math. (Penguji)
NIP.19671022 199003 2 002



Dekan
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember


Prof. Dr. Basuki Widodo, M.Sc.
NIP.19650605 198903 1 002

**Pemodelan Pemilihan Moda Transportasi pada Komuter
di Jabodetabek Menggunakan *Bayesian Network*
dan Regresi Logistik *Polytomous***

Nama : Ratih Kusuma Dewi
NRP : 06211650017003
Dosen Pembimbing : Prof. Drs. Nur Iriawan, M.Ikomp., Ph.D.
Irhamah, M.Si, Ph.D.

ABSTRAK

Fenomena *urban sprawl* yang terjadi di DKI Jakarta menyebabkan pemekaran daerah disekitarnya. *Urban sprawl* merupakan proses perembetan kenampakan fisik kota ke arah luar yang disebabkan oleh pesatnya perkembangan ekonomi dan pertumbuhan penduduk. Kawasan “Bodetabek” berubah fungsi menjadi pemukiman yang padat penduduk bahkan menjadi kawasan industri. Oleh karena itu, sebagian penduduk Bodetabek maupun DKI Jakarta meninggalkan rumah untuk melakukan aktivitasnya di luar wilayah administrasi tempat tinggalnya dan kembali pada hari itu juga. Pelaku migrasi semacam ini biasanya disebut sebagai komuter. Pergerakan rutin para komuter Jabodetabek menjadi penyebab utama kemacetan lalu lintas. Peralihan moda transportasi komuter dari kendaraan pribadi ke angkutan umum merupakan solusi untuk mengurangi kemacetan. Meski demikian, belum semua komuter berminat menggunakan angkutan umum. Hasil dari Survei Komuter 2014 menunjukkan bahwa hanya 27 persen komuter Jabodetabek yang menggunakan kendaraan umum untuk pergi ke tempat tujuan. *Bayesian Network* (BN) dan regresi logistik digunakan dalam pemodelan pemilihan moda transportasi di Jabodetabek. Regresi logistik sudah banyak digunakan dalam pemodelan klasifikasi. Sementara BN merupakan model yang mampu menjelaskan struktur hubungan antara berbagai variabel random yang kompleks ke dalam bentuk diagram berdasarkan teori peluang bersyarat. Pada penelitian ini akan dibandingkan BN dan Regresi Logistik *Polytomous* dalam pemodelan pemilihan moda transportasi di Jabodetabek. Penggunaan Regresi Logistik *Polytomous* disebabkan karena jumlah kategori pada variabel respon lebih dari dua. Hasil penelitian menunjukkan bahwa pada data *testing*, metode Regresi Logistik *Polytomous* (*enter*) mempunyai *Correct Classification Rate* (CCR) dan *Area Under The Curve* (AUC) tertinggi, yang berarti metode tersebut memberikan akurasi yang tertinggi.

Kata kunci: *Bayesian Network*, Regresi Logistik *Polytomous*, komuter, Jabodetabek

**Transportation Choice Modeling on Commuters
in Jabodetabek Using Bayesian Network
and Polytomous Logistic Regression**

Name : Ratih Kusuma Dewi
NRP : 06211650017003
Supervisor : Prof. Drs. Nur Iriawan, M.Ikomp., Ph.D.
Irhamah, M.Si, Ph.D.

ABSTRACT

The urban sprawl phenomenon that occurred in DKI Jakarta led to the expansion of the surrounding area. Urban sprawl is a process of physical exposure to the outer city physical appearance caused by the rapid economic development and population growth. The "Bodetabek" area has turned into a densely populated residence and even an industrial estate. Compromise made by citizens of Bodetabek and DKI Jakarta leave the house to carry out its activities outside the administrative area of his residence and back on that day. People who do this kind of migration are called commuters. The routine movement of Jabodetabek commuter is the main cause of traffic jam. The transition of commuter transport modes from private vehicles to public transport is a solution to reduce congestion. However, not all commuters are interested in using public transportation. The results from the 2014 Komuter Survey show that only 27 percent of Jabodetabek commuters use public transport to go to their destination. Bayesian Network (BN) and logistic regression are used in modeling of transportation choice in Jabodetabek. Logistic regression is widely used in classification modeling. While BN is a capable model of explaining the structure of relationships between complex random variables into diagrammatic forms based on conditional probability theory. In this research, BN will be compared with polytomous logistic regression in modeling of transportation choice in Jabodetabek. The use of polytomous logistic regression is due to the number of categories in the response variable more than two. The results showed that the highest value of *Correct Classification Rate* (CCR) and *Area Under The Curve* (AUC) is achieved by Polytomous Logistic Regression (enter) in testing data, it means this method produce the best accuracy.

Keywords : *Bayesian Network, Polytomous Logistic Regression* , commuter,
Jabodetabek

KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT yang telah melimpahkan karunia nikmat dan rahmat kepada penulis sehingga tesis yang berjudul “Pemodelan Pemilihan Moda Transportasi pada Komuter di Jabodetabek Menggunakan *Bayesian Network* dan Regresi Logistik *Polytomous*” ini dapat diselesaikan sesuai dengan waktu yang diharapkan. Tesis ini disusun dalam rangka memenuhi salah satu syarat untuk meraih gelar Magister Sains di Program Pasca Sarjana Statistika Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Badan Pusat Statistik (BPS) yang telah memberi kesempatan, dukungan dan beasiswa kepada penulis untuk melanjutkan studi program S2 di ITS.
2. Bapak Prof. Drs. Nur Iriawan, M.Ikomp, Ph.D dan Ibu Irhamah, M.Si, Ph.D selaku pembimbing atas segala bimbingan, arahan dan motivasi dalam penyusunan tesis ini.
3. Ibu Dr. Kartika Fithriasari, M.Si. dan Dr. Erni Tri Astuti, M. Math. selaku dosen penguji yang telah memberikan saran dan koreksi atas penulisan tesis ini.
4. Bapak Dr. Suhartono, M.Sc selaku Ketua Jurusan Statistika ITS dan Bapak Dr. rer.pol. Heri Kuswanto, M.Si selaku Kaprodi Pasca Sarjana Statistika ITS beserta jajarannya atas kemudahan dan fasilitas yang diberikan selama studi.
5. Bapak dan Ibu dosen Statistika ITS yang telah mencurahkan ilmu dan pengalamannya selama proses studi.
6. Ibu, bapak, adikku tercinta atas doa, bantuan dan dukungannya selama masa studi, Mas Uta yang telah memberikan izin penulis untuk kuliah, dan seluruh keluarga di Magelang atas doa dan motivasi yang diberikan.
7. Sahabat-sahabatku, Mbak Prih, Mbak Mike, Mbak Imra, Mas Taufiq, Bang Fieldri, Sony, Mbak Nana, Mbak Tika, Bang Fael, Mbak Reni, Mas Fendy, Mas Prapto, Mas Umam dan Mas Aniq atas semua dukungan kepada penulis. Tak lupa teman sebangkuan dan seperjuangan Mbak Dapih, Prizka, Indah, Febri, dan Silvi.

8. Ibu Sarni dan Ibu Arifatus atas bimbingan, doa, dan dukungannya kepada penulis.
9. Almarhumah Ibu Anday, Random people, Dawels yang selalu ada dan memberikan dukungan kepada penulis.
10. Semua pihak yang telah membantu penyelesaian tesis ini. Semoga tesis ini dapat bermanfaat dan menambah wawasan keilmuan. Saran dan kritik yang membangun dari semua pihak, sangat penulis harapkan untuk perbaikan selanjutnya.

DAFTAR ISI

ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	5
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	6
BAB 2 TINJAUAN PUSTAKA	7
2.1 <i>Bayesian Network</i>	7
2.1.1 Teorema Bayesian	7
2.1.2 Struktur <i>Bayesian Network</i>	8
2.1.3 Kebebasan Bersyarat	10
2.1.4 Estimasi Parameter <i>Bayesian Network</i>	11
2.2 Uji Independensi	15
2.2.1 Uji <i>Chi-Square</i>	16
2.2.2 Uji <i>Conditional Independence</i>	17
2.3 <i>Naïve Bayes</i> (NB)	18
2.4 <i>Hierarchical Naïve Bayes</i> (HNB)	20
2.5 Asumsi Regresi Logistik	25
2.6 Regresi Logistik	26
2.6.1 Estimasi Parameter Regresi Logistik <i>Polytomous</i>	31

2.6.2 Uji Signifikansi Parameter	32
2.6.3 Seleksi Variabel	33
2.6.4 <i>Akaike Information Criterion</i> (AIC)	34
2.7 <i>Correct Clasification Rate</i>	35
2.8 <i>Area Under The Curve</i> (AUC)	35
2.9 Migrasi Penduduk	37
2.10 Migrasi Ulang Alik (Komuter)	39
2.11 Moda Transportasi	40
METODOLOGI PENELITIAN	45
3.1 Sumber Data	45
3.2 Metode Pengumpulan Data	45
3.3 Kerangka Pikir	46
3.4 Variabel Penelitian	47
3.5 Struktur Data	51
3.6 Metode dan Tahapan Penelitian	51
HASIL DAN PEMBAHASAN	55
4.1 Karakteristik Komuter Jabodetabek	55
4.1.1 Karakteristik Sarana Prasarana Kota Tujuan	57
4.1.2 Karakteristik Perjalanan	60
4.1.3 Karakteristik Sosio-Demografi Komuter	62
4.2 Pemodelan dengan <i>Naïve Bayes</i> (NB)	63
4.2.1 Estimasi Parameter Model <i>Naïve Bayes</i>	63
4.2.2 Klasifikasi dengan Metode <i>Naïve Bayes</i>	65
4.3 Pemodelan dengan <i>Hierarchical Naïve Bayes</i> (HNB)	71
4.3.1 Independensi Antar Variabel Atribut	72
4.3.2 Pembentukan Variabel Laten	75
4.3.3 Klasifikasi dengan Metode <i>Hierarchical Naïve Bayes</i>	78
4.4 Pemodelan dengan Regresi Logistik <i>Polytomous</i>	82
4.4.1 Regresi Logistik <i>Polytomous</i> dengan Metode <i>Enter</i>	83
4.4.2 Regresi Logistik <i>Polytomous</i> dengan Metode <i>Stepwise</i>	86

4.5	Pemodelan <i>Naïve Bayes</i> dengan Input Variabel Hasil Regresi Logistik <i>Polytomous</i>	92
4.5.1	Input Variabel pada Metode <i>Naïve Bayes</i>	92
4.5.2	Klasifikasi <i>Naïve Bayes</i> dengan Input Variabel Hasil Regresi Logistik <i>Polytomous</i>	92
4.6	Perbandingan Hasil Klasifikasi	97
KESIMPULAN DAN SARAN.....		101
5.1	Kesimpulan.....	101
5.2	Saran	102
DAFTAR PUSTAKA		105
LAMPIRAN.....		109

DAFTAR TABEL

Tabel 2.1	Tabel Kontingensi Frekuensi Dua Arah pada Variabel A dan B ...	16
Tabel 2.2	Tabel Kontingensi ROC	35
Tabel 3.1	Variabel Penelitian	47
Tabel 3.2	Struktur Data	51
Tabel 4.1	Status Komuter Penduduk Jabodetabek	56
Tabel 4.2	Banyaknya Komuter Menurut PDRB Kota Tujuan	59
Tabel 4.3	Jarak Tempuh, Waktu Tempuh dan Biaya Transportasi Komuter	60
Tabel 4.4	Sebaran Jarak Tempuh, Waktu Tempuh dan Biaya Transportasi Komuter	61
Tabel 4.5	Karakteristik Individu Komuter Jabodetabek	62
Tabel 4.6	Estimasi Peluang Prior $\hat{P}(Y)$ <i>Naïve Bayes</i>	67
Tabel 4.7	Estimasi Parameter Peluang Bersyarat $\hat{P}(X Y)$ <i>Naïve Bayes</i>	67
Tabel 4.8	Nilai <i>p-value</i> Hasil Uji <i>Chi-square</i> pada 11 Variabel Atribut	73
Tabel 4.9	Nilai <i>p-value</i> Hasil Uji <i>Conditional Independen</i> pada 11 Variabel Atribut	74
Tabel 4.10	Estimasi Peluang Prior $\hat{P}(Y)$ <i>Hierarchical Naïve Bayes</i>	79
Tabel 4.11	Estimasi Parameter $\hat{P}(X Y)$ <i>Hierarchical Naïve Bayes</i>	79
Tabel 4.12	Perbandingan AIC pada Regresi Logistik <i>Polytomous</i>	87
Tabel 4.13	Estimasi Peluang Prior $\hat{P}(Y)$ dengan Input Variabel Hasil Regresi Logistik <i>Polytomous</i>	94
Tabel 4.14	Estimasi Parameter $\hat{P}(X Y)$ dengan Input Variabel Hasil Regresi Logistik <i>Polytomous</i>	94
Tabel 4.15	Perbandingan Nilai <i>Correct Clasification Rate</i> (CCR) dan <i>Area</i> <i>Under The Curve</i> (AUC) – <i>Data Training</i>	97
Tabel 4.16	Perbandingan Nilai <i>Correct Clasification Rate</i> (CCR) dan <i>Area</i> <i>Under The Curve</i> (AUC) – <i>Data Training</i>	98
Tabel 4.17	Perbandingan Lama Waktu Proses <i>Running Syntax</i>	98

DAFTAR GAMBAR

Gambar 2.1 Contoh Struktur DAG	9
Gambar 2.2 Contoh Hubungan antar <i>Node</i>	9
Gambar 2.3 Struktur DAG	11
Gambar 2.4 Contoh Struktur <i>Naïve Bayes</i>	18
Gambar 2.5 Contoh Struktur HNB	21
Gambar 2.6 Contoh Kurva ROC	36
Gambar 2.7 Hubungan antara Kebutuhan dan Pola Migrasi Penduduk	37
Gambar 2.8 Migrasi Non Permanen	39
Gambar 3.1 Kerangka Pikir	46
Gambar 3.2 Diagram Alir Penelitian	53
Gambar 4.1 Persentase Komuter di Jabodetabek	55
Gambar 4.2 Penggunaan Moda Transportasi oleh Komuter Jabodetabek	57
Gambar 4.3 Kota Tujuan Komuter Jabodetabek	58
Gambar 4.4 Persentase Jumlah Pergantian Moda Transportasi	59
Gambar 4.5 Struktur <i>Naïve Bayes</i> Klasifikasi Moda Transportasi	66
Gambar 4.6 Pembentukan Variabel Laten	76
Gambar 4.7 Struktur <i>Hierarchical Naïve Bayes</i> Klasifikasi Moda Transportasi	78
Gambar 4.8 Struktur NB dengan Input Hasil Regresi Logistik <i>Polytomous</i> Klasifikasi Moda Transportasi	93

DAFTAR LAMPIRAN

Lampiran 1	Moda Transportasi yang Digunakan.....	109
Lampiran 2	<i>Syntax</i> Pengambilan Sampel Data <i>Training - Testing</i>	110
Lampiran 3	<i>Syntax Naïve Bayes</i>	111
Lampiran 4	<i>Syntax Naïve Bayes</i>	112
Lampiran 5	Output <i>Naïve Bayes</i>	113
Lampiran 6	<i>Syntax Hierarchical Naïve Bayes</i>	118
Lampiran 7	Output <i>Hierarchical Naïve Bayes</i>	120
Lampiran 8	<i>Syntax</i> Regresi Logistik <i>Polytomous</i>	129
Lampiran 9	Output Regresi Logistik <i>Polytomous</i>	130
Lampiran 10	<i>Syntax</i> Regresi Logistik <i>Polytomous – Naïve Bayes</i>	135
Lampiran 11	Output Regresi Logistik <i>Polytomous – Naïve Bayes</i>	136

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Provinsi DKI Jakarta sebagai Ibukota Negara Indonesia tidak hanya merupakan pusat pemerintahan tetapi juga merupakan pusat perekonomian nasional. Kepadatan penduduk di Provinsi DKI Jakarta sebesar 14.518 per km² (hasil Sensus Penduduk 2010) merupakan yang tertinggi di Indonesia. Hal ini sangat timpang dibandingkan provinsi terpadat kedua dan ketiga yaitu Provinsi Jawa Barat dengan kepadatan penduduk sebesar 1.222 per km² dan Provinsi D.I Yogyakarta dengan kepadatan penduduk sebesar 1.107 per km².

Tamin (2000) menjelaskan bahwa semakin tingginya harga lahan di pusat perkotaan sebagai pusat kegiatan perekonomian menyebabkan lahan permukiman semakin bergeser ke pinggiran kota. Kemudian dipertegas oleh Giyarsih (2001), perkembangan pesat suatu kota, khususnya dalam bidang ekonomi, menyebabkan padatnya penduduk dan bangunan sehingga kota tersebut tidak mampu menampung lebih banyak penduduk lagi. Akibatnya, ada pergeseran fungsi-fungsi kota ke daerah pinggiran kota (*urban fringe*) atau proses perembetan kenampakan fisik kota ke arah luar (*urban sprawl*).

Dengan kepadatan penduduk yang sangat tinggi dan lahan yang sudah tidak berkembang lagi di Provinsi DKI Jakarta, maka perumahan di Provinsi DKI Jakarta menjadi sangat mahal. Hal ini menyebabkan penduduk yang beraktivitas di Provinsi DKI Jakarta tetapi tidak mampu membeli perumahan di Jakarta memilih tinggal di kota-kota sekitarnya. Kota Bogor, Depok, Tangerang, maupun Bekasi akhirnya menjadi pilihan untuk bermukim. Kota-kota seperti ini sering disebut kota satelit, yaitu kota kecil ditepi sebuah kota besar yang meskipun merupakan komunitas mandiri, sebagian besar penduduknya tergantung dengan kehidupan di kota besar. Kompromi yang dilakukan menyebabkan penduduk melakukan kegiatan di luar wilayah tempat tinggalnya. Penduduk yang melakukan perjalanan dan meninggalkan rumah untuk pergi ke tempat lain dan mereka akan

selalu berusaha untuk kembali pada hari itu juga disebut sebagai komuter (Mantra, 2000).

Migrasi penduduk merupakan kebutuhan naluriah manusia. Migrasi dilakukan untuk mempertahankan hidup dan disebabkan karena adanya kesenjangan antara satu wilayah dengan wilayah yang lain. Mengingat pentingnya migrasi, pemerintah mengatur migrasi penduduk dalam Undang-Undang Nomor 52 Tahun 2009 tentang Perkembangan Kependudukan dan Pembangunan Keluarga dan dijabarkan dalam Peraturan Pemerintah Nomor 57 Tahun 2009 tentang Pengelolaan Perkembangan Kependudukan.

Menurut Ananta dan Chotib (1996), migrasi penduduk Indonesia terbagi menjadi dua yaitu migrasi permanen (menetap di daerah tujuan) dan non permanen (yang tidak menetap di daerah tujuan). Hal senada juga diungkapkan Mantra dan Sunarto (1986), perbedaan migrasi permanen dan migrasi non permanen terletak pada ada atau tidaknya niat untuk bertempat tinggal menetap di daerah tujuan. Migrasi permanen dapat dikatakan bertujuan pindah tempat tinggal secara tetap dan migrasi non permanen tidak mempunyai tujuan pindah tetapi memiliki tempat aktivitas yang berbeda dengan tempat tinggalnya.

Migrasi non permanen dapat dibedakan menjadi dua kategori berdasarkan aspek rutinitas yaitu sirkuler (migrasi sirkulasi) dan komuter (migrasi ulang alik). Apabila penduduk yang melakukan perpindahan sejak semula sudah bermaksud untuk kembali ke daerah asalnya secara periodik disebut sebagai migrasi sirkuler. Mantra (1981) menegaskan bahwa komuter adalah pergerakan yang dilakukan dalam waktu satu hari dengan melintasi batas wilayah dan kembali ketempat asal.

Keberadaan komuter dapat memberikan dampak positif bagi pusat kota, yaitu mengurangi kepadatan penduduk kota besar serta semakin berkembangnya daerah pinggiran sebagai tempat tinggal para komuter. Selain itu, keberadaan komuter akan membawa pengaruh terhadap kehidupan ekonomi dan sosial budaya pada penduduk daerah tujuan maupun daerah asal. Perilaku masyarakat yang ditemui komuter di daerah tujuan akan dibawa kembali pada masyarakat di daerah asalnya. Sementara itu kebiasaan-kebiasaan yang terdapat pada daerah asal

komuter juga akan terbawa ke daerah tujuan. Keberadaan komuter ini membutuhkan infrastruktur dan sarana transportasi yang memadai.

Berdasarkan Survei Komuter Jabodetabek 2014, sekitar 61,09 persen komuter yang berasal dari Bodetabek beraktivitas di Provinsi DKI Jakarta. Sebaliknya, sekitar 19,64 persen komuter yang berasal dari Provinsi DKI Jakarta beraktivitas di Bodetabek. Jalan-jalan menuju pintu masuk DKI Jakarta dipadati oleh kendaraan baik roda empat ataupun roda dua pada setiap pagi hari. Kendaraan-kendaraan dari kota satelit DKI Jakarta seperti Bogor, Depok, Tangerang, dan Bekasi akan kembali memadati jalan pada sore atau malam hari untuk kembali ke daerah asal. Sama halnya seperti fenomena kepadatan arus kendaraan di jalan raya, kepadatan juga terjadi di dalam gerbong kereta serta bus Transjakarta yang mempunyai rute dalam kota Jakarta. Oleh karena itu, Pemerintah Provinsi DKI Jakarta bekerja sama dengan pemerintahan daerah di sekitarnya berupaya menyelenggarakan transportasi umum lintas batas administrasi seperti KRL dari Bodetabek menuju Jakarta atau sebaliknya dan juga angkutan masal berbasis bus, seperti angkutan perbatasan terintegrasi bus Transjakarta (APTJ) serta bus Transjabodetabek. Selain itu, *Mass Rapid Transit* (MRT) sebagai pilihan moda transportasi baru di Jabodetabek juga sedang dalam tahap pembangunan. Di dalam kota Jakarta sendiri, Pemerintah Provinsi DKI terus menambah jumlah armada bus Transjakarta.

Meski demikian, belum semua komuter berminat menggunakan angkutan umum. Hasil Survei Komuter 2014 menunjukkan bahwa hanya sebesar 27 persen komuter Jabodetabek yang menggunakan kendaraan umum untuk pergi ke tempat tujuan. Selebihnya, sebanyak 70,96 persen menggunakan kendaraan pribadi dan lainnya.

Berdasarkan uraian di atas, kajian pemilihan moda transportasi pada komuter di Jabodetabek sangat diperlukan mengingat pergerakan rutin para komuter itu menjadi salah satu penyumbang kemacetan lalu lintas. Peralihan moda transportasi komuter dari kendaraan pribadi ke angkutan umum merupakan solusi untuk mengurangi kemacetan. Dalam hal ini, penulis terdorong untuk melakukan pemodelan pemilihan moda transportasi pada komuter di Jabodetabek.

Beberapa penelitian terdahulu terkait migrasi dilakukan oleh Sanjaya (2012). Penelitian tersebut menggunakan regresi logit hirarki yaitu nama lain dari multilevel logistik dalam memodelkan pembuatan keputusan migrasi di Indonesia menggunakan data Survei Penduduk Antar Sensus (SUPAS) 2005. Level pertama pada penelitian ini adalah karakteristik individu dan karakteristik rumah tangga sedangkan level kedua yaitu kabupaten/kota. Pengembangan penelitian spasial khususnya untuk migrasi ulang-alik atau komuter dilakukan oleh Lovelace, Ballas dan Watson (2014). Mereka mengkombinasi data level individu dan area untuk meneliti pola komuter di UK dengan metode *spatial microsimulation*. Roslan, Abdullah, dan Abdullah (2015) menggunakan analisis jalur untuk memodelkan keluarga komuter, tentang hubungan kesiapan orang tua, lingkungan keluarga dan kinerja sekolah remaja. Kemudian, Ma (2015) menerapkan *Bayesian Network* (BN) untuk pemodelan pilihan mode multimodal pada pekerja di perbatasan Luxembourg. Hasilnya dibandingkan dengan multinomial logit model (MNL) dan didapatkan nilai prediksi yang mirip. Namun MNL tidak dapat menggambarkan hubungan sebab akibat antara variabel penjelas.

Data yang digunakan dalam penelitian ini berskala nominal dengan banyak kategori (*polytomous*) yang diperoleh dari Survei komuter Jabodetabek 2014. Metode penelitian yang selama ini digunakan untuk data yang bersifat *polytomous* adalah Regresi Logistik *Polytomous*. Pendekatan lain yang dapat digunakan untuk menganalisis data yang bersifat *polytomous* adalah *Bayesian Network* (BN). BN merupakan pendekatan yang digunakan untuk mengetahui hubungan antar variabel dalam bentuk jaringan (*networks*). Friedman, Geiger, dan Goldszmidt (1997) dalam penelitiannya menyebutkan bahwa *Naïve Bayes*, yang merupakan *Bayesian classifiers* sederhana dengan asumsi independensi yang kuat di antara variabel, bersaing dengan pengklasifikasi lainnya. Sejalan dengan penelitian tersebut, penelitian yang dilakukan oleh Ashari, Paryudi, dan Tjoa (2013) yang membandingkan kinerja antara *Naïve Bayes*, *Decision Tree* and *k-Nearest Neighbor* dalam Pencarian Desain Alternatif dalam Alat Simulasi Energi menunjukkan bahwa meskipun *Naïve Bayes* merupakan metode yang sederhana, tetapi *Naïve Bayes* bisa mengungguli metode klasifikasi yang lebih kompleks.

Berdasarkan pertimbangan tersebut, maka dalam penelitian ini akan digunakan metode *Bayesian Network* (BN) dan Regresi Logistik *Polytomous* untuk memodelkan pemilihan moda transportasi pada komuter di Jabodetabek. BN merupakan model yang mampu menjelaskan struktur hubungan antara berbagai variabel random yang kompleks ke dalam bentuk diagram berdasarkan teori peluang bersyarat. Model Regresi Logistik *Polytomous* merupakan perluasan dari model regresi logistik yang bersifat *dikotomous* (biner). Apabila regresi logistik biner mensyaratkan variabel respon terdiri dari dua kategori, maka variabel respon pada Regresi Logistik *Polytomous* terdiri dari tiga kategori atau lebih.

Salah satu klasifikasi BN yang dikenal paling efektif adalah *Naïve Bayes* (Friedman, dkk, 1997). Pendekatan ini menghubungkan semua variabel prediktor dengan variabel respon dengan asumsi bahwa semua variabel prediktor saling independen. Namun sayangnya, seringkali variabel-variabel prediktor berhubungan, bahkan mempunyai pengaruh tidak langsung akan membentuk struktur hirarki. Terkait dengan fenomena ini, Zhang, Nielsen, dan Jensen (2004) mengembangkan *Naïve Bayes* (NB) ke dalam model *Hidden Naïve Bayes* yang selanjutnya dikembangkan lagi oleh Langseth dan Nielsen (2006) menjadi *Hierarchical Naïve Bayes* (HNB). Dalam pendekatan HNB hubungan antar variabel atribut dapat diakomodasi dengan baik. Oleh karena itu, dalam penelitian ini metode BN yang digunakan selain NB juga pendekatan HNB.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, pokok permasalahan yang ingin diteliti adalah bagaimana membangun model *Bayesian Network* (BN) dalam pemodelan pemilihan moda transportasi pada komuter di Jabodetabek serta perbandingan BN dengan Regresi Logistik *Polytomous*.

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas, maka tujuan dari penelitian ini adalah sebagai berikut:

- a. Mendapatkan model *Bayesian Network* (BN) untuk pemilihan moda transportasi pada komuter di Jabodetabek
- b. Membandingkan metode *Bayesian Network* (BN) dengan metode Regresi Logistik *Polytomous* untuk pemilihan moda transportasi pada komuter di Jabodetabek.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini yaitu untuk memberikan informasi tentang faktor-faktor yang mempengaruhi pemilihan moda transportasi pada komuter di Jabodetabek sehingga dapat dijadikan masukan kepada pemerintah Provinsi DKI Jakarta serta pemerintah daerah Bodetabek sebagai solusi dalam meningkatkan minat komuter dalam penggunaan kendaraan umum dan peningkatan kualitas sarana prasarana moda transportasi mengingat keterkaitan antara wilayah dan penduduk ini membutuhkan penanganan secara menyeluruh, tidak hanya secara parsial, dimana kawasan Jabodetabek memiliki administratif pemerintahan yang berbeda.

1.5 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Wilayah yang dijadikan objek kajian adalah Jabodetabek
2. Sumber data yang digunakan hanya dari Survei Komuter Jabodetabek 2014 yang dikumpulkan oleh Badan Pusat Statistik. Oleh karena data yang digunakan dalam penelitian ini adalah data tahun 2014, maka perkembangan transportasi di Jabodetabek setelah tahun 2014 tidak tercakup dalam penelitian ini.
3. Model *Bayesian Network* yang digunakan adalah metode *Naïve Bayes* (NB) dan *Hierarchical Naïve Bayes* (HNB) dengan pembentukan variabel laten berdasarkan nilai *conditional independence*.

BAB 2

TINJAUAN PUSTAKA

2.1 *Bayesian Network*

Bayesian Network (BN) merupakan suatu metode pemodelan data berbasis probabilitas yang merepresentasikan hubungan suatu himpunan variabel dengan ketergantungan bersyarat melalui suatu DAG (*Directed Acyclic Graph*). BN dapat didefinisikan sebagai kejadian $B = (G, \Theta)$, dimana G adalah model DAG dan Θ menunjukkan himpunan parameter dari BN. Model DAG G terdiri dari himpunan *node* (V) dan *edge* (E) sehingga himpunan graph G dinotasikan $G = (V, E)$. *Node* merepresentasikan variabel random.

Menurut Korb dan Nicholson (2010), *node* merepresentasikan variabel random dan *edge* merepresentasikan hubungan langsung antar variabel yang digambarkan sebagai anak panah antara node. Setiap *node* harus memenuhi sifat *mutually exclusive* dan *mutually exhaustive*, artinya setiap kejadian memiliki kesempatan untuk menjadi anggota dari setiap *node* dan menjadi anggota tepat satu kategori dari *node* tersebut. Pemodelan BN terdiri atas pembuatan struktur jaringan dan estimasi nilai peluang setiap *node* dalam *graph* struktur jaringan yang telah dibentuk.

Secara garis besar, pemodelan dalam *Bayesian Network* meliputi:

1. Struktur *learning*, yaitu membuat struktur jaringan
2. Parameter *learning*, yaitu mengestimasi nilai peluang setiap node dalam *graph* struktur jaringan yang telah dibentuk

2.1.1 Teorema Bayes

Teorema Bayes dikemukakan oleh Thomas Bayes di Inggris pada tahun 1763. Teorema Bayes ini kemudian disempurnakan oleh Laplace. Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan pengaruh yang didapat dari hasil observasi. Dalam teorema Bayes, jika Y dan X

merupakan suatu kejadian dimana $P(X) \neq 0$, maka peluang bersyarat X dinyatakan dalam persamaan berikut:

$$P(Y|X) = \frac{P(Y \cap X)}{P(X)} = \frac{P(Y)P(X|Y)}{P(X)} \quad (2.1)$$

$P(Y|X)$ merupakan peluang Y dengan syarat X , disebut posterior Y . Sedangkan $P(Y)$ disebut sebagai peluang prior Y . $P(X)$ merupakan marginal dari $P(Y \cap X)$ dan bernilai konstan. Untuk himpunan variabel random $x_1, x_2, x_3, \dots, x_n$ yang saling independen dengan parameter θ , maka generalisasi persamaan (2.1) adalah:

$$\begin{aligned} P(\theta|x_1, x_2, x_3, \dots, x_n) &= \frac{P(\theta)P(x_1, x_2, x_3, \dots, x_n|\theta)}{\int_{\Theta} P(\theta)P(x_1, x_2, x_3, \dots, x_n|\theta)d_{\theta}} \\ &\propto P(\theta)P(x_1, x_2, x_3, \dots, x_n|\theta). \end{aligned} \quad (2.2)$$

Secara umum, distribusi posterior adalah:

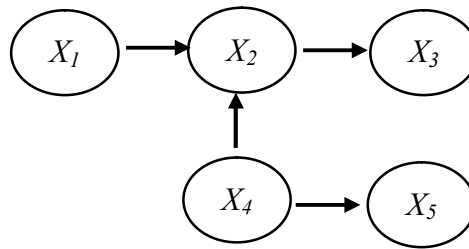
$$\begin{aligned} \text{Posterior} &= \frac{\text{likelihood.prior}}{\int \text{likelihood.prior} d_{\theta}} \\ &\propto \text{likelihood.prior} \end{aligned} \quad (2.3)$$

2.1.2 Struktur *Bayesian Network*

Cheng (2002), menyatakan bahwa *Bayesian Network* (BN) adalah sebuah DAG dengan tabel probabilitas untuk setiap *node* didalamnya. Setiap *node* dalam BN merepresentasikan variabel proposisional dalam sebuah domain dan *arc* yang ada *node* merepresentasikan hubungan ketergantungan antar variabel tersebut.

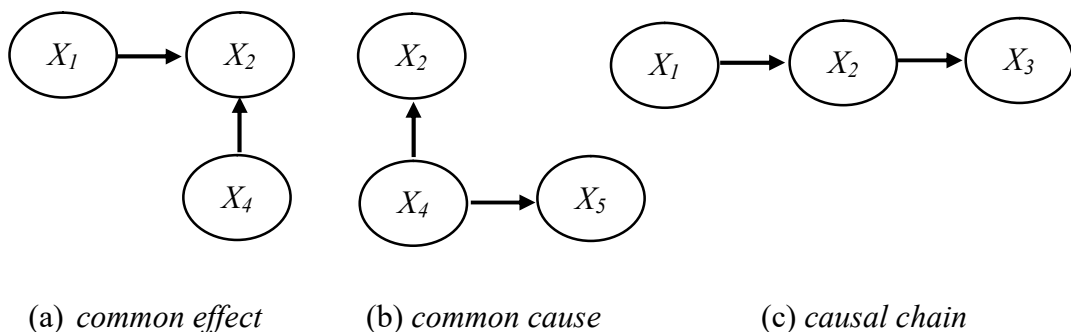
Pada Gambar 2.1, struktur BN yang terdiri atas *node* $V = \{X_1, X_2, \dots, X_5\}$ dan *edge* $E = \{e_1 = (X_1, X_2), e_2 = (X_2, X_3), e_3 = (X_4, X_2), e_4 = (X_4, X_5)\}$. Jika satu *edge* dari *node* X_i mengarah ke *node* X_j maka X_i disebut *parent*. X_j dinotasikan dengan $pa(X_j)$ dan X_j disebut *child* dari X_i . *Node* yang tidak memiliki *parent*

disebut *root node*. Sedangkan *node* yang tidak memiliki *child* disebut *leaf node* (Korb dan Nicholson, 2010). Berdasarkan Gambar 2.1 dapat dinyatakan bahwa X_2 adalah *parent* dari X_1 dan X_4 . Sebaliknya X_1 dan X_4 adalah *child* dari X_2 . Selanjutnya, X_1 dan X_4 merupakan *root node*, sedangkan X_3 dan X_5 adalah *leaf node*.



Gambar 2.1. Contoh Struktur DAG

Hubungan antar *node* terbagi atas tiga, yaitu *common cause*, *common effect* dan *causal chain*. Dari Gambar 2.1 dapat dijelaskan bahwa hubungan antar *node* X_1 , X_2 dan X_3 adalah *causal chain* dimana X_2 adalah *parent* dari X_1 sedangkan X_3 adalah *parent* dari X_2 . Jika suatu *node* menjadi *parent* lebih dari satu *node* lain maka disebut sebagai *common cause*, sedangkan jika suatu *node* menjadi *child* dari beberapa *node* (dua atau lebih) maka disebut sebagai *common effect*. Hubungan antar *node* dari Gambar 2.1 dapat diuraikan ke dalam Gambar 2.2.



Gambar 2.2. Contoh Hubungan antar Node

Salah satu kelebihan metode BN adalah mampu menjelaskan hubungan sebab akibat antar variabel. Sehingga BN dapat menjelaskan perubahan nilai variabel dalam struktur jika terdapat perubahan pada variabel lainnya.

2.1.3 Kebebasan Bersyarat

Konsep kebebasan bersyarat perlu diketahui agar dapat mengestimasi parameter model *Bayesian Network* (BN) dengan tepat. Misal X , Y dan Z merupakan variabel random, maka X dan Y bebas bersyarat, dengan syarat Z dan dituliskan dengan notasi $X \perp Y | Z$. Jika X dan Y memenuhi kebebasan dengan syarat Z , maka:

$$P(X|Y, Z) = P(X|Z). \quad (2.4)$$

Dengan demikian, diperoleh joint probability X, Y , dan Z sebagai berikut:

$$\begin{aligned} P(X, Y, Z) &= P(Z)P(Y|Z)P(X|Y, Z) \\ &= P(Z)P(Y|Z)P(X|Z). \end{aligned} \quad (2.5)$$

Dalam Gambar 2.2(a) apabila X_1 dan X_4 memenuhi kondisi bebas bersyarat diketahui X_2 , maka peluang bersama X_1 , X_2 dan X_4 sebagai berikut:

$$P(X_1, X_2, X_4) = P(X_1)P(X_4)P(X_2|X_1, X_4). \quad (2.6)$$

Struktur BN dalam Gambar 2.2(b) memenuhi kebebasan bersyarat jika X_4 diketahui. Oleh karena X_2 dan X_5 adalah saling bebas dengan syarat X_4 , maka:

$$P(X_2, X_5 | X_4) = P(X_2 | X_4)P(X_5 | X_4)$$

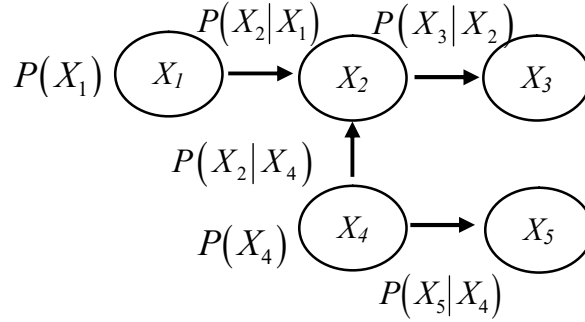
sehingga:

$$\begin{aligned} P(X_2, X_4, X_5) &= P(X_4)P(X_2, X_5 | X_4) \\ P(X_2, X_4, X_5) &= P(X_4)P(X_2 | X_4)P(X_5 | X_4). \end{aligned} \quad (2.7)$$

Apabila X_2 dalam Gambar 2.2(c) diketahui maka peluang X_1 dan X_3 memenuhi kebebasan bersyarat dimana peluang X_1 mempengaruhi peluang X_2 , sedangkan peluang X_2 mempengaruhi peluang X_3 sehingga peluang bersama X_1 , X_2 dan X_3 dapat dituliskan:

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2). \quad (2.8)$$

Kondisi distribusi peluang bersama dari Gambar 2.1 dapat diilustrasikan pada Gambar 2.3 berikut ini.



Gambar 2.3. Struktur DAG

Secara umum peluang bersama dalam BN yang menghubungkan antara *parent node(s)* dan *child node(s)* dalam struktur DAG yang memenuhi asumsi kebebasan bersyarat adalah sebagai berikut:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \quad (2.9)$$

dimana $i = 1, 2, \dots, n$ merupakan indeks untuk *node* ke-1 hingga *node* ke- n , sedangkan $P(X_i | Pa(X_i))$ merupakan distribusi peluang bersyarat *node* X_i dengan *parentnya*.

Pada beberapa kasus, struktur *Bayesian Network* (BN) tidak diketahui sehingga memerlukan eksplorasi dari data. Mempelajari struktur BN lebih sulit dibandingkan mempelajari parameter BN. Metode estimasi dari BN dipilih berdasarkan kondisi data dan struktur BN yang digunakan. Tabel 2.1 menunjukkan empat kasus untuk metode estimasi parameter BN berdasarkan kondisi data dan struktur BN.

2.1.4 Estimasi Parameter *Bayesian Network*

Apabila struktur BN dapat diketahui, maka langkah selanjutnya adalah melakukan estimasi parameter berdasarkan struktur yang telah terbentuk. Misal

kategori yang mungkin dari suatu kejadian dinotasikan k kategori, maka untuk $k = 2$ variabel diidentifikasi berdistribusi binomial dan apabila $k > 2$ diidentifikasi berdistribusi multinomial.

a. Distribusi Binomial

Misalkan X adalah variabel random berdistribusi Binomial, maka fungsi *pdf*-nya adalah:

$$P(X|\theta) = \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x}, \quad x=0,1,\dots,n \text{ dan } 0 \leq \theta \leq 1 \quad (2.10)$$

Estimasi parameter distribusi Binomial dapat dilakukan dengan *Maximum Likelihood Estimator* (MLE), dengan langkah-langkah sebagai berikut:

1. Membuat fungsi *likelihood*

Misalnya diambil m sampel random dari populasi yang berdistribusi Binomial (n, θ) , maka fungsi *likelihood*-nya adalah:

$$\begin{aligned} L(\theta|X) &= \prod_{i=1}^m P_i(X|\theta) \\ &= \prod_{i=1}^m \binom{n}{x_i} \theta^{x_i} (1-\theta)^{n-x_i} \\ &= \theta^{\sum_{i=1}^m x_i} (1-\theta)^{\sum_{i=1}^m (n-x_i)} \prod_{i=1}^m \binom{n}{x_i}. \end{aligned} \quad (2.11)$$

2. Membuat \ln fungsi *likelihood*:

$$\begin{aligned} \ln(L(\theta|X)) &= \ln \left(\theta^{\sum_{i=1}^m x_i} (1-\theta)^{\sum_{i=1}^m (n-x_i)} \prod_{i=1}^m \binom{n}{x_i} \right) \\ &= \ln \left(\theta^{\sum_{i=1}^m x_i} \right) + \ln \left((1-\theta)^{\sum_{i=1}^m (n-x_i)} \right) + \ln \left(\prod_{i=1}^m \binom{n}{x_i} \right) \\ &= \sum_{i=1}^m x_i \ln(\theta) + \sum_{i=1}^m (n-x_i) \ln(1-\theta) + \ln \left(\prod_{i=1}^m \binom{n}{x_i} \right) \end{aligned}$$

3. Menentukan nilai optimum θ dengan proses $\frac{\partial \ln(L(\theta|X))}{\partial \theta} = 0$:

$$\begin{aligned}
\frac{\partial \ln(L(\theta|X))}{\partial \theta} &= \frac{\sum_{i=1}^m x_i}{\theta} - \frac{\sum_{i=1}^m (n - x_i)}{(1-\theta)} + 0 = 0 \\
\frac{(1-\theta) \sum_{i=1}^m x_i}{\theta(1-\theta)} - \frac{\theta \sum_{i=1}^m (n - x_i)}{\theta(1-\theta)} &= 0 \\
\sum_{i=1}^m x_i - \theta \sum_{i=1}^m x_i - \theta \sum_{i=1}^m n + \theta \sum_{i=1}^m x_i &= 0 \\
\sum_{i=1}^m x_i - \theta \sum_{i=1}^m n &= 0 \\
\hat{\theta} &= \frac{\sum_{i=1}^m x_i}{\sum_{i=1}^m n} = \frac{X}{N}
\end{aligned}$$

dimana:

$\sum_{i=1}^m x_i = X$ menyatakan banyaknya kejadian sukses dari suatu sampel

$\sum_{i=1}^m n = N$ menyatakan total sampel

Estimator MLE untuk parameter θ dari distribusi Binomial (n, θ) dapat dituliskan sebagai berikut:

$$\hat{\theta}_{MLE} = \frac{X}{N} \quad (2.12)$$

b. Distribusi Multinomial

Misalkan X adalah variabel random berdistribusi Multinomial, maka fungsi *pdf*-nya adalah:

$$P(X|\theta_1, \theta_2, \dots, \theta_k) = \frac{n!}{x_1! x_2! \dots x_k!} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{1-x_1-x_2-\dots-x_{k-1}} \quad (2.13)$$

dimana $\sum_{l=1}^k x_g = n$ dan $\sum_{l=1}^k \theta_g = 1$

Estimasi parameter distribusi multinomial dapat dilakukan dengan *Maximum Likelihood Estimator* (MLE), dengan langkah sebagai berikut:

1. Membuat fungsi *likelihood*

Misalnya diambil m sampel random dari populasi yang berdistribusi

Multinomial $(n, \theta_1, \theta_2, \dots, \theta_h)$, maka fungsi *likelihood*-nya adalah:

$$\begin{aligned} L(\theta|X) &= \prod_{i=1}^m P_i(X|\theta_1, \theta_2, \dots, \theta_h) \\ &= \prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \theta_1^{x_{1i}} \theta_2^{x_{2i}} \dots \theta_h^{x_{hi}} \\ &= \theta_1^{\sum_{i=1}^m x_{1i}} \theta_2^{\sum_{i=1}^m x_{2i}} \dots \theta_h^{\sum_{i=1}^m x_{hi}} \prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \end{aligned} \quad (2.14)$$

2. Membuat \ln fungsi *likelihood*

$$\begin{aligned} \ln(L(\theta|X)) &= \ln \left(\theta_1^{\sum_{i=1}^m x_{1i}} \theta_2^{\sum_{i=1}^m x_{2i}} \dots \theta_h^{\sum_{i=1}^m x_{hi}} \prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \right) \\ &= \ln \left(\theta_1^{\sum_{i=1}^m x_{1i}} \right) + \ln \left(\theta_2^{\sum_{i=1}^m x_{2i}} \right) + \dots + \ln \left(\theta_h^{\sum_{i=1}^m x_{hi}} \right) + \ln \left(\prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \right) \\ &= \ln(\theta_1^{n_1}) + \ln(\theta_2^{n_2}) + \dots + \ln(\theta_h^{n_h}) + \ln \left(\prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \right) \\ &= n_1 \ln(\theta_1) + n_2 \ln(\theta_2) + \dots + n_h \ln(\theta_h) + \ln \left(\prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \right) \end{aligned}$$

Oleh karena $\sum_{l=1}^h \theta_l = \theta_1 + \theta_2 + \dots + \theta_h = 1$, maka dapat dituliskan

$$\theta_h = 1 - \theta_1 - \theta_2 - \dots - \theta_{h-1} = 1 - \sum_{l=1}^{h-1} \theta_l.$$

$$\ln(L(\theta|X)) = n_1 \ln(\theta_1) + n_2 \ln(\theta_2) + \dots + n_h \ln \left(1 - \sum_{g=1}^{h-1} \theta_g \right) + \ln \left(\prod_{i=1}^m \frac{n!}{x_{1i}! x_{2i}! \dots x_{hi}!} \right) \quad (2.15)$$

3. Menentukan nilai optimum θ dengan proses $\frac{\partial \ln(L(\theta|X))}{\partial \theta} = 0$:

$$\frac{\partial \ln(L(\theta|X))}{\partial \theta_1} = \frac{n_1}{\theta_1} + 0 + \dots + \frac{n_h}{1 - \sum_{g=1}^{h-1} \theta_g} + 0 = 0$$

$$\frac{\partial \ln(L(\theta|X))}{\partial \theta_2} = 0 + \frac{n_2}{\theta_2} + \dots + \frac{n_3}{1 - \sum_{g=1}^{h-1} \theta_g} + 0 = 0$$

$$\frac{\partial \ln(L(\theta|X))}{\partial \theta_{h-1}} = 0 + 0 + \dots + \frac{n_{h-1}}{\theta_{h-1}} + \frac{n_3}{1 - \sum_{g=1}^{h-1} \theta_g} + 0 = 0$$

Sehingga didapatkan $\frac{n_1}{\theta_1} = \frac{n_2}{\theta_2} = \dots = \frac{n_{g-1}}{\theta_{g-1}} = \frac{n_3}{1 - \sum_{g=1}^{h-1} \theta_g}$. Apabila rasio

perbandingan ini dinotasikan dengan simbol λ maka

$\lambda = \frac{n_1}{\theta_1} = \frac{n_2}{\theta_2} = \dots = \frac{n_{h-1}}{\theta_{h-1}} = \frac{n_3}{1 - \sum_{g=1}^{h-1} \theta_g}$. Jika $\sum_{g=1}^h \theta_g = 1$, maka diperoleh

$\sum_{g=1}^h \theta_g = \frac{n_1}{\lambda} + \frac{n_2}{\lambda} + \dots + \frac{n_h}{\lambda} = 1$. Apabila $\sum_{g=1}^h n_g = N$, maka nilai λ akan sama dengan

nilai N . Secara umum, estimasi parameter distribusi multinomial dapat dituliskan:

$$\hat{\theta}_g = \frac{n_g}{N} \quad (2.16)$$

dimana: n_g adalah banyaknya sampel untuk kategori

N adalah total sampel.

Oleh karena itu, estimasi parameter untuk peluang bersyarat adalah sebagai berikut:

$$\hat{\theta}_{g,l} = \frac{n_{g,l}}{n_l} \quad (2.17)$$

2.2 Uji Independensi

Model *Naïve Bayes* dibangun dengan asumsi seluruh atribut saling independen bersyarat variabel klasifikasi. Menurut Walpole (1995), uji independensi dilakukan untuk mengetahui ada atau tidaknya hubungan antara variabel dua variabel. Pengujian tersebut dilakukan dengan menggunakan uji *Chi-square* atau uji *Conditional Independence* (CI). Secara lebih jauh CI dapat

digunakan untuk melihat independensi dari dua variabel dengan mempertimbangkan keberadaan satu variabel lainnya.

2.2.1 Uji *Chi-Square*

Proses pengujian *Chi-square* dilakukan dengan membuat tabel kontingensi dua arah. Tabel kontingensi dua arah merupakan tabel kontingensi dengan melibatkan dua variabel kualitatif yang masing-masing mempunyai p kategori dan q kategori. Tabel 2.1 merupakan contoh tabel kontingensi dua arah dengan total frekuensi sebanyak n , dimana $j=1,2,3,\dots,p$ dan $j^*=1,2,3,\dots,q$. Namun perlu diketahui bahwa uji ini bekerja dengan baik bila data yang digunakan besar yaitu dapat memenuhi syarat maksimal 20% sel kontingensi yang memiliki nilai harapan kurang dari 5 (Agresti, 2002).

Tabel 2.1 Tabel Kontingensi Frekuensi Dua Arah pada Variabel X_j dan X_{j^*}

Kategori Variabel X_j	Kategori Variabel X_{j^*}						Total
	1	2	...	H	...	Q	
1	n_{11}	n_{12}	...	n_{1h}	...	n_{1q}	n_{1+}
2	n_{21}	n_{22}	...	n_{2h}	...	n_{2q}	n_{2+}
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
I	n_{i1}	n_{i2}	...	n_{ih}	...	n_{iq}	n_{i+}
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
P	n_{p1}	n_{p2}	...	n_{ph}	...	n_{pq}	n_{p+}
Total	n_{+1}	n_{+2}	...	n_{+h}	...	n_{+q}	n

Adapun estimasi probabilitas untuk tiap sel diperoleh dengan membagi nilai pada tiap sel dengan N , atau dapat ditulis sebagai berikut:

$$\hat{p}_{jj^*} = \frac{n_{jj^*}}{n}.$$

Dalam pengujian independensi antar dua variabel dapat digunakan uji *Pearson's Chi-square* (Dowdy, Weardon, dan Chilko, 2004). Hipotesis nol (H_0) untuk pengujian tersebut adalah variabel X_j independen terhadap variabel X_{j^*}

$(\rho_{jj^*} = \rho_{j+}\rho_{+j^*})$, dengan hipotesis alternatifnya (H_1) adalah variabel X_j tidak independen terhadap variabel X_{j^*} ($\rho_{jj^*} \neq \rho_{j+}\rho_{+j^*}$). Statistik uji yang digunakan adalah:

$$\chi^2 = \sum_{j^*=1}^q \sum_{j=1}^p \frac{(n_{jj^*} - e_{jj^*})^2}{e_{jj^*}} \quad (2.18)$$

dimana $e_{jj^*} = \frac{n_{j+}n_{+j^*}}{n}$. Berdasarkan nilai statistik uji maka keputusan akan tolak

H_0 jika $\chi^2 > \chi^2_{(p-1)(q-1); \alpha}$.

2.2.2 Uji *Conditional Independence*

Selain dengan uji *Chi-square*, untuk melihat hubungan independensi antara dua variabel digunakan uji *Conditional Independence* (CI). Hipotesis nol dalam uji CI adalah dua variabel atribut saling independen bersyarat variabel klasifikasi, sedangkan hipotesis alternatifnya yaitu dua variabel atribut tidak saling independen bersyarat variabel klasifikasi. Untuk data diskrit, statistik uji yang digunakan (Nagarajan, Scutari, dan Lebre, 2013):

$$CI(X, X^* | Y) = \sum_{j=1}^p \sum_{j^*=1}^q \sum_{l=1}^k \frac{n_{jj^*l}}{n} \log \frac{n_{jj^*l}n_{++l}}{n_{j+l}n_{+j^*l}}. \quad (2.19)$$

Nilai ini proportional dengan *log-likelihood ratio test* G^2 dimana dinyatakan tolak H_0 jika $p\text{-value} < \alpha = 0,05$. Selain itu bisa juga menggunakan statistik uji *Pearson's χ^2* untuk tabel kontingensi sebagai berikut:

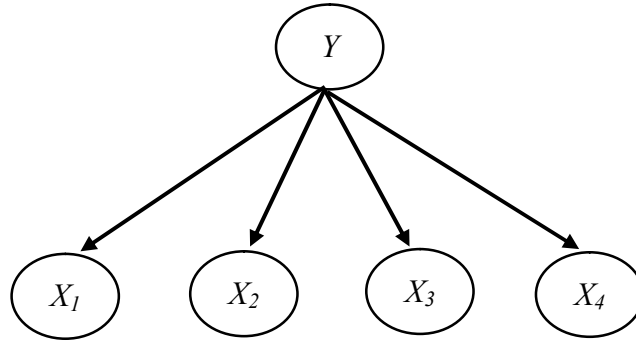
$$\chi^2(X, X^* | Y) = \sum_{j=1}^p \sum_{j^*=1}^q \sum_{l=1}^k \frac{(n_{jj^*l} - e_{jj^*l})^2}{e_{jj^*l}}. \quad (2.20)$$

dimana $e_{jj^*l} = \frac{n_{j+l}n_{+j^*l}}{n_{++l}}$. Berdasarkan nilai statistik uji maka keputusan akan tolak H_0

jika $\chi^2 > \chi^2_{(p-1)(q-1)k; \alpha}$.

2.3 Naïve Bayes (NB)

Naïve Bayes (NB) adalah bentuk klasifikasi *Bayesian Network* yang paling sederhana dimana variabel prediksi/ klasifikasinya diasumsikan bebas bersyarat (Bielza dan Larranaga, 2014) . Variabel klasifikasi (Y) menjadi *root node* dan variabel atribut (X) menjadi *leaf nodes*. Hubungan variabel yang disimbolkan dengan anak panah dihubungkan dari *root node* (Y) ke setiap *leaf nodes* (X). Asumsi yang mendasari pembentukan struktur jaringannya adalah independensi dari masing-masing kondisi/ kejadian. Gambar 2.4 memberikan contoh struktur NB pada variabel klasifikasi Y dan 4 (empat) variabel atribut $\mathbf{X}=\{X_1, X_2, X_3, X_4\}$.



Gambar 2.4 Contoh Struktur *Naïve Bayes*

Secara umum, struktur NB untuk p buah variabel atribut X dan variabel klasifikasi Y dapat diformulasikan sebagai berikut:

$$P(X_1, X_2, \dots, X_j, \dots, X_p, Y) = P(\mathbf{X}, Y) = P(Y) \prod_{j=1}^p P(X_j | Y) \quad j=1, 2, \dots, p \quad (2.21)$$

maka peluang Y pada saat diketahui atribut X adalah sebagai berikut (Friedman, dkk, 1997):

$$P(Y | X_1, X_2, \dots, X_j, \dots, X_p) = P(Y | \mathbf{X}) = \alpha \cdot P(Y) \prod_{j=1}^p P(X_j | Y) \quad j=1, 2, \dots, p \quad (2.22)$$

dimana α merupakan *normalized constant*.

Penentuan klasifikasi berdasarkan pada peluang terbesar $P(Y|\mathbf{X})$.

Misalnya pada saat sampel data \mathcal{X}_i diperoleh peluang posterior yang maksimum adalah peluang Y dengan kategori ke- k ($Y=y_k$), maka sampel data tersebut diklasifikasikan sebagai \mathcal{Y}_k . Formula untuk mendapatkan peluang kategori maksimum adalah sebagai berikut:

$$\arg \max P(Y=y_k|\mathbf{X}) = \arg \max P(Y) \prod_{j=1}^p P(X_j|Y) \quad j=1,2,\dots,p \quad (2.23)$$

Dari persamaan (2.23) dapat dihitung nilai maksimumnya apabila diketahui *conditional probability*. Oleh karena itu, diperlukan langkah-langkah estimasi parameter. Salah satu metode estimasi parameter adalah dengan MLE seperti pada persamaan (2.16) dan persamaan (2.17).

Langkah-langkah estimasi parameter NB dengan metode MLE adalah sebagai berikut:

1. Membentuk fungsi likelihood persamaan NB

$$L(\Theta|D) = \prod_{i=1}^n P(Y_i, X_{1i}, X_{2i}, \dots, X_{pi} : \Theta) \quad (2.24)$$

$$= \begin{bmatrix} P(Y_i : \theta_Y) \\ P(X_{1i}|Y : \theta_{X_1|Y}) \\ P(X_{2i}|Y : \theta_{X_2|Y}) \\ \vdots \\ P(X_{pi}|Y : \theta_{X_p|Y}) \end{bmatrix}$$

2. Membentuk ln fungsi likelihood dari persamaan (2.24)

$$\ln L(\Theta|D) = \ln \left(\prod_{i=1}^n P(Y_i, X_{1i}, X_{2i}, \dots, X_{pi} : \Theta) \right) \quad (2.25)$$

$$= \sum_{i=1}^n \begin{bmatrix} P(Y_i : \theta_Y) \\ P(X_{1i} | Y : \theta_{X_1|Y}) \\ P(X_{2i} | Y : \theta_{X_2|Y}) \\ \vdots \\ P(X_{pi} | Y : \theta_{X_p|Y}) \end{bmatrix}$$

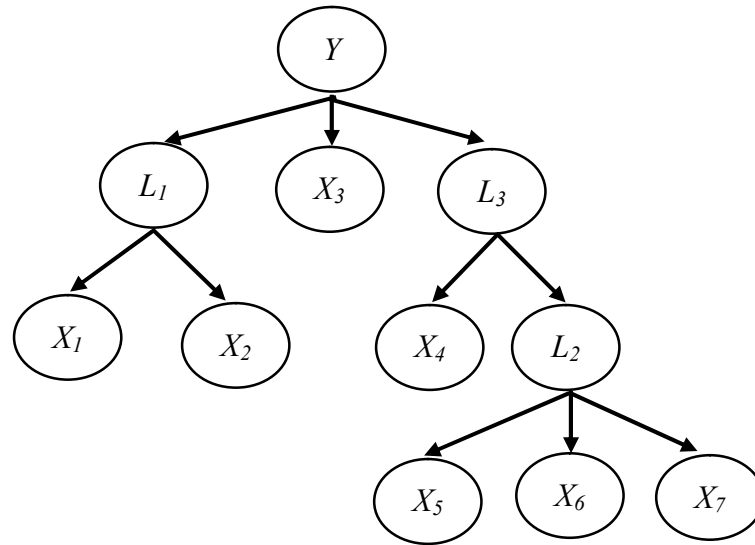
3. Mengestimasi persamaan (2.25) sehingga diperoleh hasil estimasi dalam persamaan (2.16) dan persamaan (2.17).
4. Mengestimasi persamaan NB menggunakan hasil estimasi masing-masing parameter dengan persamaan:

$$\hat{P}(Y = y_k | X) = \hat{P}(Y = y_k) \prod_{j=1}^p \hat{P}(X_j | Y = y_k) \quad j = 1, 2, \dots, p \quad (2.26)$$

2.4 Hierarchical Naïve Bayes (HNB)

Hierarchical Naïve Bayes Model (HNB) termasuk kelas spesial dari *Bayesian Network*. HNB merupakan bentuk *Bayesian Network* dimana variabel dibagi menjadi 3 (tiga) jenis variabel, yaitu variabel klasifikasi (Y), variabel atribut (X), dan variabel laten (L) (Langseth dan Nielsen, 2006). Variabel C merupakan *root node*, variabel X merupakan *leaves nodes* dan variabel L merupakan *internal nodes*. Model HNB dapat digambarkan secara ringkas dalam Gambar 2.5. Atribut X_5 , X_6 dan X_7 tidak independen bersyarat variabel klasifikasi (Y) sehingga dibentuk laten L_2 . Variabel laten L_2 merupakan variabel yang tidak teramati. Variabel laten L_1 terbentuk karena atribut X_1 dan X_2 tidak independen bersyarat variabel klasifikasi (Y). Begitu pula variabel laten L_3 terbentuk karena atribut X_4 dan L_2 tidak independen bersyarat variabel klasifikasi (Y). Perhatikan bahwa jika tidak ada variabel laten di HNB, maka akan menjadi model NB.

Model ini dirancang dengan mempertimbangkan adanya keterkaitan antar atribut. Atribut yang diduga kuat memiliki keterkaitan bersyarat selanjutnya akan dikombinasikan untuk menjadi variabel laten. Kategori pada variabel laten merupakan kombinasi kategori pada variabel pembentuknya seperti diilustrasikan pada Gambar 2.5.



Gambar 2.5 Contoh Struktur HNB (Langseth dan Nielsen, 2006)

Selanjutnya akan dilakukan *collapsing* atau penggabungan kategori pada variabel laten dengan suatu kriteria. Algoritma metode ini adalah sebagai berikut:

1. Membangun model Naïve Bayes tanpa variabel laten sebagai model awal (H_0)
2. Menentukan partisi data menjadi S subset $D^{(1)}, \dots, D^{(v)}$
3. Menentukan jumlah variabel laten. Jumlah variabel laten maksimal adalah sebesar $p-1$, dimana p adalah banyaknya variabel atribut
4. Untuk $r = 0, 1, 2, \dots, p-1$:
 - a. Untuk $s = 1, 2, \dots, v$
 - i. Mencari kandidat variabel laten $L^{(s)}$ pada subset data ke- s pada masing-masing subset data $D^{(s)}$
 - ii. Mencari jumlah kategori optimal untuk masing-masing $L^{(s)}$

- iii. Membentuk model $H^{(s)}$ yang melibatkan $L^{(s)}$ pada H_a
- b. Menghitung skor untuk seluruh model dengan rumus: Memilih $H^{(s)}$ yang paling baik

$$H' = \arg \max_{s=1, \dots, v} \text{score}(H^{(s)} | D_N)$$

Dimana D_N merupakan data training secara keseluruhan. Kemudian memilih H' dengan skor terbesar

- c. Jika $\text{score } H'$ lebih besar dibandingkan dengan skor model sebelumnya maka H' menjadi model baru. Lanjut ke pencarian variabel laten lainnya. Jika $\text{score } H'$ tidak lebih besar maka model sebelumnya adalah model final dan proses pencarian variabel laten berhenti.

Algoritma pencarian kandidat variabel laten (penjelasan 4.a.i):

1. Melakukan uji *conditional independence* pada seluruh atribut yang menjadi *child* variabel klasifikasi

$H_0 : X_j \text{ dan } X_{j^*} \text{ independent bersyarat } Y$

$H_1 : X_j \text{ dan } X_{j^*} \text{ tidak independent bersyarat } Y$

Menurut Whittaker (1990) dalam (Langseth dan Nielsen, 2006) jika X_j dan X_{j^*} independen bersyarat Y maka:

$$2N.I(X_j, X_{j^*} | Y) \xrightarrow{L} \chi^2_{(j-1)(j^*-1)} \quad (2.27)$$

dimana : $v = |sp(Y)|(|sp(X_j)| - 1)(|sp(X_{j^*})| - 1)$

$sp(Y)$ = banyaknya kategori pada variabel Y

$sp(X_j)$ = banyaknya kategori pada variabel Y .

Jika $i(X_j, X_{j^*})$ merupakan estimasi untuk $I(X_j, X_{j^*} | Y)$ maka:

$$Q(X_j, X_{j^*} | D_N) = P(Z \geq 2N.i(X_j, X_{j^*})) \quad (2.28)$$

dimana Z berdistribusi $\chi^2_{(j-1)(j^*-1)}$ dan $Q(X_j, X_{j^*} | D_N)$ memberikan p -value untuk H_0 .

2. Pilih dua variabel yang memiliki p -value terkecil dan dibawah nilai α yang ditentukan.
3. Buat variabel laten berdasarkan dua variabel pembentuknya. Dengan jumlah kategori (L) sebanyak:

$$|sp(L)| = \prod_{X \in ch(L)} |sp(X)|. \quad (2.29)$$

Algoritma penggabungan (*collapsing*) kategori (*state space*) pada variabel laten adalah sebagai berikut:

1. Menentukan kategori variabel laten sebanyak $|sp(X_j)|$ kali $|sp(X_{j^*})|$ sebagaimana pada persamaan (2.29) yaitu kombinasi unik satu-satu antar kategori X_j dan X_{j^*} .
2. Untuk setiap kategori ke- j (l_j) dan kategori ke- j^* ($l_{j^*} \in sp(L)$) hitung $\Delta_L(l_j, l_{j^*} | D_N)$
3. Pilih $l_j, l_{j^*} \in sp(L)$ dengan $\Delta_L(l_j, l_{j^*} | D_N)$ paling besar
4. Jika $\Delta_L(l_j, l_{j^*} | D_N) > 0$, maka kategori l_j, l_{j^*} digabung. Ulangi kembali ke langkah ke-2.

Formula $\Delta_L(l_j, l_{j^*} | D_N)$ adalah sebagai berikut ((Langseth dan Nielsen, 2006):

$$\begin{aligned} \Delta_L(l_j, l_{j^*}) \approx & \log_2 \frac{|sp(Y)|}{2} - \sum_{c \in sp(Y)} N(c, l_j) \log_2 \left(\frac{N(c, l_j)}{N(c, l_j) + N(c, l_{j^*})} \right) + \\ & - \sum_{c \in sp(Y)} N(c, l_{j^*}) \log_2 \left(\frac{N(c, l_{j^*})}{N(c, l_j) + N(c, l_{j^*})} \right) + \\ & N(l_j) \log \left(\frac{N(l_j)}{N(l_j) + N(l_{j^*})} \right) + N(l_{j^*}) \log \left(\frac{N(l_{j^*})}{N(l_j) + N(l_{j^*})} \right) \end{aligned} \quad (2.30)$$

dimana:

l_j = kategori ke- j

l_{j^*} = kategori ke- j^*

$|sp(Y)|$ = banyaknya kategori variabel Y (klasifikasi)

$N(c, l_j)$ = banyaknya kejadian (frekuensi) kategori ke- j pada saat $Y=c$

$N(c, l_{j^*})$ = banyaknya kejadian (frekuensi) kategori ke- j^* pada saat kategori $Y=c$

$N(l_j)$ = banyaknya kejadian (frekuensi) kategori ke- j

$N(l_{j^*})$ = banyaknya kejadian (frekuensi) kategori ke- j^*

\log = log basis bilangan natural - e

\log_2 = log basis 2 (dua)

Setelah didapatkan variabel laten maka struktur HNB untuk $p-r$ buah variabel atribut X dan r variabel laten serta variabel klasifikasi Y dapat diformulasikan sebagai berikut:

$$P(\mathbf{X}, \mathbf{L}, Y) = P(Y) \prod_{j=1}^{p-r} P(X_j | Y) \prod_{u=1}^r P(L_u | Y) \quad u = 1, 2, \dots, r \quad (2.31)$$

$$j = 1, 2, \dots, p - r$$

Klasifikasi model HNB hampir sama dengan model NB. Hanya saja pada model HNB melibatkan variabel laten sebagai ganti dari variabel atribut pembentuk laten. Estimasi parameter hampir sama dengan estimasi parameter model NB. Variabel atribut yang membentuk laten tidak dilibatkan secara langsung dalam model klasifikasi. Untuk mendapatkan variabel laten, maka data perlu ditransformasi terlebih dahulu berdasarkan struktur HNB yang terbentuk.

2.5 Asumsi Regresi Logistik

Regresi logistik tidak membutuhkan banyak asumsi seperti regresi linier umumnya yang mensyaratkan linieritas, normalitas, dan homoskedastisitas. Namun beberapa asumsi lainnya masih berlaku (Hosmer dan Lemeshow, 2000) dan (Lani, 2017):

Pertama, tidak memerlukan hubungan linier antara variabel dependen dan independen. Regresi logistik dapat menangani segala macam hubungan, karena menerapkan transformasi log non-linear ke rasio odds yang diprediksi. Kedua, variabel independen tidak perlu multivariat normal, walaupun normalitas multivariat menghasilkan solusi yang lebih stabil. Ketiga, homoskedastisitas tidak diperlukan. Regresi logistik tidak memerlukan varians untuk menjadi heteroskedastis untuk setiap tingkat variabel independen. Terakhir, dapat menangani data ordinal dan nominal sebagai variabel bebas. Variabel independen tidak perlu berskala metrik, dapat berskala interval atau rasio yang dikategorikan.

Kedua, penentuan model harus tepat. Tidak ada variabel penting yang diabaikan dan tidak ada variabel asing yang disertakan. Pendekatan yang baik untuk memastikan hal ini adalah dengan menggunakan metode reduksi variabel bertahap untuk memperkirakan regresi logistik.

Ketiga, regresi logistik mengharuskan setiap pengamatan independen. Hal ini dapat dilihat dari *error*-nya yang independen. Demikian pula dengan independensi variabel independen dimana harus independen satu sama lain sehingga tidak terjadi multikolinearitas. Multikolinearitas yang sempurna membuat estimasi menjadi tidak mungkin, sementara multikolinearitas yang kuat membuat perkiraan tidak tepat. Memasukkan efek interaksi maupun melakukan analisis faktor sebelum regresi logistik dilakukan dapat menjadi solusi dalam mengatasi multikolinearitas.

Keempat, regresi logistik mengasumsikan linearitas variabel independen dan log *odds*. Solusi untuk masalah ini adalah kategorisasi variabel independen yaitu mengubah variabel metrik ke tingkat nominal atau ordinal dan kemudian memasukkannya ke dalam model. Pendekatan lain adalah dengan menggunakan

analisis diskriminan, jika asumsi homoskedastisitas, normalitas multivariat, dan tidak adanya multikolinearitas terpenuhi.

2.6 Regresi Logistik

Analisis regresi menjelaskan bentuk hubungan antar variabel melalui suatu persamaan. Analisis regresi dapat mengukur seberapa besar suatu variabel mempengaruhi variabel lain. Analisis regresi juga dapat digunakan untuk melakukan peramalan nilai suatu variabel berdasarkan variabel lain. Dalam analisis regresi dikenal dua jenis variabel. Pertama, *dependent variable* /variabel respon (Y) merupakan variabel yang nilainya ditentukan oleh variabel lain. Variabel Y ini diasumsikan bersifat random. Kedua, *independent variable* /variabel penjelas (X) merupakan variabel yang nilainya ditentukan secara bebas (variabel yang diduga mempengaruhi variabel respon). Variabel X diasumsikan bersifat *fixed*. Variabel penjelas nilainya dapat ditentukan atau dapat diamati tetapi tidak dapat dikontrol (Draper dan Smith, 1998).

Dalam penelitian yang dihadapkan pada variabel respon Y , yang berdistribusi Binomial dimana hanya memiliki dua kemungkinan atau biner (dikotomi), seperti sukses atau gagal tidak dapat digunakan regresi linier berganda.

Misalnya, model regresi linier untuk y kontinu dan satu variabel penjelas

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

dengan ε_i merupakan residual dari masing-masing y_i dimana $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Residual ε_i mencerminkan karakteristik pengamatan ke $-i$ yang tidak teramati yang mempengaruhi respon y_i dan Diasumsikan ε_i untuk pengamatan berbeda saling independen. Dengan asumsi rata-rata ε_i nol sehingga β_0 adalah rata-rata y ketika $x=0$. Dengan asumsi ini, rata-rata atau nilai harapan dari nilai y pada masing-masing pengamatan adalah

$$E(y_i) = \beta_0 + \beta_1 x_i$$

Jika variabel respon $E(y_i) = \pi_i = P(y_i = 1)$ maka akan menjadi

$$\pi_i = \beta_0 + \beta_1 x_i$$

dan didapatkan model untuk y_i sebagai berikut:

$$y_i = \pi_i + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.32)$$

dengan $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Persamaan (2.32) tersebut merupakan model peluang linier (*linear probability model*) untuk y biner, dimana $E(y_i)$ merupakan peluang respon.

Residual pada persamaan (2.32) tersebut hanya dapat memiliki satu dari dua kemungkinan nilai untuk x_i , yaitu $1 - (\beta_0 + \beta_1 x_i)$ untuk $y_i = 1$ dan $-(\beta_0 + \beta_1 x_i)$ untuk $y_i = 0$. Hal ini melanggar asumsi regresi linier bahwa ε_i berdistribusi normal untuk suatu x_i . Selain itu, model regresi linier mengasumsikan bahwa ε_i memiliki varians konstan σ_ε^2 (asumsi homoskedastik). Namun, dari persamaan tersebut, varians y_i untuk x_i yaitu:

$$\text{var}(y_i) = \pi_i (-\pi_i) = (\beta_0 + \beta_1 x_i) [1 - (\beta_0 + \beta_1 x_i)]$$

yang berarti varians residual tergantung pada x_i sehingga asumsi homoskedastik terlanggar.

Oleh karena itu, pada umumnya dilakukan transformasi non-linier dari $\beta_0 + \beta_1 x_i$ yang memiliki jaminan bahwa prediksi peluang akan terletak nol dan satu. Bentuk umum dari model yang akan digunakan adalah

$$\pi_i = F(\beta_0 + \beta_1 x_i) \quad (2.33)$$

dengan F adalah fungsi transformasi yang memetakan peluang π antara 0 dan 1 untuk setiap nilai dalam rentang $(-\infty, \infty)$. Dimana F merupakan *cumulative distribution function* (cdf) dari suatu distribusi. Pilihan yang paling umum untuk distribusi ini adalah distribusi normal dan distribusi logistik karena bentuknya yang simetris. Meskipun untuk distribusi logistik, terdapat proporsi yang lebih besar pada ekor distribusinya dibandingkan dengan distribusi normal.

Model regresi dari suatu variabel respon yang berdistribusi Binomial adalah tidak linier. Untuk menjadikannya linier dapat digunakan *generalized*

linear model (GLM). GLM membuat suatu *link function* yang menghubungkan antara variabel prediktor dengan variabel respon dengan hubungan yang linier. Dari persamaan (2.33) dapat diperoleh GLM

$$F^{-1}(\pi_i) = \beta_0 + \beta_1 x_i \quad (2.34)$$

dengan F^{-1} adalah invers dari fungsi F . Fungsi invers F^{-1} diaplikasikan pada peluang respon π disebut *link function* karena menghubungkan π ke variabel penjelas.

Bentuk sederhana dari *link function* adalah *identity link* yang tidak mentransformasikan π sama sekali, yaitu $F^{-1} = \pi$ sehingga *identity link* akan membawa model ke *linear probability model*. Sifat penting dari kriteria *link* yang akan digunakan adalah bahwa $\pi_i = F(\beta_0 + \beta_1 x_i)$ akan selalu terletak antara 0 dan 1 berapapun nilai $\beta_0 + \beta_1 x_i$, sehingga dapat dipastikan bahwa model merupakan *linear probability model*. *Logit link* merupakan *link function* yang ditemukan oleh Joseph Berkson, seorang doktor medis dan statistisi. Berdasarkan transformasi logistik $z = \beta_0 + \beta_1 x$ yang didefinisikan sebagai

$$\pi = F(z) = \frac{\exp(z)}{1 + \exp(z)} = \frac{e^z}{1 + e^z} \quad (2.35)$$

dengan ‘exp’ dan ‘e’ keduanya menunjukkan fungsi eksponensial. $F(z)$ disebut transformasi logistik dari z dan merupakan cdf dari distribusi logistik. Karena $\exp(z)$ akan selalu lebih besar dari nol untuk berapapun nilai z maka nilai π pada persamaan (2.35) akan selalu terletak antara 0 dan 1. Jika z semakin besar dalam arah negatif ($z \rightarrow -\infty$) maka nilai π mendekati 0 dan jika z semakin besar dalam arah positif ($z \rightarrow \infty$) maka nilai π mendekati 1.

Persamaan (2.35) menyatakan peluang respon sebagai fungsi non-linier dari z . Untuk memperoleh GLM dari persamaan (2.35) pada sisi kanan dari persamaan (2.34), diperoleh:

$$\log\left(\frac{\pi}{1-\pi}\right) = z = \beta_0 + \beta_1 x \quad (2.36)$$

$\pi/(1-\pi)$ disebut sebagai *odds* bahwa $y = 1$ dan $\log(\pi/1-\pi)$ adalah *log-odds*, juga disebut sebagai *logit*. Model pada (2.36) disebut model logit atau model logistik.

Sifat menarik lain dari *link* logit adalah transformasi sederhana sehingga koefisien β memiliki interpretasi yang sederhana. Dengan melakukan *invers* dari log dari kedua sisi pada persamaan(2.36) diperoleh

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x) = \exp(\beta_0) \exp(\beta_1 x). \quad (2.37)$$

Untuk peningkatan satu unit x , dengan mengganti $x+1$ pada persamaan(2.37) maka:

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 (x+1)) = \exp(\beta_0) \exp(\beta_1 x) \exp(\beta_1). \quad (2.38)$$

Dengan membandingkan persamaan(2.37) dan persamaan(2.38), terlihat bahwa peningkatan 1 unit dalam x membuat *odds* $y = 1$ dikalikan dengan $\exp(\beta_1)$. Sebagai alternatif, dapat dikatakan bahwa peningkatan 1 unit dalam x meningkatkan *odds* sebesar $\exp(\beta_1)$. Dapat juga diinterpretasikan bahwa $\exp(\beta_1)$ adalah *odds ratio* karena persamaan (2.38) dibagi dengan persamaan (2.37) merupakan rasio dari *odds* $y = 1$ untuk dua individu dengan nilai x yang berjarak 1 unit.

Model regresi logistik ada yang bersifat *dikotomous* (biner) yang mensyaratkan variabel respon terdiri dari dua kategori, dan bersifat *polytomous* dengan variabel respon lebih dari dua kategori (Hosmer dan Lemeshow, 2000). Model regresi logistik yang diperoleh adalah seperti persamaan (2.36) dengan mengganti z dengan $g(x)$ sehingga diperoleh:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2.39)$$

dengan $g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$.

Secara umum, bentuk dari fungsi logit dengan variabel respon yang terdiri dari l kategori kategori adalah sebagai berikut:

$$g_l(x) = \beta_{l0} + \beta_{l1} x_1 + \beta_{l2} x_2 + \dots + \beta_{lp} x_p \quad (2.40)$$

Cumulative logit models didapatkan dengan membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori respon ke- l pada p variabel prediktor yang dinyatakan dalam $P(Y \leq l | x_i)$, dengan peluang lebih besar dari kategori respon ke- j , $P(Y > l | x_i)$ (Hosmer dan Lemeshow, 2000). Berikut rumus dari *cumulative logit models*.

$$\text{Logit } P(Y \leq l | x_i) = \log \left(\frac{P(Y \leq l | x_i)}{P(Y > l | x_i)} \right). \quad (2.41)$$

Misalkan terdapat k kategori dari variabel respon (Y) yang diberikan kode $l = 1, 2, \dots, k$. Suatu variabel respon dengan k kategori akan membentuk $k-1$ persamaan logit, dimana masing-masing persamaan ini membentuk fungsi logit yang membandingkan suatu kelompok kategori terhadap pembanding. Dalam model k kategori, terdapat $k-1$ fungsi logit yang terbentuk. Dalam teori dapat digunakan beberapa pasangan perbandingan logit dari respon, misalkan kelompok berkode $Y=0$ akan dijadikan sebagai nilai respon referensi (acuan). Misal X adalah vektor kovariat dari panjang p dengan kategori Y yaitu $l = 1, 2, \dots, k$. Akan didefinisikan dua fungsi logit sebagai berikut (Hosmer dan Lemeshow, 2000):

$$g_l(x) = \ln \left[\frac{P(Y = l | X)}{P(Y = 0 | X)} \right] \quad (2.42)$$

$$g_l(x) = \beta_{l0} + \beta_{l1} X_1 + \beta_{l2} X_2 + \dots + \beta_{lp} X_p \quad (2.43)$$

$$g_l(x) = \mathbf{x}^T \boldsymbol{\beta}_l \quad (2.44)$$

Model Regresi Logistik *Polytomous* logit dapat dipandang sebagai suatu perluasan dari model logit yang biner. Secara umum probabilitas bersyarat untuk model *polytomous* adalah :

$$\pi_l(x) = P\langle Y = l | x \rangle = \frac{\exp(g_l(x))}{1 + \sum_{l=1}^k \exp(g_l(x))} \quad (2.45)$$

dimana :

$\pi_l(x)$: peluang terjadinya peristiwa ke- l

2.6.1 Estimasi Parameter Regresi Logistik *Polytomous*

Metode *Maximum Likelihood Estimation* (MLE) merupakan metode yang digunakan untuk menaksir parameter-parameter model regresi logistik yang memberikan nilai estimasi β dengan memaksimumkan fungsi *likelihood* untuk sampel dengan n sampel random. Langkah-langkah metode estimasi *Newton Raphson* untuk k kategori $l = 1, 2, \dots, k$ adalah sebagai berikut:

1. Membuat fungsi *likelihood*

$$L(\beta) = \prod_{i=1}^n \left[\pi_0(x_i)^{y_0} \pi_1(x_i)^{y_1} \pi_2(x_i)^{y_2} \dots \pi_k(x_i)^{y_k} \right] \quad (2.46)$$

2. Membuat *ln-likelihood*

$$\begin{aligned} \ln L(\beta) &= \ln \prod_{i=1}^n \left[\pi_0(x_i)^{y_0} \pi_1(x_i)^{y_1} \pi_2(x_i)^{y_2} \dots \pi_k(x_i)^{y_k} \right] \\ &= \ln \left[\prod_{i=1}^n \left[\left(\frac{\exp(g_0(x_i))}{1 + \sum_{l=1}^k \exp(g_l(x_i))} \right)^{y_0} \left(\frac{\exp(g_1(x_i))}{1 + \sum_{l=1}^k \exp(g_l(x_i))} \right)^{y_1} \dots \left(\frac{\exp(g_k(x_i))}{1 + \sum_{l=1}^k \exp(g_l(x_i))} \right)^{y_k} \right] \right] \\ &= \sum_{i=1}^n \left[y_0 \left((g_0(x_i)) - \ln(1 + \sum_{l=1}^k \exp(g_l(x_i))) \right) + y_1 \left((g_1(x_i)) - \ln(1 + \sum_{l=1}^k \exp(g_l(x_i))) \right) \dots + \right. \\ &\quad \left. + y_k \left((g_k(x_i)) - \ln(1 + \sum_{l=1}^k \exp(g_l(x_i))) \right) \right] \\ &= \sum_{l=1}^k \sum_{i=1}^n y_l (g_l(x_i)) - k \sum_{i=1}^n \ln(1 + \exp \sum_{l=1}^k g_l(x_i)) \end{aligned}$$

3. Fungsi *ln-likelihood* diturunkan terhadap β

Apabila menghasilkan bentuk yang tidak *close form*, maka estimasi parameter β diperoleh melalui prosedur iterasi dengan iterasi *Newton Raphson*. iterasi *Newton Raphson* diperoleh dari pendekatan deret *Taylor* sebagai berikut.

$$\frac{\partial \ln L(\beta)}{\partial \beta} + (\beta^{(1)} - \beta) \frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta^T} + \frac{1}{2} (\beta^{(1)} - \beta)(\beta^{(1)} - \beta)^T \frac{\partial^3 \ln L(\beta)}{\partial \beta^T \partial \beta \partial \beta^T} + \dots = 0$$

Matriks β merupakan nilai awal yang ditentukan. Apabila $|(\beta^{(1)} - \beta)|$ diasumsikan sangat kecil maka suku ketiga dan seterusnya dapat diabaikan. Dengan demikian ekspansi deret *Taylor* dapat ditulis sebagai berikut.

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial \beta} + (\beta^{(1)} - \beta) \frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta^T} &= 0 \\ \beta^{(1)} &= \beta - \left(\frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta^T} \right)^{-1} \frac{\partial \ln L(\beta)}{\partial \beta} \end{aligned}$$

Secara umum iterasi ke-t *Newton Raphson* adalah sebagai berikut:

$$\beta^{(t)} = \beta^{(t-1)} - \left(\frac{\partial^2 \ln L(\beta)}{\partial \beta^{(t-1)} \partial \beta^{(t-1)T}} \right)^{-1} \frac{\partial \ln L(\beta)}{\partial \beta^{(t-1)}}$$

Proses iterasi akan terhenti pada saat $\|\beta^{(t)} - \beta^{(t-1)}\| \leq \varepsilon$.

2.6.2 Uji Signifikansi Parameter

Pengujian signifikansi model dan parameter merupakan pemeriksaan untuk menentukan apakah variabel penjelas/prediktor dalam model signifikan atau berpengaruh secara nyata terhadap variabel respon. Seperti yang ada dalam regresi logistik, maka dalam Regresi Logistik *Polytomous* dapat digunakan uji *likelihood ratio* (LR) untuk menilai signifikansi dari variabel penjelas.

a. Uji Parameter Secara Keseluruhan

Dalam Hosmer dan Lemeshow (2000), pengujian parameter secara keseluruhan atau uji rasio *likelihood* diperoleh dengan cara membandingkan

fungsi log likelihood tanpa variabel bebas dengan fungsi log likelihood dari seluruh variabel bebas. Hipotesis nol: $\beta_{jl} = 0$ (tidak ada pengaruh antara variabel penjelas terhadap variabel respon) melawan hipotesis alternatif: minimal ada satu $\beta_{jl} \neq 0$ atau tidak semua $\beta_{jl} = 0$ (minimal ada satu variabel penjelas yang berpengaruh terhadap variabel respon) untuk $j=1,2,\dots,p$ dan $l=1,2,\dots,k$. Statistik uji yang digunakan adalah :

$$G = -2 \log \left[\frac{L_0}{L_1} \right]$$

Statistik G ini mengikuti distribusi *chi-square* dengan derajat bebas p . H_0 ditolak jika $G \geq \chi^2(p, \alpha)$ atau H_0 ditolak jika $p\text{-value} \leq \alpha$

b. Uji Parameter Secara Parsial

Uji parameter secara individu diperoleh dengan cara mengkuadratkan rasio estimasi parameter dengan estimasi standar errornya (Hosmer dan Lemeshow, 2000). Hipotesis nol (H_0): $\beta_{jl} = 0$ ((tidak ada pengaruh antara variabel penjelas ke- i terhadap variabel respon) dengan hipotesis alternatif (H_1): $\beta_{jl} \neq 0$ (ada pengaruh antara variabel penjelas ke- i terhadap variabel respon) untuk $j=1,2,\dots,p$ dan $l=1,2,\dots,k$.

Dalam penelitian ini digunakan adalah uji Wald yang berfungsi menguji signifikansi tiap parameter yang dirumuskan sebagai berikut:

$$W = \frac{\hat{\beta}_{jl}}{Se(\hat{\beta}_{jl})} \sim N(0,1)$$

dimana : $\hat{\beta}_{jl}$ merupakan penduga untuk parameter (β) dan $Se(\hat{\beta}_{jl})$: penduga galat baku untuk koefisien (β). Uji *Wald* diasumsikan mengikuti distribusi normal. H_0 ditolak jika $W > z_{1-\frac{\alpha}{2}}$ atau $W \leq -z_{1-\frac{\alpha}{2}}$ atau jika $p\text{-value} < \alpha$.

2.6.3 Seleksi Variabel

Metode statistik yang biasanya digunakan untuk menseleksi variabel adalah

metode *forward selection*, *backward elimination*, dan *stepwise*. Metode *forward selection* dilakukan dengan cara memasukkan variabel independen secara bertahap berdasarkan korelasi parsial terbesar. Dalam metode *forward selection*, variabel independen yang dimasukkan dalam model tidak akan dapat dikeluarkan lagi. Proses tersebut dihentikan ketika variabel-variabel independen baru tidak bisa meningkatkan berpengaruh secara signifikan terhadap variabel dependen. Metode *backward elimination* dilakukan dengan cara memasukkan semua variabel independen kemudian mengeliminasi satu persatu hingga tersisa variabel independen yang signifikan saja.

Metode *stepwise* merupakan kombinasi dari metode *forward selection* dan metode *backward elimination*. Pada tiap tahapnya dimungkinkan sebuah variabel independen dimasukkan atau dikeluarkan dari model. Variabel independen yang pertama kali masuk adalah variabel yang korelasinya tertinggi dan signifikan terhadap variabel dependen. Variabel independen kedua yang masuk adalah variabel yang korelasi parsialnya tertinggi dan masih signifikan. Setelah sejumlah variabel independen tertentu masuk kedalam model maka variabel lain yang ada didalam model dievaluasi, jika ada variabel yang tidak signifikan maka variabel tersebut dikeluarkan.

2.6.4 Akaike Information Criterion (AIC)

Yang dimaksud dengan model terbaik dalam hal ini menggunakan *principal of parsimoni*, yaitu dengan model yang sesederhana mungkin mampu menjelaskan keragaman ketahanan hidup sebanyak mungkin.

AIC merupakan salah satu ukuran yang sering digunakan dalam membandingkan *maximum likelihood model*. AIC didefinisikan sebagai:

$$AIC = -2 \log \text{likelihood} + 2p \quad (2.47)$$

dimana, p adalah jumlah parameter dalam model. AIC dapat dianggap sebagai suatu ukuran yang mengkombinasikan antara kecocokan dan kerumitan model. Kecocokan model ini diukur dengan $-2 \log \text{likelihood}$. Semakin besar nilainya, maka tingkat kecocokan model semakin rendah. Sementara itu, tingkat kerumitan diukur dengan $2p$. Semakin besar nilainya, maka model

menjadi semakin rumit sehingga interpretasinya menjadi lebih sulit. Model dengan nilai *information criterion* yang lebih kecil dianggap lebih baik.

2.7 Correct Clasification Rate

Correct Clasification Rate(CCR) merupakan perbandingan hasil klasifikasi (prediksi) dengan klasifikasi aktual (sebenarnya). CCR dapat digunakan untuk menjelaskan ketepatan klasifikasi suatu metode. Setiap baris dalam tabel tersebut merepresentasikan klasifikasi aktual sedangkan setiap kolomnya merupakan klasifikasi hasil prediksi suatu model atau sebaliknya.

$$CCR = \frac{\sum_{l^*} (OC)_{ll^*}}{\sum_{l=1}^k \sum_{l^*=1}^{k^*} (OC)_{ll^*}} . 100\% \quad (2.48)$$

dengan: $(OC)_{ll^*}$ merupakan frekuensi bersama data aktual kategori ke- l , data prediksi kategori ke- l^* , k adalah jumlah kategori data aktual dan k^* adalah jumlah kategori data prediksi.

2.8 Area Under The Curve (AUC)

Area under the curve atau *area under(a ROC) curve* (AUC) adalah ukuran keakuratan uji diagnostik kuantitatif. AUC merupakan plot dari tes sensitivitas yang di plot pada y axis dibandingkan dengan (1 - spesifisitas) yang di plot pada x axis (Zhou, Obuchowski, dan McClish, 2002).

Receiver Operating Characteristic (ROC) adalah hasil pengukuran klasifikasi dalam bentuk 2-dimensi. Berikut ada empat peluang yang dapat diformulasikan dalam tabel kontingensi 2 x 2 untuk menganalisis ROC :

Tabel 2.2 Tabel Kontingensi ROC

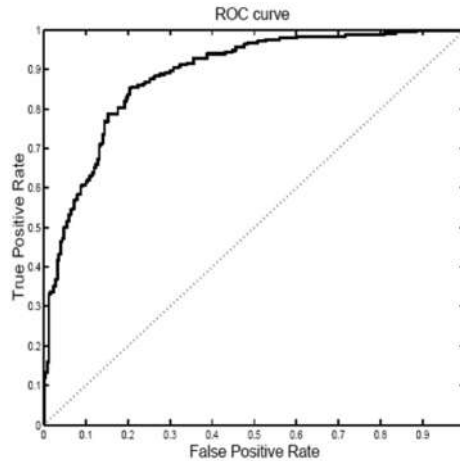
ROC		Aktual	
		Benar	Salah
(1)	(2)	(3)	(4)
Prediksi	Positif	Benar Positif (BP)	Salah Positif (SP)
	Negatif	Benar Negatif (BN)	Salah Negatif (SN)

Selanjutnya, dari Tabel 2.2 dapat dihitung nilai dari:

True Positive Rate disebut juga Sensitivity (TPR)=BP/(BP+SN)

True Negative Rate disebut juga Specifity (TNR)= BN/(BN+SP)

Area Under The Curve (AUC) adalah luas daerah di bawah kurva ROC seperti pada Gambar 2.6.



Gambar 2.6 Contoh Kurva ROC

AUC sama dengan probabilitas bahwa $f(x)$ memberikan nilai lebih tinggi pada input positif yang ditarik secara acak x^+ daripada input negatif yang ditarik secara acak x^- (Tian, Shi, Chen dan Chen, 2011). Apabila nilai AUC sama dengan 1, maka *classifier* mencapai akurasi sempurna. AUC dapat diestimasi dengan():

$$AUC(f) = P\left(f(x^+) > f(x^-)\right)$$

$$AUC = \frac{\sum_{l=1}^{n^+} \sum_{l=1}^{n^-} 1_{(f(x^+) > f(x^-))}}{n^+ n^-} \quad (2.49)$$

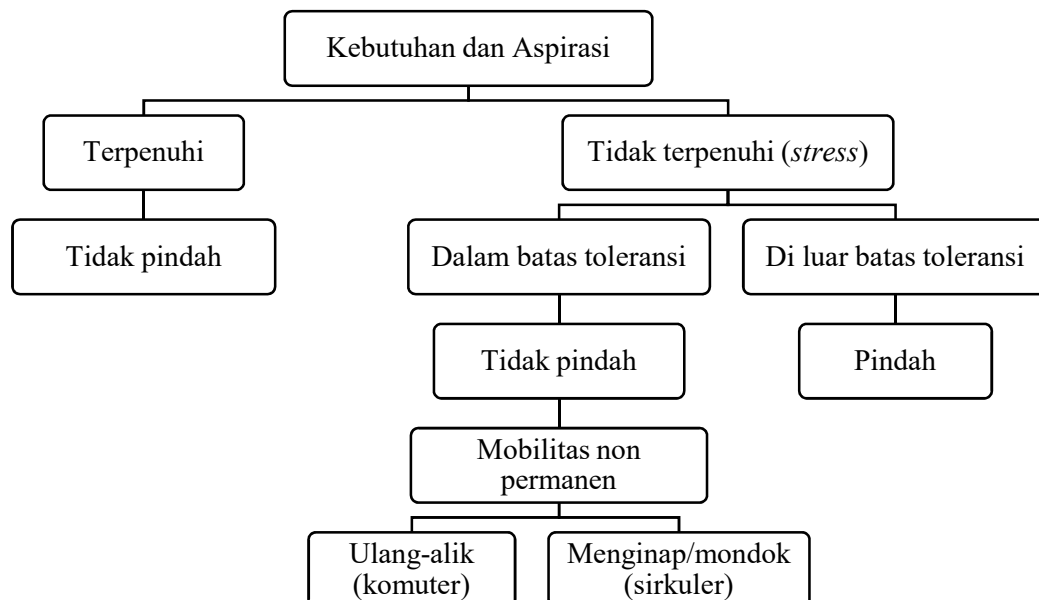
dengan: $f(.)$ merupakan nilai suatu fungsi yang diasumsikan berdistribusi normal, x^+ dan x^- merupakan sampel positif dan sampel negatif, sedangkan n^+ dan n^- merupakan jumlah sampel positif dan sampel negatif.

Dalam penelitian ini digunakan *package* (rms) dari perangkat lunak R versi 4.3.2. Dalam *package* ini jika prediktor memiliki lebih dari 2 kategori, maka variabel indikator dibentuk dari kategori yang paling sering dan sisanya digabung menjadi kategori lainnya.

2.9 Migrasi Penduduk

Migrasi merupakan salah satu fenomena dasar yang ada sejak awal peradaban manusia dan akan selalu ada yang disebabkan adanya sumber kesempatan dan pandangan yang baru bagi para migran serta adanya penderitaan dan kebutuhan kemanusiaan. Menurut Mantra (2000), keputusan untuk melakukan migrasi/ mobilitas penduduk dapat dijelaskan melalui teori kebutuhan dan tekanan (*need and stress*) pada Gambar 2.7.

Kebutuhan setiap individu berupa kebutuhan ekonomi, sosial, politik maupun psikologi tidak selamanya terpenuhi. Tidak terpenuhinya kebutuhan tersebut akan menimbulkan tekanan atau stres. Tinggi rendahnya stres yang dialami oleh masing-masing individu berbanding terbalik dengan proporsi pemenuhan kebutuhan tersebut. Apabila stres masih dalam batas toleransi maka individu tersebut tidak akan pindah dan tetap menetap di daerah asal dan berusaha menyesuaikan kebutuhannya dengan lingkungan yang ada. Namun apabila stres yang dirasakan oleh seorang individu di luar batas toleransinya maka individu tersebut akan mempertimbangkan untuk pindah ke tempat dimana dia merasa kebutuhan-kebutuhan yang diperlukannya dapat terpenuhi dengan baik.



Gambar 2.7 Hubungan antara Kebutuhan dan Pola Migrasi Penduduk (Mantra, 2000)

Menurut Lee (1966) setiap tindakan migrasi selalu melibatkan daerah asal, tujuan, dan rintangan. Terdapat empat faktor yang mempengaruhi migrasi yaitu:

1. Faktor yang terdapat di daerah asal,
2. Faktor yang terdapat di daerah tujuan,
3. Rintangan-rintangan yang menghambat: biaya, jarak, undang-undang migrasi
4. Faktor-faktor pribadi: kepekaan pribadi dan kecerdasan

Widaryatmo (2009) dalam penelitiannya terhadap pekerja di Indonesia menjelaskan bahwa variabel individu yaitu umur, jenis kelamin, tingkat pendidikan, status perkawinan, status pekerjaan, sektor pekerjaan, tempat tinggal dan variabel kontekstual seperti pertumbuhan PDRB, share sektor industri terhadap total PDRB mempengaruhi keputusan pekerja Indonesia untuk melakukan mobilitas non permanen.

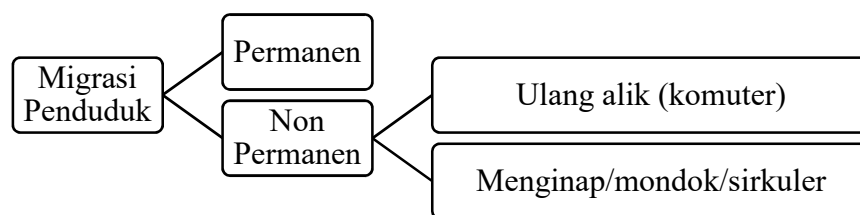
Pada penelitian yang dilakukan oleh Guntoro (2016), beberapa faktor yang signifikan mempengaruhi penduduk melakukan migrasi internal di Indonesia yaitu:

- a. Faktor karakteristik individu yaitu jenis kelamin, umur dan status perkawinan
- b. Faktor karakteristik rumah tangga yang terdiri atas variabel area tempat tinggal dan jumlah anggota rumah tangga
- c. Faktor status sosial ekonomi yaitu pendidikan responden, pendidikan tinggi ibu, dan kepemilikan rumah.

Berdasarkan hasil penelitian Sanjaya (2012), variabel-variabel yang mempengaruhi keputusan migrasi di Indonesia yaitu: umur anggota rumah tangga, pendidikan tertinggi yang ditamatkan, suku bangsa, status perkawinan, keluarga dengan anak usia 5-12 tahun, keluarga dengan anak usia 13-17 tahun, kepala keluarga perempuan dengan anak usia di bawah 18 tahun, suami istri bekerja, status kepemilikan rumah, persentase penduduk yang pindah tahun 2000-2005, persentase suku jawa tahun 2005 dan persentase luas lahan hijau.

2.10 Migrasi Ulang Alik (Komuter)

Migrasi penduduk Indonesia terbagi mejadi dua yaitu migrasi permanen dan migrasi non permanen. Migrasi permanen dapat dikatakan bertujuan pindah tempat tinggal secara tetap dan migrasi non permanen tidak mempunyai tujuan pindah tetapi memiliki tempat aktivitas yang berbeda dengan tempat tinggalnya. Migrasi non permanen sendiri dapat dibedakan menjadi dua kategori berdasarkan aspek rutinitas yaitu sirkuler (migrasi sirkulasi) dan komuter (migrasi ulang alik) sehingga dapat digambarkan pada Gambar 2.8.



Gambar 2.8. Migrasi Non Permanen (Mantra, 2000)

Dengan melihat daerah tempat tinggalnya, Hugo (1981) dalam Sanjaya (2012) membedakan komuter menjadi dua. Pertama *autochthonous commuters* adalah penduduk asli luar kota yang melakukan komutasi ke kota untuk bekerja. Hal ini disebabkan oleh faktor ekonomi dan faktor budaya. Mereka membutuhkan kota untuk mencari nafkah, tetapi mereka enggan tinggal jauh dari lingkungan keluarga dan sanak famili. Dengan adanya transportasi yang makin lancar mereka cenderung untuk tetap tinggal menetap di desa dan tiap hari pergi ke daerah tujuan untuk bekerja. Kedua *allochthonous commuters* adalah penduduk yang sebelumnya tinggal di kota tetapi dikarena alasan-alasan tertentu pindah dan menetap di luar kota, namun tetap bekerja dengan jalan melakukan komuter. *Allochthonous commuters* merupakan bagian dari proses sub-urbanisasi dimana terjadi perpindahan penduduk dari pusat kota ke pinggiran atau luar kota dengan tujuan untuk mendapatkan perumahan yang cukup baik dengan biaya yang relatif murah dan lingkungan yang bersih dan sehat.

Badan Pusat Statistik (2014) menjelaskan bahwa pelaku migrasi ulang alik/nglaju atau biasa disebut dengan komuter adalah seseorang yang melakukan suatu kegiatan bekerja/sekolah/kursus di luar kabupaten/kota dan secara rutin

pergi dan pulang pada hari yang sama. Di luar kabupaten/kota maksudnya adalah kegiatan tersebut berlokasi di kabupaten/kota yang berbeda dengan tempat tinggal komuter sekarang. Komuter tidak mempunyai tempat tinggal di tempat tujuan. Sementara yang dimaksud dengan rutin tidak harus selalu setiap hari melakukan kegiatan di luar kabupaten/kota, asalkan kegiatan tersebut sudah menjadi kebiasaan yang masih berlangsung.

Dalam penelitiannya, Sahara (2010) menyimpulkan bahwa jenis kelamin, status perkawinan, tingkat pendidikan, kelompok umur, pendapatan, jenis pekerjaan, alat transportasi, dan jarak signifikan mempengaruhi waktu tempuh yang digunakan untuk perjalanan pekerja migrasi ulang-alik. Variabel jenis kelamin dan status perkawinan juga berpengaruh secara signifikan terhadap probabilitas keputusan tenaga kerja menjadi komuter dalam penelitian Kusumaningrum (2014).

2.11 Moda Transportasi

Moda transportasi menjadi masalah tersendiri bagi para komuter. Ma (2015) dalam penelitiannya menyimpulkan bahwa waktu tempuh, jarak tempuh, mudah/sulit untuk menemukan tempat parkir, tujuan utama komuter, jumlah mobil dalam rumah tangga, pendapatan rumah tangga, jumlah anak, status pernikahan dan jenis pekerjaan (pekerjaan penuh waktu atau paruh waktu) memiliki pengaruh signifikan terhadap probabilitas pilihan moda transportasi komuter.

Menurut Tamin (2000), faktor yang mempengaruhi pemilihan moda dapat dikelompokkan menjadi tiga, yaitu:

- a. Ciri pengguna jalan, yang meliputi:
 - 1) Ketersediaan atau kepemilikan kendaraan pribadi
 - 2) Kepemilikan Surat Izin Mengemudi (SIM)
 - 3) Struktur rumah tangga
 - 4) Pendapatan
 - 5) Faktor lain misal keharusan menggunakan mobil ke tempat bekerja dan keperluan mengantar anak sekolah.

- b. Ciri pergerakan, yang meliputi:
 - 1) Tujuan pergerakan
 - 2) Waktu terjadinya pergerakan
 - 3) Jarak perjalanan.
- c. Ciri fasilitas moda transportasi, yang meliputi:
 - Faktor kuantitatif:
 - 1) Waktu perjalanan
 - 2) Biaya transportasi
 - 3) Ketersediaan ruang dan tarif parkir
 - Faktor kualitatif:
 - 1) Kenyamanan dan keamanan
 - 2) Keandalan dan keteraturan
- c. Ciri kota atau zona, yang meliputi:
 - Jarak dari pusat kota dan kepadatan penduduk

Penggunaan model pemilihan moda akan menghasilkan besarnya pergerakan setiap moda. Ciri pribadi yang berbeda-beda digunakan untuk memperkirakan pemilihan moda. Pemodelan pemilihan moda sangat ditentukan oleh persepsi seseorang dalam membandingkan biaya perjalanan ataupun waktu tempuh dalam memilih moda yang akan digunakannya.

Pemilihan moda transportasi dipengaruhi oleh faktor yang berbeda-beda (Ma, 2015):

- Karakteristik perjalanan: tujuan perjalanan, rantai perjalanan, kondisi cuaca, waktu sampai, waktu perjalanan, jarak perjalanan, biaya perjalanan, dll.
- Karakteristik sosio-demografi: Surat Ijin Mengemudi (SIM), ketersediaan mobil, kepemilikan tiket musiman, anak yang dimiliki, jenis kelamin, umur, pendapatan rumah tangga, komposisi rumah tangga, pekerjaan, dll
- Karakteristik spasial: kepadatan penduduk dari wilayah tempat tinggal, kedekatan dengan infrastruktur dan pelayanan, parkir, frekuensi dari transportasi umum.
- Faktor sosio-psikologi: kebiasaan, gaya hidup, persepsi tentang moda transportasi, pengalaman positif/negatif

Berdasarkan teori-teori penelitian yang telah dijelaskan dan dengan mempertimbangkan ketersediaan data, maka variabel pada penelitian ini dibatasi pada karakteristik perjalanan, karakteristik sosio-demografi dan sarana prasarana kota tujuan.

a. Karakteristik perjalanan.

- Waktu Tempuh

Waktu tempuh merupakan salah satu faktor yang paling utama yang harus sangat diperhatikan. Menurut Tamin (2000), waktu tempuh merupakan daya tarik utama dalam pemilihan moda transportasi. Jelas, bertambahnya waktu tempuh akan menurunkan jumlah penggunaan moda transportasi tertentu dan dengan sendirinya akan menurunkan pula tingkat pendapatannya.

- Jarak antara Tempat Tinggal dan Kegiatan Utama

Pada lokasi yang kepadatan penduduknya lebih tinggi daripada kesempatan kerja yang tersedia, terjadi surplus penduduk, dan mereka harus melakukan perjalanan ke pusat kota untuk bekerja (Tamin, 2000). Di sini terlihat bahwa makin jauh jarak dari pusat kota, makin banyak daerah perumahan dan makin sedikit kesempatan kerja yang berakibat makin banyak perjalanan yang terjadi antara daerah tersebut yang menuju pusat kota.

- Biaya Transportasi

Salah satu unsur utama yang menyatakan tingkat pelayanan adalah waktu tempuh, biaya perjalanan (tarif dan bahan bakar), dan juga hal lain seperti kenyamanan dan keamanan penumpang (Tamin, 2000).

b. Karakteristik sosio-demografi

Menurut Ma (2015), salah satu faktor yang mempengaruhi pemilihan moda transportasi adalah karakteristik sosio-demografi, antara lain: jenis kelamin, umur, status perkawinan, dan pendidikan. Misalnya untuk faktor umur, penduduk usia lanjut akan cenderung memilih angkutan yang lebih

nyaman dan kurang memperhatikan waktu tempuh, sementara penduduk usia muda sangat memperhitungkan masalah waktu tempuh.

Tamin (2000), tujuan pergerakan yang sering digunakan adalah pergerakan ke tempat kerja, pergerakan ke sekolah atau universitas (pergerakan dengan tujuan pendidikan), pergerakan ke tempat belanja dan pergerakan untuk kepentingan sosial dan rekreasi, dll. Dua tujuan pergerakan pertama (bekerja dan pendidikan) disebut tujuan pergerakan utama yang merupakan keharusan untuk dilakukan oleh setiap orang setiap hari, sedangkan tujuan pergerakan lain sifatnya hanya pilihan dan tidak rutin dilakukan.

c. Sarana prasarana kota tujuan.

- Jumlah Pergantian Moda Transportasi

Ma (2015) menyatakan bahwa rantai perjalanan dalam hal ini jumlah pergantian moda transportasi merupakan salah satu karakteristik perjalanan yang mempengaruhi pemilihan moda transportasi.

- PDRB Daerah Tujuan

Tamin (2000), faktor sosio-ekonomi yang teridentifikasi cukup signifikan sebagai peubah bebas pembangkit dan penarik pergerakan barang adalah populasi, PDRB, indeks kontribusi sektor industri, serta jumlah surplus (produksi transportasi barang) dan jumlah defisit (tarikan transportasi barang).

BAB 3

METODOLOGI PENELITIAN

3.1 Sumber Data

Data dalam penelitian ini berasal dari Survei Komuter Jabodetabek 2014 yang dilakukan oleh BPS. Survei Komuter Jabodetabek 2014 mencakup 13.120 rumah tangga sampel yang diperoleh dari 1.312 blok sensus (BS) yang tersebar di wilayah Jabodetabek yang terdiri dari tiga belas kabupaten/kota yaitu Kota Jakarta Selatan, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Timur, Kota Jakarta Utara, Kabupaten Bogor, Kota Bogor, Kota Depok, Kabupaten Tangerang, Kota Tangerang, Kota Tangerang Selatan, Kabupaten Bekasi, dan Kota Bekasi.

3.2 Metode Pengumpulan Data

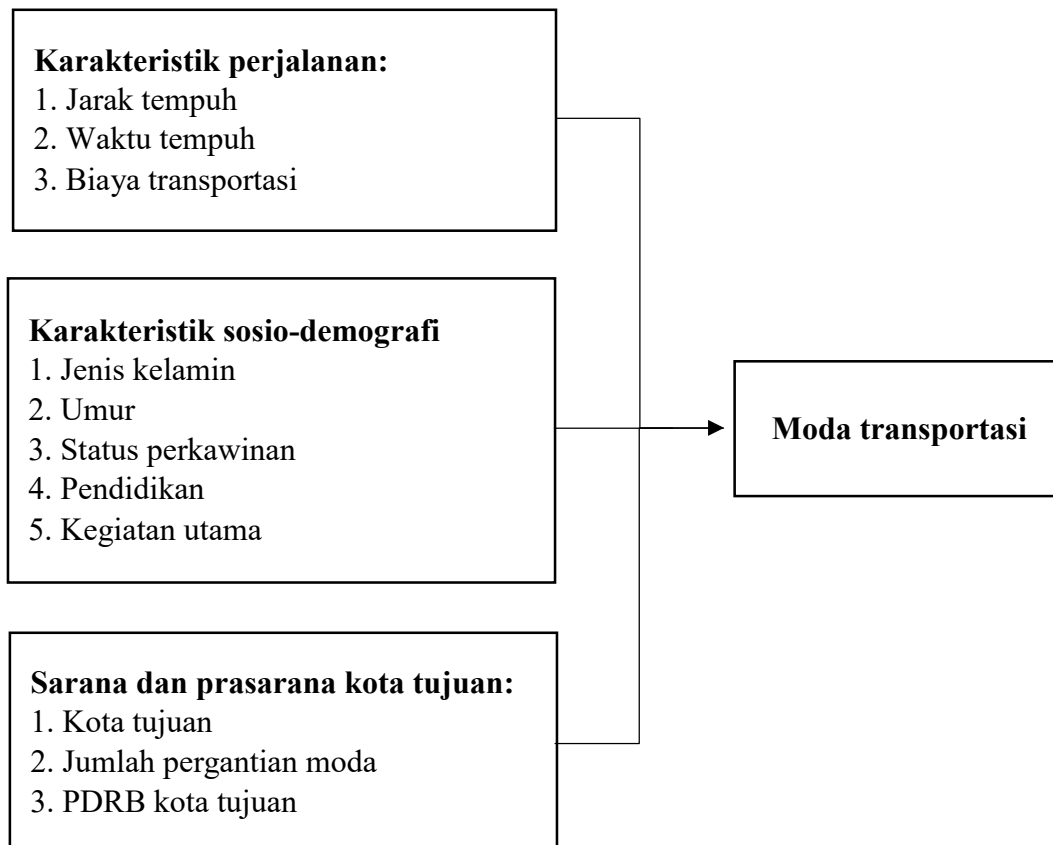
Pada tahun 2014, Badan Pusat Statistik melaksanakan survei komuter pertama yang terlengkap yaitu Survei Komuter di Jabodetabek. Kerangka sampel yang digunakan pada Survei Komuter 2014 terdiri dari kerangka sampel blok sensus dan kerangka sampel rumah tangga, dengan definisi sebagai berikut:

- 1) Kerangka sampel blok sensus adalah daftar blok sensus pada masing-masing Kabupaten/Kota terpilih survei komuter di setiap strata kecamatan.
Stratifikasi kecamatan dibuat dengan tujuan agar sebaran sampel lebih representatif terhadap sebaran penduduk komuter berdasarkan kecamatan. Stratifikasi dibedakan menurut strata rendah, sedang, dan tinggi. Variabel yang dijadikan dasar stratifikasi adalah penduduk berumur 15 tahun ke atas yang bekerja dari hasil Sensus Penduduk 2010 dan dievaluasi berdasarkan hasil Survei Penduduk Antar Sensus (SUPAS) dan Survei Angkatan Kerja Nasional (Sakernas) dengan asumsi semakin besar jumlah penduduk berumur 15 tahun ke atas yang bekerja maka jumlah komuter (termasuk bekerja) juga semakin besar.
- 2) Kerangka sampel rumah tangga adalah daftar seluruh rumah tangga hasil pemutakhiran di setiap blok sensus terpilih.

Sampel dipilih dengan cara *two stages sampling*. Langkah pertama dipilih sejumlah blok sensus per strata menggunakan *Probability Poportional to Size* (PPS) sistematis dengan size jumlah penduduk usia 15 tahun ke atas yang bekerja. Selanjutnya dari blok sensus yang terpilih, diambil 10 rumah tangga sampel menggunakan sistematis sampling setelah sebelumnya dilakukan pemutakhiran rumah tangga. Pencacahan dilaksanakan pada bulan Mei 2014. Pengumpulan data dilakukan secara *door-to-door* dengan mendatangi rumah tangga yang terpilih sebagai sampel.

3.3 Kerangka Pikir

Berdasarkan kajian teori serta ketersediaan data, maka disusunlah kerangka pikir untuk pemodelan pemilihan moda transportasi pada komuter di Jabodetabek sebagai berikut:



Gambar 3.1 Kerangka Pikir

3.4 Variabel Penelitian

Dengan mempertimbangkan literatur, penelitian terdahulu, serta ketersediaan data, maka variabel yang digunakan dalam penelitian ini terdiri 1 (satu) variabel kelas kategori moda transportasi dan 11 (sebelas) variabel atribut. Variabel respon dan variabel prediktor yang digunakan dalam penelitian ini tersaji dalam Tabel 3.1.

Tabel 3.1. Variabel Penelitian

Variabel	Deskripsi	Skala Data	Keterangan
(1)	(2)	(3)	(4)
Y	Moda Transportasi	Nominal	1 – 99
X ₁	Jenis kelamin	Nominal	1 : laki-laki 2 : perempuan
X ₂	Umur	Nominal	1 : kurang dari 16 tahun 2 : 16 – 25 tahun 3 : 26 – 35 tahun 4 : 36 – 45 tahun 5 : lebih dari 45 tahun
X ₃	Status perkawinan	Nominal	1 : belum kawin 2 : kawin/ cerai
X ₄	Pendidikan	Ordinal	1: SMP/ sederajat atau dibawahnya 2: SMA/ sederajat 3: perguruan tinggi
X ₅	Kegiatan utama	Nominal	1 : Pengusaha 2 : PNS 3 : karyawan swasta/pekerja bebas 4 : sekolah/ kursus
X ₆	Kota tujuan	Nominal	1 : DKI Jakarta 2 : Luar DKI Jakarta
X ₇	Jarak antara tempat tinggal dan kegiatan utama (km)	Nominal	1 : 0 – 9 2 : 10 – 19 3 : 20 – 29 4 : 30 – 39 5 : 40 atau lebih
X ₈	Waktu tempuh (menit)	Nominal	1 : <30 2 : 30 – 59 3 : 60 – 89 4 : 90 – 119 5 : 120 atau lebih
X ₉	Jumlah moda transportasi	Nominal	1 : 1 kali 2 : 2 kali

Lanjutan Tabel 3.1

Variabel (1)	Deskripsi (2)	Skala Data (3)	Keterangan (4)
			3 : 3 kali atau lebih
X ₁₀	Biaya transportasi (rupiah)	Nominal	1 : 0 – 6.000 2 : 6.001 – 12.000 3 : 12.001 – 18.000 4 : 18.001 – 24.000 5 : > 24.000
X ₁₁	PDRB kota tujuan (juta rupiah)	Ordinal	1 : 0 – 150.000 2 : 150.001 – 300.000 3 : lebih dari 300.000

Definisi operasional variabel penelitian yaitu:

1. Moda transportasi (Y) merupakan sarana transportasi yang digunakan untuk pergi/pulang ke/dari tempat kegiatan. Apabila menggunakan lebih dari satu moda transportasi, maka semua jenis moda transportasi yang digunakan dipilih/ditulisikan. Lebih lanjut kategori dari variabel respon moda transportasi diberikan pada lampiran 1.
2. Jenis Kelamin (X₁) merupakan jenis kelamin dari responden, terdapat dua kategori:
Kode 1 : laki-laki
Kode 2 : perempuan.
3. Umur (X₂) merupakan umur responden dihitung dalam tahun dengan pembulatan ke bawah atau umur pada waktu ulang tahun yang terakhir. Departemen Kesehatan RI (2009) mengelompokkan umur menjadi 9 kategori sesuai tingkat kematangan biologis seseorang, yaitu balita (0-5 tahun), kanak-kanak (5-11 tahun), remaja awal (12-16 tahun), remaja akhir (17-25 tahun), dewasa awal (26-35 tahun), dewasa akhir (36-45 tahun), lansia awal (46-55 tahun), lansia akhir (56-65 tahun), dan manula (65 tahun ke atas). Berdasarkan hal tersebut umur kepala rumah tangga dalam penelitian ini dikelompokkan menjadi 5 (lima) kategori:
Kode 1 : kurang dari 16 tahun
Kode 2 : 16 – 25 tahun
Kode 3 : 26 – 35 tahun
Kode 4 : 36 – 45 tahun

Kode 5 : lebih dari 45 tahun

4. Status perkawinan (X_3) merupakan status perkawinan dari responden, yaitu:

Kode 1 : belum kawin

Kode 2 : kawin/ cerai, yang mencakup: kawin, cerai hidup maupun cerai mati.

5. Pendidikan (X_4) merupakan tingkat pendidikan yang dicapai responden dengan mendapatkan tanda tamat (ijazah), yang dibagi menjadi empat kategori:

Kode 1 : SMP/ sederajat atau dibawahnya

Kode 2 : SMA/ sederajat

Kode 3 : perguruan tinggi

6. Kegiatan utama (X_5) merupakan kegiatan yang dilakukan komuter yang menggunakan waktu terbanyak dibandingkan kegiatan lainnya, yang dibagi menjadi empat kategori:

Kode 1 : Pengusaha

Kode 2 : PNS

Kode 3 : karyawan swasta/pekerja bebas

Kode 4 : sekolah/ kursus

7. Kota tujuan (X_6) adalah lokasi kegiatan yang dilakukan, yang dibagi menjadi empat kategori:

Kode 1 : DKI Jakarta

Kode 2 : Luar DKI Jakarta

8. Jarak antara tempat tinggal dan kegiatan utama (X_7) adalah jarak sekali jalan dari tempat tinggal sampai ke tempat kegiatan, bukan jarak pergi-pulang, dalam satuan kilometer (km) yang terdiri dari lima kategori:

Kode 1 : 0 – 10

Kode 2 : 10 – 19

Kode 3 : 20 – 29

Kode 4 : 30 – 39

Kode 5 : 40 atau lebih

9. Waktu tempuh (X_8) merupakan lama perjalanan dari rumah ke tempat kegiatan yang biasanya ditempuh komuter. Lamanya perjalanan ini dihitung

sejak komuter berangkat dari rumah sampai ke tempat kegiatan, termasuk waktu menunggu kendaraan umum (bagi yang menggunakan kendaraan umum) dalam satuan menit, terdapat lima kategori:

Kode 1 : <30

Kode 2 : 30 – 59

Kode 3 : 60 – 89

Kode 4 : 90 – 119

Kode 5 : 120 atau lebih

10. Jumlah moda transportasi (X_9) merupakan berapa kali komuter biasanya menggunakan moda transportasi untuk sampai ke tempat kegiatan yang terdiri dari tiga kategori:

Kode 1 : 1 kali

Kode 2 : 2 kali

Kode 3 : 3 kali atau lebih

11. Biaya transportasi (X_{10}) merupakan biaya yang dikeluarkan dalam sehari untuk melakukan perjalanan pergi dan pulang ke/dari tempat kegiatan dalam satuan rupiah yang dikategorikan menjadi lima yaitu:

Kode 1 : 0 – 6.000

Kode 2 : 6.001 – 12.000

Kode 3 : 12.001 – 18.000

Kode 4 : 18.001 – 24.000

Kode 5 : > 24.000

12. PDRB kota tujuan (X_{11}) merupakan jumlah nilai tambah barang dan jasa yang dihasilkan dari seluruh kegiatan perekonomian di suatu daerah tujuan komuter. Dalam penelitian ini PDRB dalam satuan juta rupiah dikategorikan menjadi:

Kode 1 : 0 – 150.000 juta rupiah

Kode 2 : 150.000 – 300.000 juta rupiah

Kode 3 : lebih dari 300.000 juta rupiah

3.5 Struktur Data

Struktur data dari penelitian ini berdasarkan variabel-variabel yang telah disebutkan sebelumnya disajikan dalam Tabel 3.2.

Tabel 3.2. Struktur Data

Individu (<i>i</i>)	Respon (Y_i)	Prediktor			
		X_1	X_2	...	X_{11}
(1)	(2)	(3)	(4)	(5)	(6)
1	Y_1	$X_{1,1}$	$X_{2,1}$...	$X_{11,1}$
2	Y_2	$X_{1,2}$	$X_{2,2}$...	$X_{11,2}$
⋮	⋮	⋮	⋮	...	⋮
n	Y_n	$X_{1,n}$	$X_{2,n}$...	$X_{11,n}$

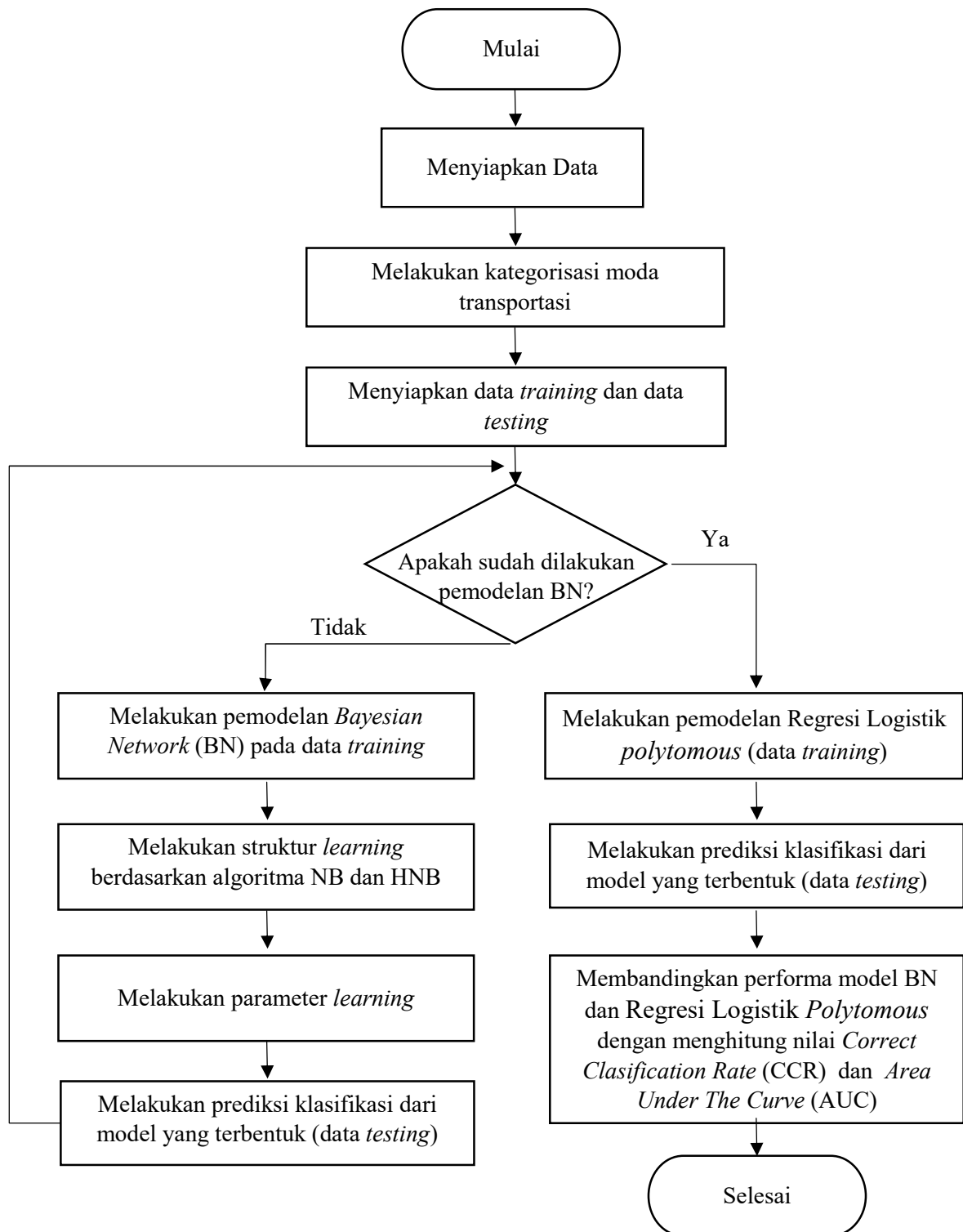
3.6 Metode dan Tahapan Penelitian

Metode dan tahapan penelitian yang akan dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut:

1. Persiapan data
 - a. Memilih variabel yang sesuai dari *micro data* Survei Komuter Jabodetabek 2014,
 - b. Melakukan penggabungan *micro data* antara keterangan individu dan keterangan rumah tangga,
 - c. Melakukan penyesuaian kategori pada variabel atribut (X),
 - d. Melakukan kategorisasi moda transportasi yang digunakan,
 - e. Menyiapkan data *training* (90 persen) dan data *testing* (10 persen).
2. Pembentukan model *Bayesian Network* pada data *training*
 - a. Melakukan struktur *learning*. Membangun struktur Bayes berdasarkan algoritma NB dan HNB.
 - b. Melakukan parameter *learning*. Menghitung probabilitas-probabilitas berdasarkan struktur yang terbentuk.
 - i. Menghitung probabilitas awal/ *prior probability* $\hat{P}(Y)$

- ii. Menghitung nilai probabilitas independen kelas Y dari semua fitur dalam vektor $\prod_{j \in p} \hat{P}(X_j|Y)$
 - iii. Menghitung nilai *posterior probability* untuk masing-masing klasifikasi $\hat{P}(Y|X)$
 - iv. Menghitung prediksi yang didapat dari nilai maksimum dalam perhitungan *posterior probability*
3. Pembentukan model regresi logistik pada data *training*. Meregresikan variabel kategori individu (Y) dan variabel atribut (X) dengan menggunakan model Regresi Logistik *Polytomous*.
- a. Membentuk model Regresi Logistik *Polytomous* menggunakan data training
 - b. Menguji signifikansi parameter secara individu dan secara keseluruhan
 - c. Melakukan prediksi dari model dengan data *testing*.
4. Perbandingan performa model
- a. Membuat tabulasi ketepatan klasifikasi moda transportasi berdasarkan metode *Bayesian Network* (NB dan HNB) pada data *testing*
 - b. Membuat tabulasi ketepatan klasifikasi moda transportasi berdasarkan metode Regresi Logistik *Polytomous* pada data *testing*
 - c. Menghitung nilai *Correct Clasification Rate* (CCR) dan *Area Under The Curve* (AUC) pada point 4.a dan 4.b
 - d. Membandingkan performa model.

Untuk lebih jelasnya metode dan tahapan penelitian dituangkan dalam Diagram Alir Penelitian pada Gambar 3.2.



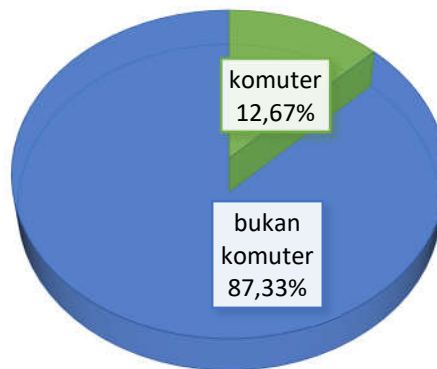
Gambar 3.2 Diagram Alir Penelitian

BAB 4

HASIL DAN PEMBAHASAN

4.1 Karakteristik Komuter Jabodetabek

Survei Komuter Jabodetabek 2014 adalah survei komuter pertama yang dilakukan BPS dan bertujuan untuk memberikan gambaran pola dan karakteristik komuter di Jabodetabek pada tahun 2014. Survei tersebut mencakup 13.120 rumah tangga sampel di 13 kabupaten/kota di Jabodetabek. Dari 47.421 responden, sebanyak 5.768 merupakan komuter. Cakupan wilayah adalah Provinsi DKI Jakarta (Jakarta Selatan, Jakarta Timur, Jakarta Pusat, Jakarta Barat, Jakarta Utara), Provinsi Jawa Barat (Kabupaten Bogor, Kota Bogor, Kota Depok, Kabupaten Bekasi). Dari sekitar 28 juta penduduk Jabodetabek yang berumur 5 tahun ke atas, sebanyak 12,67 persennya merupakan komuter.



Gambar 4.1 Persentase Komuter di Jabodetabek

Persentase komuter tertinggi terdapat di Kota Depok (19,88 persen), diikuti Kota Bekasi (19,82 persen), dan Kota Tangerang Selatan (18,37 persen), sedangkan persentase komuter terendah terdapat di Kabupaten Tangerang (5,95 persen), Kabupaten Bekasi (7,98 persen), dan Kota Bogor (8,24 persen). Untuk wilayah provinsi DKI Jakarta, persentase komuter hampir merata disemua wilayahnya dengan persentase tertinggi terdapat di wilayah Jakarta Pusat (15,11 persen), namun secara jumlah merupakan yang terendah.

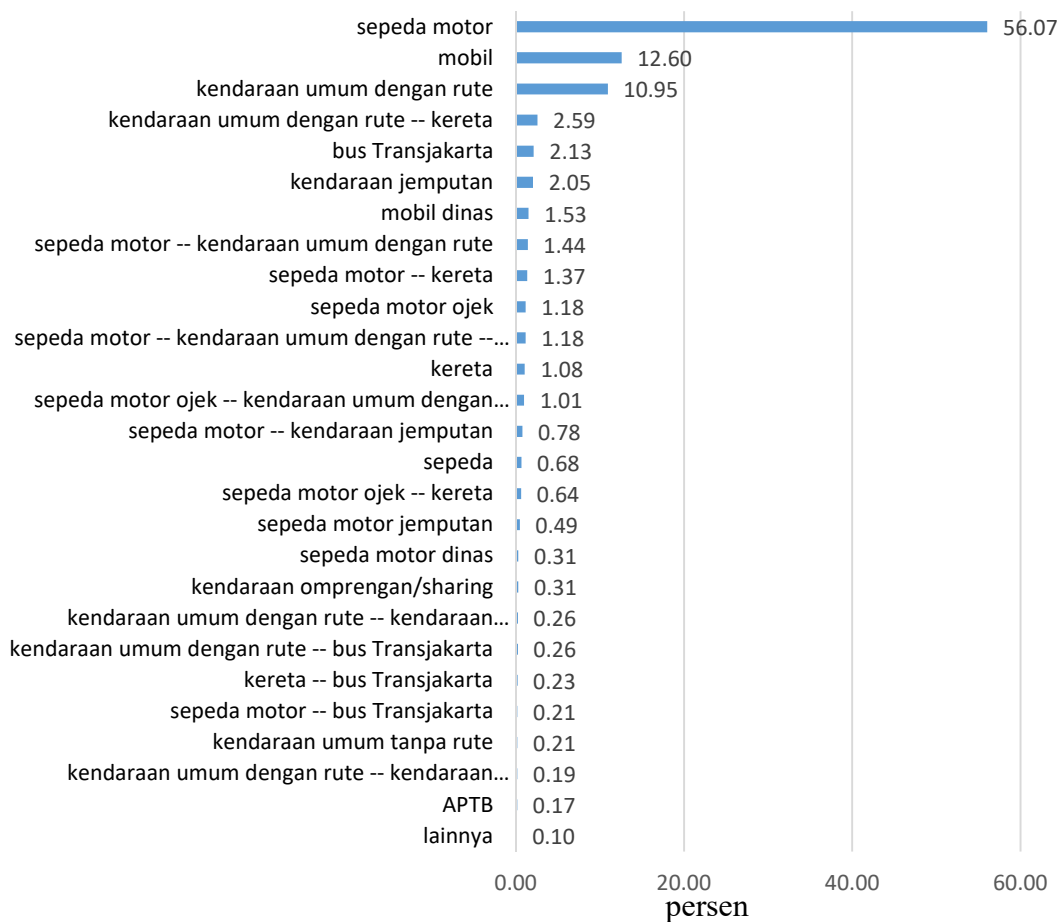
Tabel 4.1 Status Komuter Penduduk Jabodetabek

Kota Asal		Status Komuter		Jumlah
Kode Wilayah	Nama Wilayah	Komuter	Non Komuter	
(1)	(2)	(3)	(4)	(5)
3171	Jakarta Selatan	13,95%	86,05%	100,00%
3172	Jakarta Timur	14,79%	85,21%	100,00%
3173	Jakarta Pusat	15,11%	84,89%	100,00%
3174	Jakarta Barat	14,86%	85,14%	100,00%
3175	Jakarta Utara	11,28%	88,72%	100,00%
3201	Kabupaten Bogor	8,86%	91,14%	100,00%
3216	Kabupaten Bekasi	7,98%	92,02%	100,00%
3271	Kota Bogor	8,24%	91,76%	100,00%
3275	Kota Bekasi	19,82%	80,18%	100,00%
3276	Kota Depok	19,88%	80,12%	100,00%
3603	Kabupaten Tangerang	5,95%	94,05%	100,00%
3671	Kota Tangerang	15,67%	84,33%	100,00%
3674	Kota Tangerang Selatan	18,37%	81,63%	100,00%
Jumlah		12,67%	87,33%	100,00%

Sumber: Survei Komuter Jabodetabek 2014, Badan Pusat Statistik (diolah)

Pergerakan rutin para komuter Jabodetabek menjadi penyebab utama kemacetan lalu lintas. Peralihan moda transportasi komuter dari kendaraan pribadi ke angkutan umum merupakan solusi untuk mengurangi kemacetan. Meski demikian, para komuter belum semuanya berminat menggunakan angkutan umum. Hasil Survei Komuter 2014 menunjukkan bahwa hanya 27 persen komuter Jabodetabek yang menggunakan kendaraan umum untuk pergi ke tempat tujuan.

Moda transportasi yang digunakan oleh komuter yaitu sepeda motor (56,07 persen), mobil (12,60 persen), kendaraan umum dengan rute (10,95 persen), Kendaraan umum dengan rute – kereta (2,59 persen), bus Transjakarta (2,13 persen), kendaraan jemputan (2,05 persen), mobil dinas (1,53 persen), sepeda motor — kendaraan umum dengan rute (1,44 persen), sepeda motor—kereta (1,37 persen), sepeda motor ojek (1,18 persen), sepeda motor – kendaraan umum dengan rute – kereta (1,18 persen), serta lainnya yang dapat dilihat pada Gambar 4.2.



Gambar 4.2 Penggunaan Moda Transportasi oleh Komuter Jabodetabek

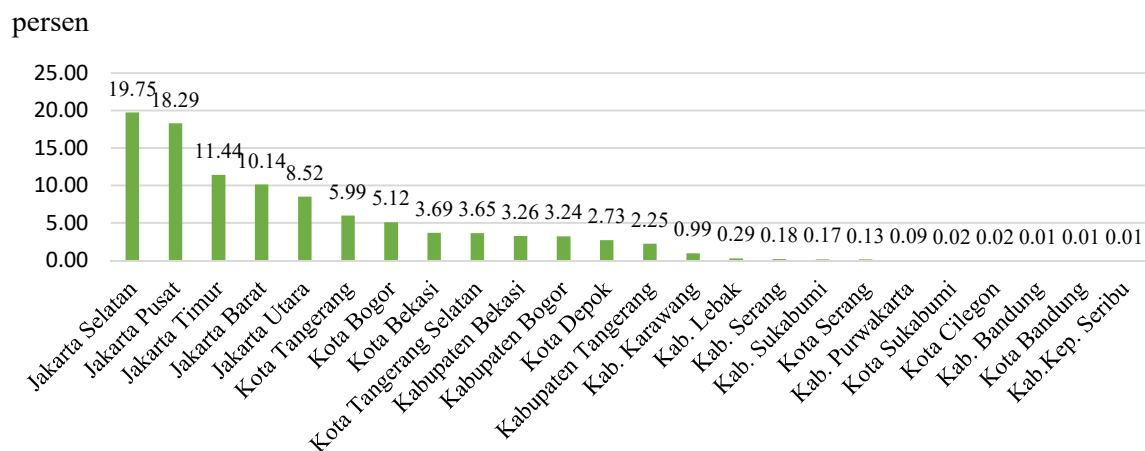
4.1.1 Karakteristik Sarana Prasarana Kota Tujuan

a. Kota Tujuan Komuter

Sebagian besar komuter yang tinggal di Jabodetabek melakukan kegiatan rutинnya di wilayah Jabodetabek juga. Hanya 1,93 persen komuter Jabodetabek yang melakukan kegiatan rutинnya di luar wilayah Jabodetabek. Kota-kota tujuan komuter Jabodetabek di luar wilayah Jabodetabek yaitu: Kota Karawang (0,99 persen), Kabupaten Lebak (0,29 persen), Kabupaten Serang (0,18 persen), Kabupaten Sukabumi (0,17 persen), Kota Serang (0,13 persen), Kab. Purwakarta (0,09 persen), Kota Sukabumi (0,17 persen), Kota Serang (0,13 persen), dst. Hal tersebut dimungkinkan karena berbagai alasan. Misalnya, Kabupaten Karawang

menjadi daerah tujuan komuter antara lain dari penduduk Kab. Bekasi, Kota Bekasi, Jakarta Timur, Jakarta Utara, dan Jakarta Barat. Hal ini dikarenakan Karawang merupakan kawasan industri dimana terdapat banyak pabrik dengan banyak karyawan sehingga dimungkinkan beberapa karyawannya yang telah memiliki rumah di luar Karawang memilih untuk menjadi komuter.

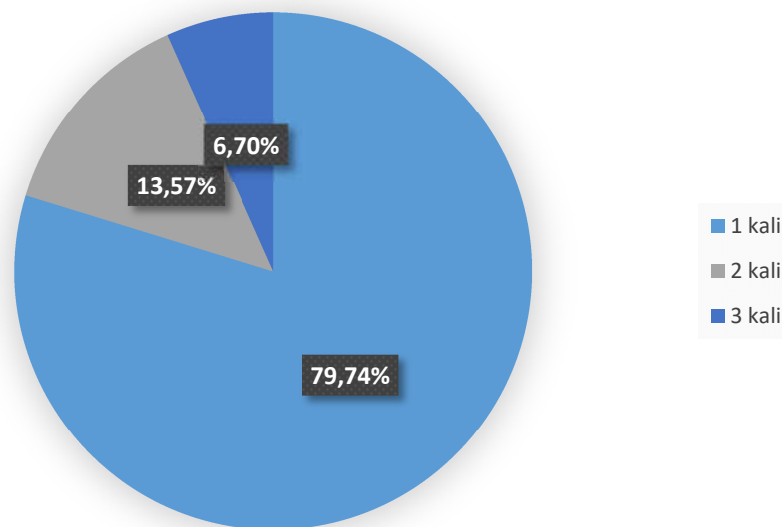
Pada Gambar 4.3 diperlihatkan kota tujuan komuter Jabodetabek. Kota tujuan utama komuter Jabodetabek adalah Kota Jakarta Selatan, Kota Jakarta Pusat, Kota Jakarta Timur, Kota Jakarta Barat, dan Kota Jakarta Utara.



Gambar 4.3 Kota Tujuan Komuter Jabodetabek

b. Jumlah Moda Transportasi

Penggunaan moda transportasi pada Gambar 4.2 memberikan gambaran bahwa sebagian komuter Jabodetabek melakukan pergantian jenis moda transportasi saat berangkat menuju tempat kegiatan. Berdasarkan Gambar 4.4 dapat dilihat bahwa 13,57 persen komuter berganti jenis moda transportasi sebanyak 2 kali, 6,70 persen komuter berganti sebanyak 3 kali atau lebih. Selebihnya sebanyak 79,74 persen hanya menggunakan moda transportasi sebanyak 1 kali. Sebagian besar komuter lebih menyukai menggunakan satu moda atau melakukan pergantian moda sebanyak satu kali dikarenakan lebih praktis dan tidak memerlukan waktu menunggu saat pergantian moda.



Gambar 4.4 Persentase Jumlah Pergantian Moda Transportasi

c. PDRB Kota Tujuan

Transportasi memiliki peranan yang penting dalam rangka mendorong peningkatan pertumbuhan ekonomi. Dimana pertumbuhan ekonomi merupakan tujuan utama pembangunan setiap daerah. Ukuran yang umum dilihat adalah Produk Domestik Regional Bruto (PDRB). Pada kota dengan PDRB tinggi diasumsikan sarana dan prasarana transportasi memadai. Dalam penelitian ini tinggi rendahnya PDRB menjadi salah satu indikator sarana dan prasarana kota tujuan. Pada Tabel 4.2 terlihat bahwa 20,14 persen komuter mempunyai kota tujuan dengan PDRB kurang dari 150 milyar, 18,60 persen mempunyai kota tujuan dengan PDRB 150 – 300 milyar, dan 61,26 persen mempunyai kota tujuan dengan PDRB lebih dari 300 milyar.

Tabel 4.2 Banyaknya Komuter Menurut PDRB Kota Tujuan

PDRB Kota Tujuan	Jumlah	Persentase
(1)	(2)	(3)
<150 milyar	1.161	20,14
150 - 300 milyar	1.072	18,60
>300 milyar	3.531	61,26
Jumlah	5.764	100,00

Sumber: Survei Komuter Jabodetabek 2014, Badan Pusat Statistik (diolah)

4.1.2 Karakteristik Perjalanan

Dari Tabel 4.3 terlihat bahwa jarak tempuh komuter Jabodetabek sangat beragam. Sebesar 25,45 persen komuter menempuh jarak kurang dari 10 km, sebesar 27,53 persen komuter menempuh jarak 10 – 19 km, sebesar 22,54 persen komuter menempuh jarak 20 – 29 km, sebesar 13,62 persen komuter menempuh jarak 30 – 39 km, sedangkan sisanya menempuh jarak lebih dari 40 km. Rata-rata waktu tempuh komuter Jabodetabek dari tempat tinggal ke tempat tujuan adalah 60,87 menit atau satu jam. Sebesar 31,45 persen komuter menempuh waktu 30 – 59 menit, sebesar 27,45 persen komuter menempuh waktu 60 – 89 menit, sebesar 15,60 persen komuter menempuh waktu 90 – 119 menit, sedangkan sisanya menempuh waktu kurang dari 30 menit dan lebih dari 120 menit dengan persentase masing-masing 12,75 persen dan 12,78 persen.

Tabel 4.3 Jarak Tempuh, Waktu Tempuh dan Biaya Transportasi Komuter

Variabel	Kategori	Jumlah	Persentase
(1)	(2)	(3)	(4)
Jarak (km)	0 – 9 km	1.467	25,45
	10 – 19 km	1.587	27,53
	20 – 29 km	1.299	22,54
	30 – 39 km	785	13,62
	40 km atau lebih	626	10,86
	Total	5.764	100,00
Waktu tempuh (menit)	< 30 menit	737	12,78
	30 – 59 menit	1.814	31,45
	60 – 89 menit	1.582	27,45
	90 – 119 menit	900	15,61
	120 menit atau lebih	735	12,75
	Total	5.764	100,00
Biaya (rupiah)	Rp 0,00 – Rp 6.000,00	943	16,36
	Rp 6.001,00 – Rp 12.000,00	2.128	36,92
	Rp 12.001,00 – Rp 18.000,00	1.031	17,89
	Rp 18.001,00 – Rp 24.000,00	567	9,84
	> Rp 24.000,00	1.095	19,00
	Total	5.764	100,00

Sumber: Survei Komuter Jabodetabek 2014, Badan Pusat Statistik (diolah)

Rata-rata biaya yang dikeluarkan oleh komuter untuk transportasi ke tempat kegiatan adalah sebesar Rp 18.193,77 dengan persentase pengeluaran terbanyak

pada rentang biaya Rp 6.001,00 – Rp 12.000,00. Terdapat 19,00 persen komuter dengan pengeluaran biaya transportasi lebih dari Rp 24.000,- untuk sekali pergi dari tempat tinggal ke tempat kegiatan.

Dari Tabel 4.4, jarak tempuh maksimal pada komuter Jabodetabek adalah 104 km, dengan rata-rata jarak yang harus ditempuh oleh komuter 19,95 km. Sebanyak 50 persen komuter menempuh jarak 17 km atau kurang, 25 persen komuter menempuh jarak 17 km sampai 28 km, dan variasi jarak terbesar ada pada 25 persen kelompok komuter yang menempuh jarak 28 km sampai 104 km.

Tabel 4.4 Sebaran Jarak Tempuh, Waktu Tempuh dan Biaya Transportasi Komuter

Variabel	Minimal	Q1	Median	Q3	Maksimal	Rata-rata
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Jarak (km)	1,00	9,00	17,00	28,00	104,00	19,9563
Waktu (menit)	3,00	30,00	60,00	90,00	240,00	60,8730
Biaya (rupiah)	0,00	7.000,00	11.000,00	20.000,00	310.000,00	18.193,77

Sumber: Survei Komuter Jabodetabek 2014, Badan Pusat Statistik (diolah)

Waktu tempuh maksimal pada komuter Jabodetabek adalah 240 menit, dengan rata-rata waktu tempuh 60,87 menit. Sebanyak 25 persen komuter menempuh lama perjalanan selama 30 menit atau kurang, 25 persen komuter menempuh lama perjalanan selama 30 menit sampai 60 menit, 25 persen komuter menempuh lama perjalanan selama 60 menit sampai 90 menit dan variasi lama perjalanan terbesar ada pada 25 persen kelompok komuter yang menempuh lama perjalanan 90 menit sampai 240 menit.

Biaya transportasi maksimum Rp 310.000, 00 dengan rata-rata biaya transportasi Rp 18.193,77. Sebanyak 25 persen komuter mengeluarkan biaya transportasi sebesar Rp 7.000,00 atau kurang, 25 persen komuter mengeluarkan biaya transportasi sebesar Rp 7.000,00 sampai Rp 11.000,00 , 25 persen komuter mengeluarkan biaya transportasi sebesar Rp 11.000,00 sampai Rp 20.000,00 dan variasi biaya transportasi terbesar ada pada 25 persen kelompok komuter mengeluarkan biaya transportasi sebesar Rp 20.000,00 sampai Rp 310.000,00.

4.1.3 Karakteristik Sosio-Demografi Komuter

Berdasarkan Tabel 4.5, sebanyak 68,10 persen komuter berjenis kelamin laki-laki dan sisanya sebanyak 31,90 persen berjenis kelamin perempuan. Komuter Jabodetabek yang berstatus kawin/ cerai jumlahnya lebih banyak dibandingkan yang belum kawin, yaitu 58,57 persen dibandingkan 41,43 persen. Status kawin/ ceraidari komuter ini mencakup komuter yang berstatus cerai, baik cerai hidup maupun. cerai mati.

Sebagian besar komuter yaitu sebanyak 46,22 persen merupakan lulusan SMA/ sederajat, 31,23 persen merupakan lulusan perguruan tinggi. Namun, tidak semua komuter berpendidikan tinggi. Sebanyak 22,55 persen komuter merupakan tamatan SMP/ sederajat , SD/ sederajat, tidak tamat SD, maupun belum pernah sekolah. Termasuk di dalamnya komuter yang masih bersekolah.

Tabel 4.5 Karakteristik Individu Komuter Jabodetabek

Variabel	Kategori	Jumlah	Persentase
(1)	(2)	(3)	(4)
Jenis Kelamin	Laki-laki	3.925	68,10
	Perempuan	1.839	31,90
	Total	5.764	100,00
Status Kawin	Belum Kawin	2.388	41,43
	Kawin/ cerai	3.376	58,57
	Total	5.764	100,00
Pendidikan	SMP/ sederajat atau dibawahnya	1.300	22,55
	SMA/ sederajat	2.664	46,22
	Perguruan tinggi	1.800	31,23
	Total	5.764	100,00
Umur	5 – 15 tahun	407	7,06
	16 – 25 tahun	1.505	26,11
	26 – 35 tahun	1.441	25,00
	36 – 45 tahun	1.317	22,85
	lebih dari 45 tahun	1.094	18,98
	Total	5.764	100,00
Kegiatan Utama	Pengusaha	366	6,35
	PNS	402	6,97
	Karyawan Swasta/ Pekerja Bebas	3.934	68,25
	Sekolah/ kursus	1.062	18,42
	Total	5.764	100,00

Sumber: Survei Komuter Jabodetabek 2014, Badan Pusat Statistik (diolah)

Sebesar 26,11 persen komuter berusia 16 – 25 tahun, 25,00 persen komuter berusia 26 – 35 tahun, 22,85 persen komuter berusia 36 – 45 tahun, 18,98 persen komuter berusia lebih dari 45 tahun, dan 7,06 persen komuter berusia 5 – 15 tahun. Kegiatan utama komuter baik pada laki-laki maupun perempuan adalah bekerja, yaitu 68,25 persen merupakan karyawan swasta/ pekerja bebas, 6,97 persen merupakan PNS dan 6,35 persen merupakan pengusaha. Selain bekerja, tujuan utama komuter adalah untuk sekolah dan kursus.

4.2 Pemodelan dengan *Naïve Bayes* (NB)

4.2.1 Estimasi Parameter Model *Naïve Bayes*

Klasifikasi termasuk dalam kelompok *supervised learning*, dimana akan mengestimasi fungsi $f: X \rightarrow Y$ atau ekuivalen dengan $P(Y|X)$. Misalkan Y adalah random variabel dan X adalah vektor yang terdiri dari j atribut. Dengan kata lain, $\mathbf{X} = (X_1, X_2, \dots, X_p)$, $j = 1, 2, \dots, p$, masing-masing X_j terdiri dari variabel random dari sebanyak n .

Aplikasi Teorema Bayes pada $P(Y = y_k | X)$ dapat dijabarkan dalam:

$$P(Y = y_k | X = x_h) = \frac{P(X = x_h | Y = y_k)}{\sum_l P(X = x_h | Y = y_l) P(Y = y_l)} \quad (4.1)$$

dimana y_k merupakan seluruh kemungkinan nilai untuk Y dan x_h merupakan seluruh kemungkinan nilai vektor untuk X .

Menurut Mitchell (2017), salah satu cara untuk mendapatkan $P(Y|X)$ adalah dengan menggunakan data *training* untuk mengestimasi $P(X|Y)$ dan $P(Y)$. Hasil estimasi tersebut bersama-sama dengan Teorema Bayes di atas guna menentukan $P(Y|X = x_h)$ untuk x_h data baru. Parameter yang akan diestimasi untuk dua set parameter untuk:

$$\hat{\pi}_l = P(Y = y_l) \quad (4.2)$$

$$\hat{\theta}_{gl} = P(X_j = x_g | Y = y_l) \quad (4.3)$$

Asumsi *conditional independence* diperlukan untuk mengurangi parameter yang harus diestimasi pada $P(X|Y)$.

$$(\forall g, l, m) P(X = x_g | Y = y_j, Z = z_k) = P(X = x_g | Z = z_k) \quad (4.4)$$

Penurunan Algoritma *Naïve Bayes*

$$\begin{aligned} P(X|Y) &= P(X_1, X_2|Y) \\ &= P(X_1|X_2, Y) P(X_2|Y) \\ &= P(X_1|Y) P(X_2|Y) \end{aligned}$$

Dengan asumsi conditional independen maka:

$$P(X_1, \dots, X_p | Y) = \prod_{i=1}^p P(X_i | Y) \quad (4.5)$$

Probabilitas Y saat kemungkinan nilai k :

$$P(Y = y_k | X_1, \dots, X_p) = \frac{P(Y = y_k) P(X_1, \dots, X_p | Y = y_k)}{\sum_l P(Y = y_l) P(X_1, \dots, X_p | Y = y_l)} \quad (4.6)$$

$$P(Y = y_k | X_1, \dots, X_p) = \frac{P(Y = y_k) \prod_j P(X_j | Y = y_k)}{\sum_l P(Y = y_l) \prod_j P(X_j | Y = y_l)} \quad (4.7)$$

$$Y \leftarrow \arg \max_{y_k} \frac{P(Y = y_k) \prod_j P(X_j | Y = y_k)}{\sum_l P(Y = y_l) \prod_j P(X_j | Y = y_l)} \quad (4.8)$$

Karena denominator atau *normalized constant* tidak bergantung pada y_k sehingga bisa disederhanakan menjadi:

$$Y \leftarrow \arg \max_{y_k} P(Y = y_k) \prod_j P(X_j | Y = y_k) \quad (4.9)$$

Untuk data diskrit, maka estimasi dua set parameter NB pada persamaan (4.2) dan persamaan (4.3) menggunakan hasil MLE dari persamaan (2.16) dan persamaan (2.17) maka menjadi:

$$\theta_{jgl} \equiv P(X_j = x_{gl} | Y = y_k) \quad (4.10)$$

sehingga estimasi dari θ_{jgl} adalah:

$$\hat{\theta}_{jgl} = \hat{P}(X_j = x_{gl} | Y = y_k) = \frac{\#D\{X_j = x_{gl} \wedge Y = y_k\}}{\#D\{Y = y_k\}} \quad (4.11)$$

dimana $\#D\{x\}$ merupakan banyaknya elemen dari set data D yang memenuhi x .

Selanjutnya untuk satu set parameter kedua yaitu:

$$\pi_l \equiv P(Y = y_l) \quad (4.12)$$

Maka estimasi dari π_l adalah:

$$\hat{\pi}_k = \hat{P}(Y = y_k) = \frac{\#D\{Y = y_k\}}{|D|} \quad (4.13)$$

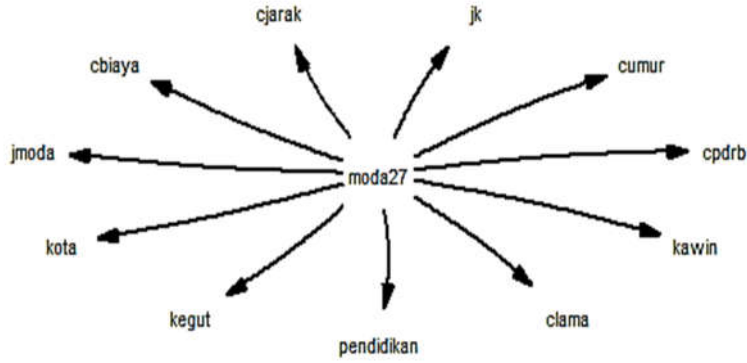
dimana $\#D\{x\}$ merupakan banyaknya elemen dari set data D yang memenuhi x .

4.2.2 Klasifikasi dengan Metode *Naïve Bayes*

Model NB dibangun dengan asumsi seluruh atribut saling independen bersyarat variabel klasifikasi. Dalam *Bayesian Network*, secara umum terdapat dua hal dalam pemodelan, yaitu membangun struktur dan mengestimasi parameter.

a. Struktur *Naïve Bayes*

NB dibangun dengan asumsi seluruh atribut independen bersyarat variabel klasifikasi. Dengan data training yang memiliki 11 atribut dan satu variabel klasifikasi maka struktur NB yang terbentuk dapat digambarkan pada Gambar 4.5



Gambar 4.5 Struktur *Naïve Bayes* Klasifikasi Moda Transportasi

Berdasarkan struktur yang terbentuk pada Gambar 4.5 tersebut, kemudian dapat dibuat peluang bersama seperti pada persamaan (4.14) dengan melibatkan seluruh variabel atribut.

$$\begin{aligned}
 P(Y|X_1, X_2, \dots, X_{11}) = & P(Y) * P(X_1|Y) * P(X_2|Y) * P(X_3|Y) * P(X_4|Y) * P(X_5|Y) * \\
 & P(X_6|Y) * P(X_7|Y) * P(X_8|Y) * P(X_9|Y) * P(X_{10}|Y) * \\
 & P(X_{11}|Y)
 \end{aligned} \quad (4.14)$$

b. Estimasi Parameter Model *Naïve Bayes*

Persamaan (4.14) merupakan peluang bersama model NB yang terbentuk. Langkah selanjutnya adalah melakukan estimasi parameter. Parameter dalam model *Bayesian Network* menggambarkan distribusi peluang sesuai dengan struktur yang terbentuk. Dengan metode MLE maka memaksimumkan fungsi likelihood sama dengan memaksimumkan masing-masing peluang lokal. Secara praktis besarnya peluang lokal adalah proporsi kejadian pada kategori ke-*l* variabel ke-*j* dan kelas ke-*c* atau dengan kata lain *cross* tabulasi antara variabel atribut dengan variabel klasifikasi yang diproporsikan terhadap kolom. Tabel 4.6 menampilkan hasil estimasi parameter berdasarkan data *training* untuk peluang prior $\hat{P}(Y)$ NB.

Tabel 4.6 Estimasi Peluang Prior $\hat{P}(Y)$ Naïve Bayes

$\hat{P}(Y=l)$							
$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=13$	$l=16$	$l=27$
(1)	(2)	(3)	(4)	(5)	(13)	(16)	(27)
0,0015	0,0222	0,0214	0,0031	0,1099	0,1268	0,5588	0,0010

Sedangkan hasil estimasi parameter berdasarkan data *training* untuk peluang bersyarat $\hat{P}(X|Y)$ NB tersaji pada Tabel 4.7. Untuk lebih lengkapnya, hasil estimasi parameter HNB dapat dilihat pada lampiran 4.

Tabel 4.7. Estimasi Parameter Peluang Bersyarat $\hat{P}(X|Y)$ Naïve Bayes

$\hat{P}(X_j = h Y = l)$			Y						
			$l = 1$	$l = 2$	$l = 3$	$l = 5$	$l = 13$	$l = 16$	$l = 27$
(1)			(2)	(3)	(4)	(6)	(14)	(17)	(28)
X_1	Jenis kelamin	$h=1$	0,5000	0,4261	0,4955	0,3965	0,7006	0,7961	0,6000
		$h=2$	0,5000	0,5739	0,5045	0,6035	0,2994	0,2039	0,4000
X_2	Umur	$h=1$	0,0000	0,0000	0,2883	0,1439	0,0942	0,0469	0,4000
		$h=2$	0,1250	0,3826	0,1622	0,3509	0,0851	0,2853	0,0000
		$h=3$	0,3750	0,2957	0,1261	0,1719	0,1596	0,2984	0,0000
		$h=4$	0,2500	0,1826	0,1982	0,1632	0,2857	0,2328	0,2000
		$h=5$	0,2500	0,1391	0,2252	0,1702	0,3754	0,1366	0,4000
X_3	Status kawin	$h=1$	0,3750	0,6174	0,4775	0,5842	0,2629	0,4146	0,4000
		$h=2$	0,6250	0,3826	0,5225	0,4158	0,7371	0,5854	0,6000
X_4	Pendidikan	$h=1$	0,0000	0,1043	0,4685	0,3912	0,1489	0,2059	0,8000
		$h=2$	0,5000	0,5130	0,2432	0,4158	0,1824	0,5612	0,2000
		$h=3$	0,5000	0,3826	0,2883	0,1930	0,6687	0,2328	0,0000
X_5	Kegiatan utama	$h=1$	0,0000	0,0261	0,0270	0,0333	0,1535	0,0566	0,4000
		$h=2$	0,0000	0,0174	0,2432	0,0368	0,1170	0,0500	0,0000
		$h=3$	1,0000	0,7739	0,4144	0,5982	0,5638	0,7330	0,2000
		$h=4$	0,0000	0,1826	0,3153	0,3316	0,1656	0,1604	0,4000
X_6	Kota tujuan	$h=1$	0,8750	1,0000	0,6216	0,6875	0,8085	0,7051	0,5000
		$h=2$	0,1250	0,0000	0,3784	0,3125	0,1915	0,2949	0,5000
X_7	Jarak tempuh	$h=1$	0,0000	0,1913	0,2523	0,3947	0,2006	0,2663	1,0000
		$h=2$	0,0000	0,4087	0,2613	0,2737	0,2553	0,3139	0,0000
		$h=3$	0,2500	0,2261	0,1622	0,1596	0,2644	0,2273	0,0000
		$h=4$	0,2500	0,1478	0,1622	0,0982	0,1535	0,1221	0,0000
		$h=5$	0,5000	0,0261	0,1622	0,0737	0,1261	0,0704	0,0000
X_8	Waktu tempuh	$h=1$	0,0000	0,0348	0,0811	0,1281	0,0745	0,1587	1,0000
		$h=2$	0,0000	0,3130	0,2432	0,3719	0,2416	0,3549	0,0000
		$h=3$	0,1250	0,2696	0,3063	0,2281	0,3040	0,2746	0,0000
		$h=4$	0,3750	0,2087	0,1171	0,1263	0,1960	0,1342	0,0000
		$h=5$	0,5000	0,1739	0,2522	0,1456	0,1839	0,0776	0,0000
X_9	Jumlah moda	$h=1$	0,0000	0,2261	0,8468	0,4105	0,9863	0,9948	1,0000
		$h=2$	0,6250	0,6348	0,1441	0,4263	0,0137	0,0045	0,0000
		$h=3$	0,3750	0,1391	0,0090	0,1632	0,0000	0,0007	0,0000

Lanjutan Tabel 4.7

$\hat{P}(X_j = h Y = l)$			Y						
			$l = 1$	$l = 2$	$l = 3$	$l = 5$	$l = 13$	$l = 16$	$l = 27$
(1)			(2)	(3)	(4)	(6)	(14)	(17)	(28)
X_{10}	Biaya transportasi	$h=1$	0,0000	0,0870	0,1351	0,2281	0,0228	0,1987	1,0000
		$h=2$	0,0000	0,2957	0,4685	0,3316	0,0258	0,5005	0,0000
		$h=3$	0,0000	0,4087	0,0991	0,1667	0,0380	0,2039	0,0000
		$h=4$	0,1250	0,1217	0,1532	0,1491	0,0775	0,0676	0,0000
		$h=5$	0,8750	0,0870	0,1441	0,1246	0,8359	0,0293	0,0000
X_{11}	PDRB kota tujuan	$h=1$	0,0000	0,0000	0,2162	0,2281	0,1444	0,2273	0,2000
		$h=2$	0,2500	0,1565	0,1892	0,2140	0,1565	0,1818	0,0000
		$h=3$	0,7500	0,8435	0,5946	0,5579	0,6991	0,5909	0,8000

Hasil dari estimasi parameter dapat menjelaskan kecenderungan komuter memilih moda transportasi apabila nilai variabel atribut diketahui. Tiga peluang prior terbesar yaitu sepeda motor ($Y=16$), mobil ($Y=13$), dan kendaraan umum dengan rute ($Y=5$). Komuter dengan jenis kelamin laki-laki cenderung memilih menggunakan sepeda motor dibandingkan komuter yang berjenis kelamin perempuan $\hat{P}(X_1 = 1 | Y = 16) > \hat{P}(X_1 = 2 | Y = 16)$, dimana diketahui dari Tabel 4.7 bahwa $\hat{P}(X_1 = 1 | Y = 16) = 0,7961$ dan $\hat{P}(X_1 = 2 | Y = 16) = 0,2039$. Moda transportasi mobil juga cenderung dipilih oleh komuter dengan jenis kelamin laki-laki tetapi tidak demikian dengan moda transportasi kendaraan umum dengan rute yang cenderung diminati oleh komuter dengan jenis kelamin perempuan.

Pada variabel umur, sepeda motor juga cenderung dipilih oleh kelompok umur 26-35 tahun, 36-45 tahun, dan 16-25 tahun, sedangkan mobil cenderung dipilih oleh komuter dengan kelompok umur lebih dari 45 tahun $\hat{P}(X_2 = 5 | Y = 13) = 0,3754$. Sementara kendaraan umum dengan rute cenderung dipilih oleh komuter dengan kelompok umur 16-25 tahun $\hat{P}(X_2 = 2 | Y = 5) = 0,3509$.

Sepeda motor dan mobil cenderung dipilih oleh komuter yang sudah menikah/ cerai, sedangkan kendaraan umum dengan rute cenderung dipilih oleh komuter yang belum kawin. Dari segi pendidikan, sepeda motor dan kendaraan umum dengan rute cenderung dipilih oleh komuter dengan pendidikan SMA/ sederajat, sedangkan mobil cenderung dipilih oleh komuter yang lulus perguruan

tinggi. Dari segi kegiatan utama, tujuan menuju DKI Jakarta, dan PDRB kota tujuan baik sepeda motor, mobil maupun kendaraan umum dengan rute dipilih oleh komuter dengan kegiatan utama karyawan swasta/pekerja bebas, bertujuan ke DKI Jakarta dan PDRB kota tujuan lebih dari 300 Milyar.

Dari segi jarak tempuh dan waktu tempuh, sepeda motor cenderung dipilih komuter dengan rentang jarak 10-19 km dan waktu tempuh 30-59 menit, mobil cenderung dipilih oleh komuter dengan rentang jarak 20-29 km dan waktu tempuh 60-89 menit, sedangkan kendaraan umum dengan rute dipilih komuter dengan rentang jarak pendek yaitu 0-9 km dan waktu tempuh kurang dari 30 menit. Komuter yang memilih moda transportasi sepeda motor dan mobil memilih kepraktisan karena jumlah moda hanya sekali. Komuter yang memilih moda transportasi sepeda motor dan kendaraan umum dengan rute cenderung mengeluarkan biaya transportasi dengan rentang Rp 6.001,00 – Rp 12.000,00. Komuter yang memilih moda transportasi mobil cenderung mengeluarkan biaya transportasi lebih dari Rp 24.000,00. Begitu pula interpretasi untuk moda transportasi lainnya.

c. Kriteria Klasifikasi Model *Naïve Bayes*

Pada model NB, klasifikasi dilakukan dengan menghitung *posterior probability* $\hat{P}(X|Y)$ sesuai dengan struktur NB dan estimasi parameter berdasarkan karakteristik komuter. Komuter terklasifikasi menggunakan moda kategori tertentu, misalnya dihitung nilai posterior untuk semua kategori (I) pada komuter dengan data atributnya, dan didapatkan peluang posterior yang maksimum adalah peluang Y dengan kategori ke- k ($Y = y_k$), maka komuter tersebut diklasifikasikan sebagai y_k . Formula untuk mendapatkan peluang kategori maksimum pada persamaan (2.23). Estimasi peluang posterior NB untuk $Y = 1$ (APTB), $Y = 2$ (bus Transjakarta), $Y = 5$ (kendaraan umum dengan rute), $Y = 13$ (mobil), $Y = 16$ (sepeda motor) dan $Y = 99$ (lainnya) adalah sebagai berikut:

$$\begin{aligned}
\hat{P}(Y=1|X_1, X_2, \dots, X_{11}) &\propto P(Y=1) * P(X_1|Y=1) * P(X_2|Y=1) * P(X_3|Y=1) * \\
&\quad P(X_4|Y=1) * P(X_5|Y=1) * P(X_6|Y=1) * P(X_7|Y=1) * \\
&\quad P(X_8|Y=1) * P(X_9|Y=1) * P(X_{10}|Y=1) * P(X_{11}|Y=1) \\
\hat{P}(Y=2|X_1, X_2, \dots, X_{11}) &\propto P(Y=2) * P(X_1|Y=2) * P(X_2|Y=2) * P(X_3|Y=2) * \\
&\quad P(X_4|Y=2) * P(X_5|Y=2) * P(X_6|Y=2) * P(X_7|Y=2) * \\
&\quad P(X_8|Y=2) * P(X_9|Y=2) * P(X_{10}|Y=2) * P(X_{11}|Y=2) \\
&\quad \vdots \\
\hat{P}(Y=5|X_1, X_2, \dots, X_{11}) &\propto P(Y=5) * P(X_1|Y=5) * P(X_2|Y=5) * P(X_3|Y=5) * \\
&\quad P(X_4|Y=5) * P(X_5|Y=5) * P(X_6|Y=5) * P(X_7|Y=5) * \\
&\quad P(X_8|Y=5) * P(X_9|Y=5) * P(X_{10}|Y=5) * P(X_{11}|Y=5) \\
&\quad \vdots \\
\hat{P}(Y=13|X_1, X_2, \dots, X_{11}) &\propto P(Y=13) * P(X_1|Y=13) * P(X_2|Y=13) * P(X_3|Y=13) * \\
&\quad P(X_4|Y=13) * P(X_5|Y=13) * P(X_6|Y=13) * P(X_7|Y=13) * \\
&\quad P(X_8|Y=13) * P(X_9|Y=13) * P(X_{10}|Y=13) * P(X_{11}|Y=13) \\
&\quad \vdots \\
\hat{P}(Y=16|X_1, X_2, \dots, X_{11}) &\propto P(Y=16) * P(X_1|Y=16) * P(X_2|Y=16) * P(X_3|Y=16) * \\
&\quad P(X_4|Y=16) * P(X_5|Y=16) * P(X_6|Y=16) * P(X_7|Y=16) * \\
&\quad P(X_8|Y=16) * P(X_9|Y=16) * P(X_{10}|Y=16) * P(X_{11}|Y=16) \\
&\quad \vdots \\
\hat{P}(Y=99|X_1, X_2, \dots, X_{11}) &\propto P(Y=99) * P(X_1|Y=99) * P(X_2|Y=99) * P(X_3|Y=99) * \\
&\quad P(X_4|Y=99) * P(X_5|Y=99) * P(X_6|Y=99) * P(X_7|Y=99) * \\
&\quad P(X_8|Y=99) * P(X_9|Y=99) * P(X_{10}|Y=99) * P(X_{11}|Y=99)
\end{aligned}$$

d. Klasifikasi Model *Naïve Bayes* pada Data *Testing*

Berdasarkan model NB yang telah terbentuk (struktur dan estimasi parameternya), klasifikasi dilakukan pada data *testing* yang berjumlah 576 komuter. Terdapat 191 komuter yang salah prediksi moda transportasi yang digunakannya dan sebanyak 385 moda transportasi komuter yang diprediksi secara benar sehingga nilai *Correct Clasification Rate* (CCR) sebesar 66,8403

persen. Nilai *Area Under The Curve* (AUC) sebesar 64,5369 persen yang artinya metode NB mengklasifikasikan dengan benar 64,5369 persen moda transportasi yang digunakan oleh komuter. *Syntax* R dan output yang digunakan untuk melakukan pemodelan klasifikasi dengan metode NB dapat dilihat pada Lampiran 4 dan Lampiran 5.

4.3 Pemodelan dengan *Hierarchical Naïve Bayes* (HNB)

Model HNB dibangun dengan mempertimbangkan adanya hubungan antar variabel atribut. Algoritma pemodelan dengan HNB adalah sebagai berikut:

1. Menentukan partisi data menjadi S subset subset $D^{(1)}, D^{(2)}, \dots, D^{(v)}$, (*overlapping*) dimana $s = 1, 2, \dots, v$.

Dalam penelitian ini akan dibagi dalam 10 subset ($v = 10$). Data training sebanyak DN=5188. $D^{(s)}$ merupakan sampel data yang diambil dari kombinasi 90 persen, 85 persen, 80 persen, 75 persen dan 70 persen data *training* secara random dan *overlapping*.

2. Untuk $r = 0, 1, 2, \dots, p-1$, lakukan:
 - a. Hitung nilai *conditional independence* (CI) pada setiap pasangan variabel atribut pada masing-masing subset data $D^{(s)}$, dengan hipotesis nol dalam uji CI adalah dua variabel atribut saling independen, sedangkan hipotesis alternatifnya dua variabel atribut tidak saling independen. Statistik Uji mengikuti distribusi $\chi^2_{(j-1)(j^*-1)}$ yang telah dibuktikan pada persamaan (2.27) dan persamaan (2.28).
 - b. Pembentukan struktur variabel atribut menggunakan nilai signifikansi *conditional independence* (CI) pada masing-masing $D^{(s)}$ dengan konsep berfikir hierarki clustering *complete linkage*.
 - i. Membentuk matrik CI antar variabel atribut
 - ii. Gabung dua kelompok terdekat berdasarkan nilai signifikansi CI yang terkecil
 - iii. Perbaharui matriks CI setelah penggabungan kelompok pada langkah 2.a.ii dimana jarak yang diambil adalah yang paling maksimum

- iv. Ulangi langkah 2.a.ii dan 2.a.iii sampai terbentuk satu kelompok.
- c. Penentuan variabel laten yang terbentuk pada masing-masing $D^{(s)}$ dengan cutting point $\alpha = 0,05$. Penentuan cutting point $\alpha = 0,05$ dimana, apabila nilai signifikansi sudah lebih dari 0,05 dianggap sudah independen.
- d. Penggabungan kategori pada masing-masing variabel laten yang terbentuk.
 - Penggabungan kategori dilakukan untuk pasangan variabel.
 - Apabila terdapat tiga atau lebih calon variabel laten, maka dilakukan penggabungan pada dua calon variabel laten (sepasang calon variabel laten) dahulu kemudian hasil penggabungan digabung lagi dengan calon variabel laten lain.
- e. Membangun model HNB dengan cara menggunakan *syntax* NB dengan input variabel atribut dan variabel laten yang terbentuk pada masing-masing $D^{(s)}$
- f. Menghitung akurasi dengan *Correct Clasification Rate* (CCR) dan *Area Under The Curve* (AUC) pada masing-masing struktur yang terbentuk
3. Pemilihan model HNB yang terbaik berdasarkan akurasi terbesar.
4. Menghitung nilai probabilitas prior $\hat{P}(Y)$
5. Menghitung nilai probabilitas independen kelas Y dari semua fitur dalam vektor $\prod_{j \in p} \hat{P}(X_j|Y)$
6. Menghitung nilai *posterior probability* untuk masing-masing klasifikasi $\hat{P}(Y|X)$

4.3.1 Independensi Antar Variabel Atribut

Untuk melihat hubungan independensi antara dua variabel dapat dilakukan uji *Chi-square* atau uji *Conditional Independence* (CI). Secara lebih jauh CI dapat digunakan untuk melihat independensi dari dua variabel dengan mempertimbangkan keberadaan satu variabel lainnya.

a. Hasil Uji *Chi-square*

Dalam penelitian ini digunakan sebelas variabel atribut karakteristik komuter Jabodetabek. Untuk melihat hubungan independensi antara dua variabel digunakan uji *Chi-square*. Hipotesis nol dalam uji *Chi-square* adalah dua variabel atribut saling independen, sedangkan hipotesis alternatifnya dua variabel atribut tidak saling independen.

Dalam penelitian ini diharapkan dapat mencakup sebanyak mungkin karakteristik pemilihan moda untuk menggambarkan transportasi di Jabodetabek. Namun, hal tersebut menyebabkan asumsi untuk uji *Chi-square* ini belum dapat terpenuhi karena masih terdapat 26,32 persen sel kontingensi yang memiliki nilai harapan dibawah 5 dimana batas maksimal yang diperbolehkan adalah 20 persen. Agresti (2007) menyebutkan dalam uji *Chi-Square* diperlukan sampel yang besar.

Tabel 4.8 Nilai *p-value* Hasil Uji *Chi-square* pada 11 Variabel Atribut

χ^2	1	2	3	4	5	6	7	8	9	10	11
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0969	0,0000	0,0042	0,0000	0,0000	0,1947
2		0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
3			0,0000	0,0000	0,0000	0,0091	0,0000	0,0000	0,0046	0,0000	0,0000
4				0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
5					0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
6						0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
7							0,0000	0,0000	0,0000	0,0000	0,0000
8								0,0000	0,0000	0,0000	0,0000
9									0,0000	0,0000	0,0000
10										0,0000	0,0000
11											0,0000

Keterangan:

1 : jenis kelamin	5 : kegiatan utama	9 : jumlah moda
2 : umur	6 : kota tujuan	10 : biaya transportasi
3 : status kawin	7 : jarak tempuh	11 : PDRB kota tujuan
4 : pendidikan	8 : waktu tempuh	

Hasil uji *Chi-square* pada Tabel 4.8 menunjukkan terdapat banyak nilai *p-value* yang sangat kecil. Jika menggunakan batas $\alpha = 0,05$, maka hanya terdapat satu nilai *p-value* yang lebih besar dari *alpha* yaitu pada indeks [1,6] dan [1,11] yang menunjukkan gagal tolak H_0 karena nilai *p-value* di atas *alpha* yang ditentukan. Sehingga dapat disimpulkan bahwa variabel ke-1 (jenis kelamin)

independen terhadap variabel ke-6 (kota tujuan) dan variabel ke-11 (PDRB kota tujuan), sedangkan variabel-variabel lainnya telah terjadi tidak saling bebas.

b. Hasil Uji *Conditional Independence*

Selain dengan uji *Chi-square* sebagaimana yang telah disajikan sebelumnya, untuk melihat hubungan independensi antara dua variabel digunakan uji conditional independence (CI). Hipotesis nol dalam uji CI adalah dua variabel atribut saling independen, sedangkan hipotesis alternatifnya dua variabel atribut tidak saling independen.

Tabel 4.9 Nilai *p-value* Hasil Uji *Conditional Independence* pada Variabel Atribut

χ^2	1	2	3	4	5	6	7	8	9	10	11
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0010	0,0000	0,0001	0,9992	0,0013	0,0057
2		0,0000	0,0000	0,0000	0,0000	0,0012	0,0000	0,0000	1,0000	0,0000	0,0013
3			0,0000	0,0000	0,0000	0,2017	0,0000	0,0000	0,7338	0,0000	0,0042
4				0,0000	0,0000	0,0000	0,0000	0,0000	0,0029	0,0000	0,0000
5					0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000
6						0,0000	0,0000	0,0000	0,7323	0,0070	0,0000
7							0,0000	0,0000	0,0842	0,0000	0,0000
8								0,0000	0,0000	0,0000	0,0000
9									0,0010	0,0000	0,0000
10										0,0000	0,5771
11											0,0146

Keterangan:

1 : jenis kelamin	5 : kegiatan utama	9 : jumlah moda
2 : umur	6 : kota tujuan	10 : biaya transportasi
3 : status kawin	7 : jarak tempuh	11 : PDRB kota tujuan
4 : pendidikan	8 : waktu tempuh	

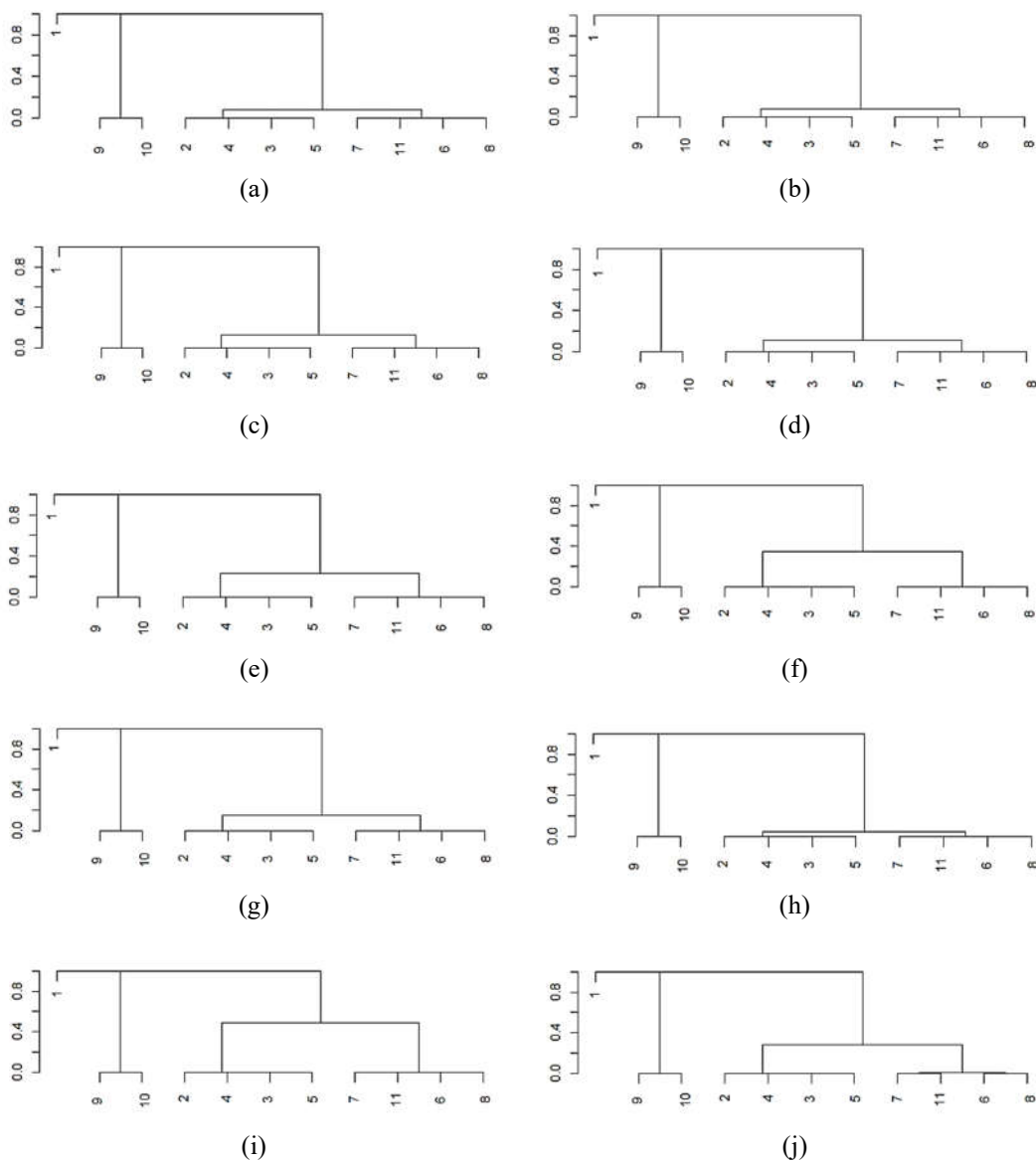
Uji CI ini mengadopsi uji *Chi-square* tetapi dilakukan untuk setiap kategori moda transportasi. Seperti pembahasan pada uji *Chi-square*, asumsi untuk uji CI ini belum dapat terpenuhi karena masih terdapat sel kontingensi yang memiliki nilai harapan dibawah 5 per kategori moda transportasinya dimana batas maksimal yang diperbolehkan adalah 20 persen. Hal ini dapat dilihat pada lampiran 3. Tidak terpenuhinya asumsi ini dapat mengganggu akurasi dari estimasi dan aplikasi penggunaan CI pada pembentukan laten.

Hasil uji *Conditional Independence* pada Tabel 4.9 menunjukkan terdapat banyak nilai *p-value* yang tidak signifikan. Jika menggunakan batas $\alpha = 0,05$, maka terdapat 8 nilai *p-value* yang lebih besar dari *alpha* yaitu pada indeks [1,9],

[2,9], [3,6], [3,9], [5,9], [6,9], [7,9], dan [10,11] yang menunjukkan gagal tolak H_0 karena nilai p -value di atas α yang ditentukan. Dapat disimpulkan bahwa variabel ke-1 (jenis kelamin) independen terhadap variabel ke-9 (jumlah moda) dengan syarat terdapat variabel klasifikasi (moda transportasi), variabel ke-2 (umur) independen terhadap variabel ke-9 (jumlah moda) dengan syarat terdapat variabel klasifikasi (moda transportasi), variabel ke-1 (jenis kelamin) independen terhadap variabel ke-5 (kota) dengan syarat terdapat variabel klasifikasi (moda transportasi), dst. Sementara variabel-variabel lainnya yang signifikan berarti terjadi tidak saling bebas.

4.3.2 Pembentukan Variabel Laten

Setelah uji independensi dilakukan untuk semua variabel atribut bersyarat variabel klasifikasi, maka dua variabel yang memiliki nilai p -value terkecil dipilih untuk dijadikan kandidat pembentuk variabel laten. Kategori variabel laten dibangun berdasarkan kombinasi seluruh kategori dari dua variabel pembentuknya. Dua variabel pembentuk laten dijadikan *childs* sekaligus variabel laten merupakan *parentnya*. Setelah terbentuk satu variabel laten, maka akan dicari variabel-variabel laten lainnya dengan langkah yang sama. Dalam pemodelan HNB untuk klasifikasi moda transportasi komuter, data *training* dibagi menjadi sepuluh data set secara random dan *overlapping*. Dalam algoritma HNB, kandidat laten terbaik dibangun dari dua variabel atribut dengan p -value dari uji *conditional independence* yang terkecil. Kandidat variabel laten kemudian dipilih berdasarkan struktur yang terbentuk. Berdasarkan hasil dari sepuluh set data *training*, struktur yang terbentuk hampir mirip dan menunjukkan struktur seperti pada Gambar 4.6. Hasil pencarian variabel laten dengan tingkat signifikansi 0,05 didapatkan tiga variabel laten yang terbentuk di dalam model, yaitu laten 1 (jumlah moda dan biaya transportasi), laten 2 (umur, status kawin, pendidikan dan kegiatan utama), serta laten 3 (kota tujuan, jarak tempuh, waktu tempuh dan PDRB kota tujuan).



Keterangan:

1 : jenis kelamin

2 : umur

3 : status kawin

4 : pendidikan

5 : kegiatan utama

6 : kota tujuan

7 : jarak tempuh

8 : waktu tempuh

9 : jumlah moda

10 : biaya transportasi

11 : PDRB kota tujuan

Gambar 4.6 Pembentukan Variabel Laten: (a) dan (b) 90 persen data *training*, (c) dan (d) 85 persen data *training*, (e) dan (f) 80 persen data *training*, (g) dan (h) 75 persen data *training*, (i) dan (j) 70 persen data *training*, (i) dan (j) 90 persen data *training*

Variabel laten 1 terbentuk dari variabel jumlah moda yang digunakan oleh komuter dan biaya transportasi. Jumlah pergantian moda dan biaya transportasi saling berhubungan, dimana semakin banyak moda transportasi yang digunakan

maka semakin banyak biaya transportasi yang harus dikeluarkan oleh komuter. Variabel laten 2 terbentuk dari variabel umur, status kawin, pendidikan dan kegiatan utama. Ketiga variabel tersebut memang saling mempengaruhi, misalnya bagi seorang pelajar dengan umur 10 tahun misalnya, pasti pendidikannya masih di bawah SMP dan kegiatan utamanya adalah sekolah/kursus dan belum menikah.

Namun hal tersebut tidak berlaku bagi kelompok umur 16-25 tahun dimana terjadi keberagaman kegiatan utama dan pendidikan. Masih terdapat komuter berusia 25 tahun yang kuliah, sementara terdapat komuter umur 16 tahun yang sudah bekerja dan bahkan mungkin sudah menikah. Variabel laten 3 terbentuk dari kota tujuan, jarak tempuh, waktu tempuh dan PDRB kota tujuan. Variabel-variabel ini menceminkan sarana dan prasarana transportasi menuju kota tujuan.

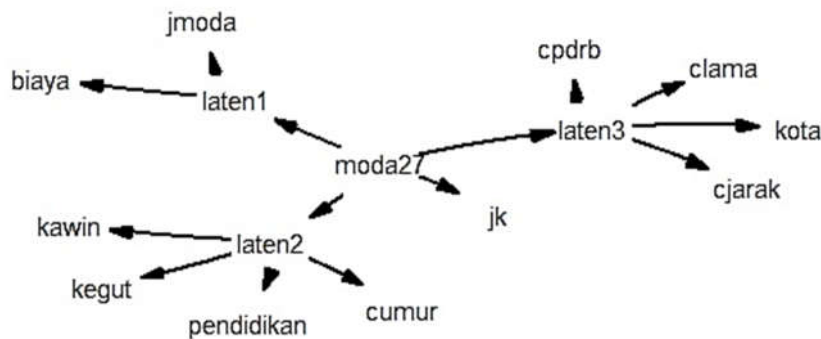
Variabel laten terbentuk dari variabel atribut/dependen yang berkorelasi kuat. Namun, dalam penelitian ini terdapat beberapa variabel pembentuk laten yang terbentuk bukan dari korelasi terbesarnya. Hal ini disebabkan karena belum terpenuhinya asumsi CI yang digunakan sebagai pembentuk laten yaitu masih terdapat sel kontingensi yang memiliki nilai harapan dibawah 5 per kategori moda transportasinya dimana batas maksimal yang diperbolehkan adalah 20 persen.

Uji independensi didapatkan dengan statistik uji *Chi-square* dimana hipotesis nol adalah dua variabel atribut saling independen, sedangkan hipotesis alternatifnya dua variabel atribut tidak saling independen. Uji independensi *Chi-square* diperoleh dengan menggunakan statistik uji χ^2 dengan membandingkan jumlah pada setiap sel dengan nilai harapan pada setiap sel. Saat jumlah pada setiap sel sama dengan nilai harapan, maka nilai χ^2 akan semakin kecil. Semakin besar perbedaan antara jumlah pada setiap sel dengan nilai harapannya, maka nilai χ^2 akan semakin besar dan semakin kuat untuk menolak hipotesis nol. Statistik χ^2 dapat didekati dengan distribusi *Chi-square* untuk nilai sampel yang besar. Pendekatan dengan distribusi *Chi-square* meningkat seiring dengan peningkatan nilai harapan. Nilai harapan 5 sudah cukup untuk layak untuk pendekatan (Agresti, 2007). Dengan tidak terpenuhinya asumsi tersebut, maka estimasi dan pemodelan pada model HNB belum reliabel karena sebenarnya statistik uji kurang tepat didekati dengan distribusi *Chi-square*.

4.3.3 Klasifikasi dengan Metode *Hierarchical Naïve Bayes*

a. Struktur *Hierarchical Naïve Bayes*

Pembentukan struktur *Hierarchical Naïve Bayes* berdasarkan seluruh atribut independen dan variabel laten bersyarat variabel klasifikasi. Dengan data *training* yang memiliki satu atribut independen, tiga variabel laten dan satu variabel klasifikasi maka struktur HNB yang terbentuk dapat dilihat pada Gambar 4.7.



Gambar 4.7 Struktur *Hierarchical Naïve Bayes* Klasifikasi Moda Transportasi

Berdasarkan struktur yang terbentuk pada Gambar 4.7 tersebut, kemudian dapat dibuat peluang bersama seperti pada persamaan (4.15) dengan melibatkan variabel atribut dan variabel laten.

$$\begin{aligned}
 P(Y|X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) = & P(Y) * P(X_1|Y) * P(\text{laten1}|Y) * \\
 & P(\text{laten2}|Y) * P(\text{laten3}|Y) * P(X_9|\text{laten1}) * P(X_{10}|\text{laten1}) * \\
 & P(X_2|\text{laten2}) * P(X_3|\text{laten2}) * P(X_4|\text{laten2}) * P(X_5|\text{laten2}) * \\
 & P(X_6|\text{laten3}) * P(X_7|\text{laten3}) * P(X_8|\text{laten3}) * P(X_{11}|\text{laten3}) * \quad (4.15)
 \end{aligned}$$

b. Estimasi Parameter Model *Hierarchical Naïve Bayes*

Persamaan persamaan (4.15) menunjukkan peluang bersama model HNB yang terbentuk. Langkah selanjutnya adalah melakukan estimasi parameter. Parameter dalam model *Bayesian Network* menggambarkan distribusi peluang sesuai dengan struktur yang terbentuk. Dengan metode MLE maka peluang bersyarat adalah proporsi kejadian pada kategori ke-*l* variabel ke-*j* dan kelas ke-*c*

atau dengan kata lain *cross* tabulasi antara variabel atribut dengan variabel klasifikasi yang diproporsikan terhadap kolom.

Dengan hadirnya variabel laten ke dalam model maka variabel klasifikasi menjadi independen terhadap dua variabel pembentuk laten. Sehingga yang diperlukan adalah estimasi terhadap parameter-parameter yang berkaitan langsung dengan variabel klasifikasi. Oleh karena itu peluang bersamanya menjadi seperti *Naïve Bayes* dengan tidak memasukkan variabel-variabel pembentuk laten, yaitu sebagai berikut:

$$P(Y|X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) = P(Y) * P(X_1|Y) * P(\text{laten1}|Y) * P(\text{laten2}|Y) * P(\text{laten3}|Y) \quad (4.16)$$

Persamaan (4.16) hanya melibatkan satu variabel atribut dan tiga variabel laten. Hasil estimasi parameter berdasarkan data *training* untuk peluang prior $\hat{P}(Y)$ HNB tersaji pada Tabel 4.10.

Tabel 4.10 Estimasi Peluang Prior $\hat{P}(Y)$ *Hierarchical Naïve Bayes*

$\hat{P}(Y=l)$							
$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=13$	$l=16$	$l=27$
(1)	(2)	(3)	(4)	(5)	(13)	(16)	(27)
0,0015	0,0222	0,0214	0,0031	0,1099	0,1268	0,5588	0,0010

Sedangkan hasil estimasi parameter berdasarkan data *training* untuk peluang bersyarat $\hat{P}(X|Y)$ HNB tersaji pada Tabel 4.11. Untuk lebih lengkapnya, hasil estimasi parameter HNB dapat dilihat pada lampiran 6.

Tabel 4.11 Estimasi Parameter $\hat{P}(X|Y)$ *Hierarchical Naïve Bayes*

$\hat{P}\left(X_j=h\middle Y=l\right)$			Y						
			$l=1$	$l=2$	$l=3$	$l=5$	$l=13$	$l=16$	$l=27$
(1)			(2)	(3)	(4)	(6)	(14)	(17)	(28)
X_1	Jenis kelamin	$h=1$	0,0000	0,4261	0,4955	0,3965	0,7006	0,7961	0,6000
		$h=2$	0,0000	0,5739	0,5045	0,6035	0,2994	0,2039	0,4000
L_1	Laten 1	$h=1$	0,0000	0,0696	0,1351	0,2105	0,0213	0,1983	1,0000
		$h=2$	0,0000	0,1391	0,4324	0,1158	0,0243	0,4981	0,0000
		$h=3$	0,0000	0,0174	0,0901	0,0263	0,0380	0,2021	0,0000
		$h=4$	0,0000	0,0000	0,1171	0,0368	0,0760	0,0673	0,0000
		$h=5$	0,0000	0,0000	0,0721	0,0211	0,8267	0,0290	0,0000

Lanjutan Tabel 4.11

$\hat{P}(X_j = h Y = l)$		Y						
		$l = 1$	$l = 2$	$l = 3$	$l = 5$	$l = 13$	$l = 16$	$l = 27$
	(1)	(2)	(3)	(4)	(6)	(14)	(17)	(28)
L_2 Laten 2	$h=6$	0,0000	0,0174	0,0000	0,0175	0,0015	0,0003	0,0000
	$h=7$	0,0000	0,1565	0,0360	0,1982	0,0015	0,0024	0,0000
	$h=8$	0,0000	0,3826	0,0090	0,1070	0,1667	0,0010	0,0000
	$h=9$	0,1250	0,0348	0,0360	0,0561	0,0000	0,0003	0,0000
	\vdots							
	$h=1$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$h=2$	0,0000	0,0000	0,2793	0,1404	0,0942	0,0466	0,4000
	$h=3$	0,0000	0,0000	0,0000	0,0018	0,0000	0,0000	0,0000
	$h=4$	0,0000	0,0000	0,0000	0,0456	0,0000	0,0003	0,0000
	\vdots							
L_3 Laten 3	$h=1$	0,0000	0,0000	0,0180	0,0105	0,0076	0,0169	0,0000
	$h=2$	0,0000	0,0087	0,0360	0,0526	0,0319	0,0580	0,8000
	$h=3$	0,0000	0,0261	0,0000	0,0263	0,0076	0,0121	0,0000
	$h=4$	0,0000	0,0966	0,0450	0,1088	0,0669	0,0576	0,0000
	\vdots							

Hasil dari estimasi parameter dapat menjelaskan kecenderungan komuter memilih moda transportasi apabila nilai variabel atribut diketahui seperti pada model NB. Tiga peluang prior terbesar yaitu sepeda motor ($Y=16$), mobil ($Y=13$), dan kendaraan umum dengan rute ($Y=5$). Komuter dengan jenis kelamin laki-laki cenderung memilih menggunakan sepeda motor dibandingkan komuter yang berjenis kelamin perempuan $\hat{P}(X_1 = 1 | Y = 16) > \hat{P}(X_1 = 2 | Y = 16)$, dimana diketahui dari bahwa $\hat{P}(X_1 = 1 | Y = 16) = 0,7961$ dan $\hat{P}(X_1 = 2 | Y = 16) = 0,2039$. Moda transportasi mobil juga cenderung dipilih oleh komuter dengan jenis kelamin laki-laki tetapi tidak demikian dengan moda transportasi kendaraan umum dengan rute yang cenderung diminati oleh komuter dengan jenis kelamin perempuan. Begitu pula interpretasi untuk variabel laten.

c. Kriteria Klasifikasi Model *Hierarchical Naïve Bayes*

Pada Model HNB, klasifikasi dilakukan dengan menghitung *posterior probability* $\hat{P}(X|Y)$ sesuai dengan struktur HNB dan estimasi parameter berdasarkan karakteristik komuter. Komuter terklasifikasi menggunakan moda kategori tertentu, misalnya dihitung nilai posterior untuk semua kategori (l) pada komuter dengan data atributnya, dan didapatkan peluang posterior yang

maksimum adalah peluang Y dengan kategori ke- k ($Y = y_k$), maka komuter tersebut diklasifikasikan sebagai y_k . Formula untuk mendapatkan peluang kategori maksimum pada persamaan(2.31). Estimasi peluang posterior HNB untuk $Y = 1$ (APTB), $Y = 2$ (bus Transjakarta), $Y = 5$ (kendaraan umum dengan rute), $Y = 13$ (mobil), $Y = 16$ (sepeda motor) dan $Y = 99$ (lainnya) adalah sebagai berikut:

$$\begin{aligned}
\hat{P}(Y = 1 | X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) &\propto (Y = 1) * P(X_1 | Y = 1) * \\
&\quad P(\text{laten1} | Y = 1) * P(\text{laten2} | Y = 1) * P(\text{laten3} | Y = 1) \\
\hat{P}(Y = 2 | X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) &\propto (Y = 2) * P(X_1 | Y = 2) * \\
&\quad P(\text{laten1} | Y = 2) * P(\text{laten2} | Y = 2) * P(\text{laten3} | Y = 2) \\
&\quad \vdots \\
\hat{P}(Y = 5 | X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) &\propto (Y = 5) * P(X_1 | Y = 5) * \\
&\quad P(\text{laten1} | Y = 5) * P(\text{laten2} | Y = 5) * P(\text{laten3} | Y = 5) \\
&\quad \vdots \\
\hat{P}(Y = 13 | X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) &\propto (Y = 13) * P(X_1 | Y = 13) * \\
&\quad P(\text{laten1} | Y = 13) * P(\text{laten2} | Y = 13) * P(\text{laten3} | Y = 13) \\
&\quad \vdots \\
\hat{P}(Y = 16 | X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) &\propto (Y = 16) * P(X_1 | Y = 16) * \\
&\quad P(\text{laten1} | Y = 16) * P(\text{laten2} | Y = 16) * P(\text{laten3} | Y = 16) \\
&\quad \vdots \\
\hat{P}(Y = 99 | X_1, X_2, \dots, X_{11}, \text{laten1}, \text{laten2}, \text{laten3}) &\propto (Y = 99) * P(X_1 | Y = 99) * \\
&\quad P(\text{laten1} | Y = 99) * P(\text{laten2} | Y = 99) * P(\text{laten3} | Y = 99)
\end{aligned}$$

d. Klasifikasi Model *Hierarchical Naïve Bayes* pada Data *Testing*

Berdasarkan model HNB yang telah terbentuk (struktur dan estimasi parameternya), klasifikasi dilakukan pada data *testing* yang berjumlah 576 komuter. Terdapat 172 komuter yang salah prediksi moda transportasi yang

digunakannya dan sebanyak 404 moda transportasi komuter yang diprediksi secara benar sehingga nilai *Correct Clasification Rate* (CCR) sebesar 70,1789 persen. Nilai *Area Under The Curve* (AUC) sebesar 68,3821 persen yang artinya metode HNB mengklasifikasikan dengan benar 68,3821 persen moda transportasi yang digunakan oleh komuter. *Syntax* R dan output yang digunakan untuk melakukan pemodelan klasifikasi dengan metode HNB dapat dilihat pada Lampiran 6 dan Lampiran 7.

4.4 Pemodelan dengan Regresi Logistik *Polytomous*

Analisis regresi logistik digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data kategorik dengan variabel bebas yang berupa data berskala interval dan atau kategorik. Jika variabel dependen memiliki dua kategori, maka kita gunakan analisis regresi logistik biner untuk memeriksa hubungan antara variabel dependen dan independennya. Sedangkan untuk variabel dependen yang memiliki kategori lebih dari dua, alat analisis yang dapat kita gunakan adalah Regresi Logistik *Polytomous* atau regresi logistik ordinal. Istilah variabel dependen pada Regresi Logistik *Polytomous* sepadan dengan istilah variabel atribut dalam *Naïve Bayes*.

Perbedaan antara Regresi Logistik *Polytomous* dan regresi logistik ordinal terletak pada skala variabel dependennya. Variabel dependen pada Regresi Logistik *Polytomous* berskala nominal, sedangkan variabel dependen pada regresi logistik ordinal berskala ordinal. Penelitian ini menggunakan analisis regresi *polytomous* untuk menjelaskan hubungan antara variabel dependen dan variabel independennya

Dalam analisis Regresi Logistik *Polytomous*, suatu variabel dependen dengan l kategori akan membentuk persamaan logit sebanyak $l-1$ dimana masing-masing persamaan ini membentuk regresi logistik biner yang membandingkan suatu kelompok kategori terhadap kategori acuan. Dalam penelitian ini, moda transportasi merupakan variabel dependen yang terdiri dari 27 kategori sehingga akan terbentuk 26 persamaan logit dengan kategori 1 (APT) sebagai kategori acuan.

4.4.1 Regresi Logistik *Polytomous* dengan Metode *Enter*

a. Pemodelan Regresi Logistik *Polytomous* dengan Metode *Enter*

Pemodelan Regresi Logistik *Polytomous* dengan menggunakan metode *enter* yaitu dengan memasukkan semua variabel dependen ke dalam analisis sekaligus. Model dibangun dengan melibatkan dua belas variabel, yaitu satu variabel dependen yaitu moda transportasi dan sebelas variabel dependen yaitu jenis kelamin, umur, status kawin, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, waktu tempuh, jumlah moda, biaya transportasi, dan PDRB kota tujuan. Variabel dependen (moda transportasi) terdiri dari 27 kategori yang dapat dilihat pada lampiran 1. Kategori terkecil menjadi *reference* kategori pada penelitian ini. Adapun persamaan logit yang diperoleh secara umum adalah sebagai berikut:

$$\begin{aligned} g_l(x) = & \beta_{l0} + \beta_{l1}X_1 + \beta_{l21}X_{21} + \beta_{l22}X_{22} + \beta_{l23}X_{23} + \beta_{l24}X_{24} + \beta_{l3}X_3 + \beta_{l41}X_{41} + \beta_{l42}X_{42} + \\ & \beta_{l51}X_{51} + \beta_{l52}X_{52} + \beta_{l23}X_{53} + \beta_{l6}X_6 + \beta_{l71}X_{71} + \beta_{l72}X_{72} + \beta_{l73}X_{73} + \beta_{l74}X_{74} + \\ & \beta_{l81}X_{81} + \beta_{l82}X_{82} + \beta_{l83}X_{83} + \beta_{l84}X_{84} + \beta_{l91}X_{91} + \beta_{l92}X_{92} + \beta_{l101}X_{101} + \beta_{l102}X_{102} + \\ & \beta_{l103}X_{103} + \beta_{l104}X_{104} + \beta_{l111}X_{111} + \beta_{l112}X_{112} \end{aligned} \quad (4.17)$$

b. Estimasi Parameter Model Regresi Logistik *Polytomous* dengan Metode *Enter*

Setelah diperoleh model logit secara umum seperti persamaan (4.17), langkah selanjutnya adalah melakukan estimasi parameter. Estimasi parameter dalam model Regresi Logistik *Polytomous* menggunakan MLE. Hasil dari estimasi parameter dari masing-masing nilai *beta* kemudian digunakan untuk menyusun persamaan logit. Persamaan logit untuk $Y = 2$ (bus Transjakarta), $Y = 5$ (kendaraan umum dengan rute), $Y = 13$ (mobil), $Y = 16$ (sepeda motor) dan $Y = 27$ (lainnya) dengan base kategori $Y = 1$ (APTB) adalah sebagai berikut:

$$\begin{aligned}
\hat{g}_2(x) &= 0,1922 - 0,2451X_1 + 6,9968X_{21} + 6,1108X_{22} + 6,8767X_{23} + 6,7756X_{24} + \\
&\quad + 0,1648X_3 - 0,6864X_{41} + 0,5837X_{42} - 1,4232X_{51} - 2,5230X_{52} + 0,6989X_{53} + \\
&\quad - 5,4173X_6 + 1,6930X_{71} - 0,0816X_{72} - 0,2119X_{73} - 2,2385X_{74} + 1,0451X_{81} + \\
&\quad - 0,0617X_{82} - 1,0326X_{83} - 0,2910X_{84} - 0,6922X_{91} - 0,7313X_{92} + 0,9298X_{101} + \\
&\quad + 1,7175X_{102} - 0,8115X_{103} - 3,6622X_{104} + 1,1785X_{111} + 1,1293X_{112} \\
\hat{g}_5(x) &= 9,7961 + 0,2493X_1 + 0,7061X_{21} - 0,5636X_{22} + 0,2858X_{23} + 0,5232X_{24} + \\
&\quad + 0,6451X_3 - 1,5528X_{41} - 0,6765X_{42} - 0,0441X_{51} - 2,1121X_{52} + 0,8939X_{53} + \\
&\quad + 1,4036X_6 + 0,9336X_{71} - 0,7700X_{72} - 1,0342X_{73} - 1,9667X_{74} + 1,0326X_{81} + \\
&\quad - 0,0175X_{82} - 1,0377X_{83} - 0,3575X_{84} - 1,2787X_{91} - 0,7913X_{92} + 0,9488X_{101} + \\
&\quad + 0,9563X_{102} - 0,3780X_{103} - 3,0647X_{104} + 0,1357X_{111} - 0,0533X_{112} \\
\hat{g}_{13}(x) &= 10,0851 - 0,7706X_1 - 0,8723X_{21} - 1,3694X_{22} - 0,0658X_{23} - 0,0013X_{24} + \\
&\quad + 1,3688X_3 - 1,1559X_{41} + 1,6859X_{42} - 0,7585X_{51} - 3,8088X_{52} + 1,4009X_{53} + \\
&\quad - 0,0215X_6 + 0,8225X_{71} - 0,9009X_{72} - 1,1237X_{73} - 2,3899X_{74} - 0,6423X_{81} + \\
&\quad - 1,7565X_{82} - 2,5534X_{83} - 2,0077X_{84} - 7,0781X_{91} - 14,5586X_{92} + 2,2292X_{101} + \\
&\quad + 4,3386X_{102} + 3,6084X_{103} + 3,9899X_{104} - 0,4873X_{111} - 1,2024X_{112} \\
\hat{g}_{16}(x) &= 12,5136 - 1,5139X_1 + 1,0623X_{21} - 2,2068X_{22} + 0,0388X_{23} - 0,3998X_{24} + \\
&\quad + 1,5442X_3 - 0,4626X_{41} + 0,7001X_{42} + 0,3254X_{51} - 2,4252X_{52} + 0,6636X_{53} + \\
&\quad - 0,7456X_6 + 1,5142X_{71} + 0,3412X_{72} + 0,2316X_{73} - 0,4969X_{74} - 0,0125X_{81} + \\
&\quad - 1,2595X_{82} - 2,1867X_{83} - 2,0016X_{84} - 7,5508X_{91} - 22,8612X_{92} + 1,9190X_{101} + \\
&\quad + 2,3044X_{102} - 0,4371X_{103} - 3,5134X_{104} - 0,3383X_{111} - 0,6809X_{112} \\
\hat{g}_{27}(x) &= 8,0560 + 0,5056X_1 - 2,6905X_{21} - 2,1362X_{22} + 0,9558X_{23} + 1,6727X_{24} + \\
&\quad + 0,1269X_3 - 0,8977X_{41} - 2,7164X_{42} - 3,9550X_{51} - 4,9093X_{52} - 2,2743X_{53} + \\
&\quad + 2,9611X_6 + 0,1029X_{71} - 0,7733X_{72} - 0,4603X_{73} - 2,2687X_{74} - 4,0192X_{81} + \\
&\quad - 3,7943X_{82} - 3,8578X_{83} - 4,6433X_{84} - 4,1559X_{91} - 2,0942X_{92} - 2,0967X_{101} + \\
&\quad - 0,6863X_{102} - 2,3523X_{103} - 4,8377X_{104} - 2,6908X_{111} + 2,7833X_{112}
\end{aligned}$$

Hasil estimasi parameter berdasarkan data *training* untuk Regresi Logistik *Polytomous* dengan metode *enter* selengkapnya tersaji pada lampiran 9. Berdasarkan persamaan diatas dapat pula dihitung nilai *odds ratio*. *Odds ratio* merupakan ukuran untuk mengetahui risiko kecenderungan untuk mengalami suatu kejadian tertentu antara kategori yang satu dengan kategori yang lain dalam suatu variabel. Perbandingan antar kategori dapat dilakukan dengan menggunakan nilai odds rasio. *Odds ratio* merupakan hasil pengitungan bilangan natural (e) yang dipangkatkan dengan nilai koefisien. Nilai *odds ratio* memberikan gambaran

seberapa besar suatu kategori variabel dependen memiliki kecenderungan untuk mengklasifikasikan moda transportasi yang digunakan oleh komuter dibandingkan dengan *reference* kategori.

Interpretasi *odds ratio* dilakukan berdasarkan pada model dengan variabel dependen yang atribut signifikan dikarenakan pada model yang tidak signifikan sering menghasilkan nilai *odds* yang tidak interpretable (interpretasinya tidak sesuai dengan hubungan antar variabel yang sesungguhnya). Salah satunya adalah karena faktor multikolinieritas. Oleh karena itu interpretasi *odds ratio* dilakukan pada subbab akan dilakukan pada Regresi Logistik *Polytomous* dengan menggunakan metode *stepwise*.

c. Kriteria Klasifikasi Model Regresi Logistik *Polytomous* dengan Metode *Enter*

Klasifikasi dapat dilakukan dengan menghitung probabilitas bersyarat untuk model *polytomous* sesuai dengan model logit yang terbentuk. Penghitungan probabilitas bersyarat untuk model *polytomous* adalah seperti pada persamaan (2.45) sehingga diperoleh:

$$\begin{aligned}\hat{\pi}_1(x) &= \hat{P}(Y=1|x) = \frac{1}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ \hat{\pi}_2(x) &= \hat{P}(Y=2|x) = \frac{\exp(g_2(x))}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ &\vdots \\ \hat{\pi}_5(x) &= \hat{P}(Y=5|x) = \frac{\exp(g_5(x))}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ &\vdots \\ \hat{\pi}_{13}(x) &= \hat{P}(Y=13|x) = \frac{\exp(g_{13}(x))}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ &\vdots\end{aligned}$$

$$\hat{\pi}_{27}(x) = \hat{P}(Y=27|x) = \frac{\exp(g_{27}(x))}{1 + \sum_{l=2}^k \exp(g_l(x))}$$

Untuk mengetahui klasifikasi moda transportasi komuter, langkah pertama yaitu memasukkan variabel dependen ke dalam 27 persamaan probabilitas bersyarat di atas. Komuter terklasifikasi menggunakan moda kategori tertentu dengan nilai probabilitas bersyarat terbesar. Misalnya didapatkan peluang bersyarat maksimum adalah peluang Y dengan kategori ke- k ($Y=y_k$), maka komuter tersebut diklasifikasikan sebagai \mathcal{Y}_k .

d. Klasifikasi Model Regresi Logistik *Polytomous* dengan Metode *Enter* pada *Data Testing*

Berdasarkan model Regresi Logistik *Polytomous* yang telah terbentuk, klasifikasi dilakukan pada data *testing* yang berjumlah 576 komuter. Terdapat 137 komuter yang salah prediksi moda transportasi yang digunakannya dan sebanyak 439 moda transportasi komuter yang diprediksi secara benar sehingga nilai *Correct Clasification Rate* (CCR) sebesar 76,2153 persen. Nilai *Area Under The Curve* (AUC) sebesar 71,3496 persen yang artinya metode model Regresi Logistik *Polytomous* dengan metode *enter* mengklasifikasikan dengan benar 71,3496persen moda transportasi yang digunakan oleh komuter. *Syntax* R dan output yang digunakan untuk melakukan pemodelan klasifikasi dengan metode Regresi Logistik *Polytomous* dapat dilihat pada Lampiran 8 dan Lampiran 9.

4.4.2 Regresi Logistik *Polytomous* dengan Metode *Stepwise*

Untuk mengurangi kemungkinan adanya multikolinearitas dari persamaan/model yang dihasilkan, maka digunakan seleksi variabel. Dalam penelitian ini digunakan metode *stepwise*. Metode *stepwise* melibatkan dua jenis proses yaitu: *forward selection* dan *backward elimination*. Metode ini dilakukan melalui beberapa tahapan. Pada masing-masing tahapan, akan diputuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model. Variabel akan ditambahkan ke dalam model selama nilai *p-valuenya* kurang dari

nilai kritis alpha (α). Kemudian variabel dengan nilai *p-value* lebih dari nilai kritis alpha (α) akan dihilangkan. Proses ini dilakukan terus menerus hingga tidak ada lagi variabel yang memenuhi kriteria untuk ditambahkan atau dihilangkan.

Model dibangun dengan melibatkan dua belas variabel, yaitu satu variabel dependen yaitu moda transportasi dan sebelas variabel dependen yaitu jenis kelamin, umur, status kawin, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, waktu tempuh, jumlah moda, biaya transportasi, dan PDRB kota tujuan. Setelah dilakukan seleksi variabel dengan metode *stepwise*, terdapat tiga variabel yang dikeluarkan dari model yaitu variabel status kawin, waktu tempuh, dan PDRB kota tujuan dengan pengurangan nilai AIC seperti pada Tabel 4.12

Tabel 4.12 Perbandingan AIC pada Regresi Logistik *Polytomous*

Metode	Jumlah Variabel Dependen	AIC
(1)	(2)	(3)
<i>Enter</i>	11	10.354,39
<i>Stepwise</i>	8	10.278,29

a. Pemodelan Regresi Logistik *Polytomous* dengan Metode *Stepwise*

Berdasarkan hasil Regresi Logistik *Polytomous* dengan menggunakan metode *stepwise*, maka diperoleh model logit dengan variabel dependen adalah moda transportasi serta variabel independennya adalah jenis kelamin, umur, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, jumlah moda, dan biaya transportasi. Adapun persamaan logit yang diperoleh secara umum adalah sebagai berikut:

$$g_i(x) = \beta_{i0} + \beta_{i1}X_1 + \beta_{i21}X_{21} + \beta_{i22}X_{22} + \beta_{i23}X_{23} + \beta_{i24}X_{24} + \beta_{i41}X_{41} + \beta_{i42}X_{42} + \beta_{i51}X_{51} + \beta_{i52}X_{52} + \beta_{i23}X_{53} + \beta_{i6}X_6 + \beta_{i71}X_{71} + \beta_{i72}X_{72} + \beta_{i73}X_{73} + \beta_{i74}X_{74} + \beta_{i91}X_{91} + \beta_{i92}X_{92} + \beta_{i101}X_{101} + \beta_{i102}X_{102} + \beta_{i103}X_{103} + \beta_{i104}X_{104} \quad (4.18)$$

b. Estimasi Parameter Model Regresi Logistik *Polytomous* dengan Metode *Stepwise*

Setelah diperoleh model logit secara umum seperti persamaan (4.18), langkah selanjutnya adalah melakukan estimasi parameter. Estimasi parameter

dalam model Regresi Logistik *Polytomous* menggunakan MLE. Hasil dari estimasi parameter dari masing-masing nilai *beta* kemudian digunakan untuk menyusun persamaan logit. Hasil dari estimasi parameter dari masing-masing nilai *beta* kemudian digunakan untuk menyusun persamaan logit. Persamaan logit untuk $Y = 2$ (bus Transjakarta), $Y = 5$ (kendaraan umum dengan rute), $Y = 13$ (mobil), $Y = 16$ (sepeda motor) dan $Y = 27$ (lainnya) dengan base kategori $Y = 1$ (APT) adalah sebagai berikut:

$$\begin{aligned}\hat{g}_2(x) = & 1,8895 - 0,2897X_1 + 8,9043X_{21} + 7,7898X_{22} + 8,6805X_{23} + 8,3448X_{24} + \\ & - 0,6556X_{41} + 0,6570X_{42} - 0,8737X_{51} - 2,5444X_{52} + 0,8244X_{53} + \\ & - 8,6864X_6 + 2,0488X_{71} - 0,3220X_{72} - 0,6673X_{73} - 2,8134X_{74} - 1,8858X_{91} \\ & - 2,5191X_{92} + 1,3055X_{101} + 1,8689X_{102} - 1,1837X_{103} - 4,2938X_{104}\end{aligned}$$

$$\begin{aligned}\hat{g}_5(x) = & 12,4835 + 0,1658X_1 + 0,2204X_{21} - 1,0969X_{22} + 0,1524X_{23} + 0,1669X_{24} + \\ & - 1,5595X_{41} - 0,7111X_{42} + 0,7581X_{51} - 2,0115X_{52} + 1,0410X_{53} + \\ & + 1,6384X_6 + 1,3225X_{71} - 1,0287X_{72} - 1,4222X_{73} - 2,4458X_{74} - 2,3940X_{91} \\ & - 2,4036X_{92} + 1,3700X_{101} + 1,1876X_{102} - 0,7274X_{103} - 3,6989X_{104}\end{aligned}$$

$$\begin{aligned}\hat{g}_{13}(x) = & 11,4999 - 0,8615X_1 - 1,4325X_{21} - 1,6470X_{22} + 0,1330X_{23} + 0,0283X_{24} + \\ & - 1,1688X_{41} + 1,5680X_{42} + 0,0145X_{51} - 3,6454X_{52} + 1,3797X_{53} + \\ & + 1,1926X_6 + 0,9404X_{71} - 1,4031X_{72} - 1,8260X_{73} - 2,9991X_{74} - 7,9504X_{91} \\ & - 31,4349X_{92} + 2,2480X_{101} + 3,8555X_{102} + 2,5535X_{103} + 2,4208X_{104}\end{aligned}$$

$$\begin{aligned}\hat{g}_{16}(x) = & 14,3330 - 1,5365X_1 + 0,4730X_{21} - 0,3719X_{22} + 0,5038X_{23} - 0,0896X_{24} + \\ & - 0,5331X_{41} + 0,5761X_{42} + 1,0801X_{51} - 2,3495X_{52} + 0,6907X_{53} + \\ & + 1,5674X_6 + 1,6309X_{71} - 0,2780X_{72} - 0,5944X_{73} - 1,4189X_{74} - 8,2083X_{91} \\ & - 9,2296X_{92} + 2,1012X_{101} + 2,2135X_{102} - 1,0104X_{103} - 4,4046X_{104}\end{aligned}$$

$$\begin{aligned}\hat{g}_{27}(x) = & 11,6494 + 0,2241X_1 - 3,7151X_{21} - 3,7917X_{22} + 0,8541X_{23} + 1,5229X_{24} + \\ & - 1,1172X_{41} - 2,9470X_{42} - 5,7035X_{51} - 4,6661X_{52} - 0,8532X_{53} + \\ & + 0,7738X_6 - 2,1664X_{71} - 3,1224X_{72} - 3,1976X_{73} - 6,5885X_{74} - 5,0939X_{91} \\ & - 3,1985X_{92} - 2,8822X_{101} - 2,2271X_{102} - 4,4654X_{103} - 8,5143X_{104}\end{aligned}$$

Hasil estimasi parameter berdasarkan data *training* untuk Regresi Logistik *Polytomous* selengkapnya tersaji pada lampiran 8. Berikut ini merupakan contoh nilai *odds ratio* pada moda transportasi kategori 16 yaitu sepeda motor apabila

dibandingkan dengan *base* kategori yaitu APTB ($Y=1$). *Odds ratio* merupakan hasil penghitungan bilangan natural (e) yang dipangkatkan dengan nilai koefisien.

Berdasarkan metode *Bayesian Network* sebelumnya, dimana peluang posterior terbesar adalah untuk moda sepeda motor ($Y=16$), mobil ($Y=13$) dan kendaraan umum dengan rute ($Y=5$), maka interpretasi odds ratio untuk model logit $\hat{g}_{16}(x)$ yaitu sepeda motor dibandingkan base kategorinya APTB adalah sebagai berikut:

- Kecenderungan komuter dengan jenis kelamin perempuan untuk memilih moda sepeda motor daripada APTB adalah 0,2151 kali dibandingkan komuter yang berjenis kelamin laki-laki, atau dapat dikatakan bahwa kecenderungan komuter dengan jenis kelamin laki-laki untuk memilih moda sepeda motor daripada APTB adalah 4,6490 kali dibandingkan komuter yang berjenis kelamin perempuan.
- Komuter dengan kelompok umur 16-25 tahun 1,6048 kali cenderung untuk memilih moda sepeda motor daripada APTB dibandingkan komuter dengan kelompok umur kurang dari 16 tahun. Komuter dengan kelompok umur 26-35 tahun 0,6894 kali cenderung untuk memilih moda sepeda motor daripada APTB dibandingkan komuter dengan kelompok umur kurang dari 16 tahun. Komuter dengan kelompok umur 36-45 tahun 1,655 kali cenderung untuk memilih moda sepeda motor daripada APTB dibandingkan komuter dengan kelompok umur kurang dari 16 tahun. Komuter dengan kelompok umur lebih dari 45 tahun 0,9143 kali cenderung untuk memilih moda sepeda motor daripada APTB dibandingkan komuter dengan kelompok umur kurang dari 16 tahun.
- Komuter dengan pendidikan SMA/ sederajat 0,5867 kali cenderung untuk memilih moda sepeda motor daripada APTB dibandingkan komuter yang berpendidikan SMP/ sederajat ke bawah. Namun, komuter lulusan perguruan tinggi yang cenderung 1,7792 kali untuk memilih moda sepeda motor daripada APTB dibandingkan komuter yang berpendidikan SMP/ sederajat ke bawah.
- Komuter dengan kegiatan utama PNS lebih cenderung 2,9451 kali memilih moda sepeda motor daripada APTB dibandingkan pengusaha. Komuter dengan

kegiatan utama karyawan swasta/ pekerja bebas lebih cenderung 0,0954 kali memilih moda sepeda motor daripada APTB dibandingkan pengusaha. Komuter dengan kegiatan utama sekolah/ kursus 1,9950 kali lebih cenderung memilih moda sepeda motor daripada APTB dibandingkan pengusaha.

- Komuter dengan kota tujuan di luar DKI Jakarta 2,9451 kali lebih cenderung memilih menggunakan moda sepeda motor daripada APTB dibandingkan komuter yang kota tujuannya DKI Jakarta.
- Untuk variabel jarak, komuter dengan jarak tempuh 10 km atau lebih cenderung menggunakan sepeda motor dari pada sepeda dibandingkan dengan rentang jarak tempuh dibawah 10 km. Komuter dengan rentang jarak lebih antara 10 – 19 km cenderung 5,1086 kali menggunakan sepeda motor daripada APTB bila dibandingkan dengan rentang jarak dibawah 10 km. Komuter dengan rentang jarak antara 20 – 29 km cenderung 7,5727 kali menggunakan sepeda motor daripada APTB bila dibandingkan dengan rentang jarak dibawah 10 km. Namun, komuter dengan rentang jarak tempuh lebih dari 39 km cenderung 0,5518 kali menggunakan sepeda motor daripada APTB bila dibandingkan dengan rentang jarak dibawah 10 km. Artinya komuter jarak jauh lebih cenderung menggunakan APTB daripada sepeda motor
- Dengan rentang biaya transportasi Rp 6.001,00 – Rp 12.000,00, komuter 8,1756 kali cenderung menggunakan sepeda motor daripada APTB bila dibandingkan rentang biaya transportasi dibawah Rp 6.000,00. Dengan rentang biaya transportasi Rp 12.001,00 – Rp 18.000,00, komuter 9,1478 kali cenderung menggunakan sepeda motor daripada APTB bila dibandingkan rentang biaya transportasi dibawah Rp 6.000,00. Begitu pula dengan rentang biaya Rp 18.001,00 – Rp 24.000,00, komuter 3,6408 kali cenderung menggunakan sepeda motor daripada APTB bila dibandingkan rentang biaya transportasi dibawah Rp 6.000,00. Selanjutnya, dengan rentang biaya lebih dari Rp 24.000,00, komuter 0,0122 kali cenderung menggunakan sepeda motor daripada APTB bila dibandingkan rentang biaya transportasi dibawah Rp 6.000,00.

Demikian pula cara interpretasi untuk moda mobil, kendaraan umum dengan rute dan moda lainnya.

c. Kriteria Klasifikasi Model Regresi Logistik *Polytomous* dengan Metode *Stepwise*

Klasifikasi dapat dilakukan dengan menghitung probabilitas bersyarat untuk model *polytomous* sesuai dengan model logit yang terbentuk. Penghitungan probabilitas bersyarat untuk model *polytomous* adalah seperti pada persamaan (2.45) sehingga diperoleh:

$$\begin{aligned}\hat{\pi}_1(x) &= \hat{P}(Y=1|x) = \frac{1}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ \hat{\pi}_2(x) &= \hat{P}(Y=2|x) = \frac{\exp(g_2(x))}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ &\vdots \\ \hat{\pi}_5(x) &= \hat{P}(Y=5|x) = \frac{\exp(g_5(x))}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ &\vdots \\ \hat{\pi}_{13}(x) &= \hat{P}(Y=13|x) = \frac{\exp(g_{13}(x))}{1 + \sum_{l=2}^k \exp(g_l(x))} \\ &\vdots \\ \hat{\pi}_{27}(x) &= \hat{P}(Y=27|x) = \frac{\exp(g_{27}(x))}{1 + \sum_{l=2}^k \exp(g_l(x))}\end{aligned}$$

Untuk mengetahui klasifikasi moda transportasi komuter, langkah pertama yaitu memasukkan variabel dependen ke dalam 27 persamaan probabilitas bersyarat di atas. Komuter terklasifikasi menggunakan moda kategori tertentu dengan nilai probabilitas bersyarat terbesar. Misalnya didapatkan peluang

bersyarat maksimum adalah peluang Y dengan kategori ke- k ($Y = y_k$), maka komuter tersebut diklasifikasikan sebagai y_k .

d. Klasifikasi Model Regresi Logistik *Polytomous* dengan Metode *Stepwise* pada Data *Testing*

Berdasarkan model Regresi Logistik *Polytomous* yang telah terbentuk, klasifikasi dilakukan pada data *testing* yang berjumlah 576 komuter. Terdapat 149 komuter yang salah prediksi moda transportasi yang digunakannya dan sebanyak 427 moda transportasi komuter yang diprediksi secara benar sehingga nilai *Correct Clasification Rate* (CCR) sebesar 74,1319 persen. Nilai *Area Under The Curve* (AUC) sebesar 68,3486 persen yang artinya metode model Regresi Logistik *Polytomous* dengan metode *stepwise* mengklasifikasikan dengan benar 68,3486 persen moda transportasi yang digunakan oleh komuter. *Syntax* R dan output yang digunakan untuk melakukan pemodelan klasifikasi dengan metode Regresi Logistik *Polytomous* dapat dilihat pada Lampiran 8 dan Lampiran 9.

4.5 Pemodelan *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

4.5.1 Input Variabel pada Metode *Naïve Bayes*

Model *Naïve Bayes* dibangun dengan asumsi seluruh atribut saling independen bersyarat variabel klasifikasi. Hal ini dilakukan salah satunya dengan melakukan reduksi variabel dengan metode *stepwise* pada Regresi Logistik *Polytmous*. Berdasarkan hasil Regresi Logistik *Polytomous* dengan menggunakan metode *stepwise*, diperoleh variabel jenis kelamin, umur, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, jumlah moda, dan biaya transportasi. Variabel – variabel inilah yang kemudian menjadi input variabel pada metode NB.

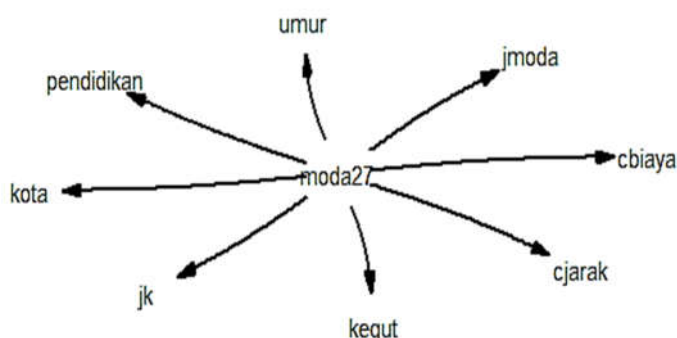
4.5.2 Klasifikasi *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

Model NB dibangun dengan asumsi seluruh atribut saling independen bersyarat variabel klasifikasi. Dalam *Bayesian Network*, secara umum terdapat

dua hal dalam pemodelan, yaitu membangun struktur dan mengestimasi parameter.

a. Struktur *Naïve Bayes*-- Input Variabel Hasil Regresi Logistik *Polytomous*

Struktur Regresi Logistik *Polytomous* -- NB dengan delapan variabel independen atau variabel atribut dan satu variabel klasifikasi dapat dilihat pada Gambar 4.8



Gambar 4.8 Struktur NB dengan Input Hasil Regresi Logistik *Polytomous*
Klasifikasi Moda Transportasi

Berdasarkan struktur yang terbentuk pada Gambar 4.8, kemudian dapat dibuat peluang bersama seperti persamaan (4.19).

$$P(Y|X_1, X_2, \dots, X_{11}) = P(Y) * P(X_1|Y) * P(X_2|Y) * P(X_4|Y) * P(X_5|Y) * P(X_6|Y) * P(X_7|Y) * P(X_9|Y) * P(X_{10}|Y) * \quad (4.19)$$

b. Estimasi Parameter *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

Persamaan (4.19) merupakan peluang bersama model *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous* yang terbentuk. Langkah selanjutnya adalah melakukan estimasi parameter. Hasil estimasi parameter berdasarkan data *training* untuk peluang prior $\hat{P}(Y)$ *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous* tersaji pada Tabel 4.13.

Tabel 4.13 Estimasi Peluang Prior $\hat{P}(Y)$ *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

$\hat{P}(Y = l)$							
$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 13$	$l = 16$	$l = 27$
(1)	(2)	(3)	(4)	(5)	(13)	(16)	(27)
0,0015	0,0222	0,0214	0,0031	0,1099	0,1268	0,5588	0,0010

Sedangkan hasil estimasi parameter berdasarkan data *training* untuk peluang bersyarat $\hat{P}(X|Y)$ *Naïve Bayes* -- Input Variabel Hasil Regresi Logistik *Polytomous* tersaji pada Tabel 4.14. Untuk lebih lengkapnya, hasil estimasi parameter Regresi Logistik *Polytomous* -- NB dapat dilihat pada lampiran 10.

Tabel 4.14 Estimasi Parameter $\hat{P}(X_j = h|Y = l)$ *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

$\hat{P}(X_j = h Y = l)$			Y							
			$l = 1$	$l = 2$	$l = 3$	$l = 5$	$l = 13$	$l = 16$	$l = 27$	
			(1)	(2)	(3)	(4)	(6)	(14)	(17)	(28)
X_1	Jenis kelamin	$h=1$	0,5000	0,4261	0,4955	0,3965	0,7006	0,7961	0,6000	
		$h=2$	0,5000	0,5739	0,5045	0,6035	0,2994	0,2039	0,4000	
X_2	Umur	$h=1$	0,0000	0,0000	0,2883	0,1439	0,0942	0,0469	0,4000	
		$h=2$	0,1250	0,3826	0,1622	0,3509	0,0851	0,2853	0,0000	
		$h=3$	0,3750	0,2957	0,1261	0,1719	0,1596	0,2984	0,0000	
		$h=4$	0,2500	0,1826	0,1982	0,1632	0,2857	0,2328	0,2000	
		$h=5$	0,2500	0,1391	0,2252	0,1702	0,3754	0,1366	0,4000	
X_4	Pendidikan	$h=1$	0,0000	0,1043	0,4685	0,3912	0,1489	0,2059	0,8000	
		$h=2$	0,5000	0,5130	0,2432	0,4158	0,1824	0,5612	0,2000	
		$h=3$	0,5000	0,3826	0,2883	0,1930	0,6687	0,2328	0,0000	
X_5	Kegiatan utama	$h=1$	0,0000	0,0261	0,0270	0,0333	0,1535	0,0566	0,4000	
		$h=2$	0,0000	0,0174	0,2432	0,0368	0,1170	0,0500	0,0000	
		$h=3$	1,0000	0,7739	0,4144	0,5982	0,5638	0,7330	0,2000	
		$h=4$	0,0000	0,1826	0,3153	0,3316	0,1656	0,1604	0,4000	
X_6	Kota tujuan	$h=1$	0,8750	1,0000	0,6216	0,6875	0,8085	0,7051	0,5000	
		$h=2$	0,1250	0,0000	0,3784	0,3125	0,1915	0,2949	0,5000	
X_7	Jarak tempuh	$h=1$	0,0000	0,1913	0,2523	0,3947	0,2006	0,2663	1,0000	
		$h=2$	0,0000	0,4087	0,2613	0,2737	0,2553	0,3139	0,0000	
		$h=3$	0,2500	0,2261	0,1622	0,1596	0,2644	0,2273	0,0000	
		$h=4$	0,2500	0,1478	0,1622	0,0982	0,1535	0,1221	0,0000	
		$h=5$	0,5000	0,0261	0,1622	0,0737	0,1261	0,0704	0,0000	
X_9	Jumlah moda	$h=1$	0,0000	0,2261	0,8468	0,4105	0,9863	0,9948	1,0000	
		$h=2$	0,6250	0,6348	0,1441	0,4263	0,0137	0,0045	0,0000	
		$h=3$	0,3750	0,1391	0,0090	0,1632	0,0000	0,0007	0,0000	
X_{10}	Biaya transportasi	$h=1$	0,0000	0,0870	0,1351	0,2281	0,0228	0,1987	1,0000	
		$h=2$	0,0000	0,2957	0,4685	0,3316	0,0258	0,5005	0,0000	
		$h=3$	0,0000	0,4087	0,0991	0,1667	0,0380	0,2039	0,0000	
		$h=4$	0,1250	0,1217	0,1532	0,1491	0,0775	0,0676	0,0000	
		$h=5$	0,8750	0,0870	0,1441	0,1246	0,8359	0,0293	0,0000	

Hasil dari estimasi parameter dapat menjelaskan kecenderungan komuter memilih moda transportasi apabila nilai variabel atribut diketahui. Tiga peluang prior terbesar yaitu sepeda motor ($Y=16$), mobil ($Y=13$), dan kendaraan umum dengan rute ($Y=5$). Interpretasi pada model ini sama seperti model NB yang sebelumnya telah dijelaskan. Misalkan komuter dengan jenis kelamin laki-laki cenderung memilih menggunakan sepeda motor dibandingkan komuter yang berjenis kelamin perempuan $\hat{P}(X_1=1|Y=16) > \hat{P}(X_1=2|Y=16)$, dimana diketahui dari Tabel 4.14 bahwa $\hat{P}(X_1=1|Y=16)=0,7961$ dan $\hat{P}(X_1=2|Y=16)=0,2039$. Moda transportasi mobil juga cenderung dipilih oleh komuter dengan jenis kelamin laki-laki tetapi tidak demikian dengan moda transportasi kendaraan umum dengan rute yang cenderung diminati oleh komuter dengan jenis kelamin perempuan.

c. Kriteria Klasifikasi Model *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

Pada model *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*, klasifikasi dilakukan dengan menghitung *posterior probability* $\hat{P}(X|Y)$ sesuai dengan struktur NB dan estimasi parameter berdasarkan karakteristik komuter. Komuter terklasifikasi menggunakan moda kategori tertentu, misalnya dihitung nilai posterior untuk semua kategori (l) pada komuter dengan data atributnya, dan didapatkan peluang posterior yang maksimum adalah peluang Y dengan kategori ke- k ($Y=y_k$), maka komuter tersebut diklasifikasikan sebagai Y_k . Formula untuk mendapatkan peluang kategori maksimum pada persamaan (2.23). Estimasi peluang posterior untuk $Y = 1$ (APTB), $Y = 2$ (bus Transjakarta), $Y = 5$ (kendaraan umum dengan rute), $Y = 13$ (mobil), $Y = 16$ (sepeda motor) dan $Y = 99$ (lainnya) adalah sebagai berikut:

$$\begin{aligned}
\hat{P}(Y=1|X_1, X_2, \dots, X_{11}) &\propto P(Y=1) * P(X_1|Y=1) * P(X_2|Y=1) * P(X_4|Y=1) * \\
&\quad P(X_5|Y=1) * P(X_6|Y=1) * P(X_7|Y=1) * P(X_9|Y=1) * \\
&\quad P(X_{10}|Y=1) \\
\hat{P}(Y=2|X_1, X_2, \dots, X_{11}) &\propto P(Y=2) * P(X_1|Y=2) * P(X_2|Y=2) * P(X_4|Y=2) * \\
&\quad P(X_5|Y=2) * P(X_6|Y=2) * P(X_7|Y=2) * P(X_9|Y=2) * \\
&\quad P(X_{10}|Y=2) \\
&\quad \vdots \\
\hat{P}(Y=5|X_1, X_2, \dots, X_{11}) &\propto P(Y=5) * P(X_1|Y=5) * P(X_2|Y=5) * P(X_4|Y=5) * \\
&\quad P(X_5|Y=5) * P(X_6|Y=5) * P(X_7|Y=5) * P(X_9|Y=5) * \\
&\quad P(X_{10}|Y=5) \\
&\quad \vdots \\
\hat{P}(Y=13|X_1, X_2, \dots, X_{11}) &\propto P(Y=13) * P(X_1|Y=13) * P(X_2|Y=13) * P(X_4|Y=13) * \\
&\quad P(X_5|Y=13) * P(X_6|Y=13) * P(X_7|Y=13) * P(X_9|Y=13) * \\
&\quad P(X_{10}|Y=13) \\
&\quad \vdots \\
\hat{P}(Y=16|X_1, X_2, \dots, X_{11}) &\propto P(Y=16) * P(X_1|Y=16) * P(X_2|Y=16) * P(X_4|Y=16) * \\
&\quad P(X_5|Y=16) * P(X_6|Y=16) * P(X_7|Y=16) * P(X_9|Y=16) * \\
&\quad P(X_{10}|Y=16) \\
&\quad \vdots \\
\hat{P}(Y=99|X_1, X_2, \dots, X_{11}) &\propto P(Y=99) * P(X_1|Y=99) * P(X_2|Y=99) * P(X_4|Y=99) * \\
&\quad P(X_5|Y=99) * P(X_6|Y=99) * P(X_7|Y=99) * P(X_9|Y=99) * \\
&\quad P(X_{10}|Y=99)
\end{aligned}$$

d. Klasifikasi Model *Naïve Bayes* dengan Input Variabel Hasil Regresi Logistik *Polytomous*

Berdasarkan model *Naïve Bayes* -- Input Variabel Hasil Regresi Logistik *Polytomous* yang telah terbentuk (struktur dan estimasi parameternya), klasifikasi dilakukan pada data *testing* yang berjumlah 576 komuter. Terdapat 183 komuter yang salah prediksi moda transportasi yang digunakannya dan sebanyak 393 moda transportasi komuter yang diprediksi secara benar sehingga nilai *Correct Clasification Rate* (CCR) sebesar 68,2292 persen. Nilai *Area Under The Curve*

(AUC) sebesar 65,6655 persen yang artinya metode *Naïve Bayes* -- Input Variabel Hasil Regresi Logistik *Polytomous* mengklasifikasikan dengan benar 65,6655 persen moda transportasi yang digunakan oleh komuter. *Syntax* R dan output yang digunakan untuk melakukan pemodelan klasifikasi dengan metode NB dapat dilihat pada Lampiran 10 dan Lampiran 11.

4.6 Perbandingan Hasil Klasifikasi

Pengukuran *Correct Clasification Rate* (CCR) dan *Area Under The Curve* (AUC) pada masing- masing metode dilakukan untuk mengukur kebaikan model. Tabel 4.15 menunjukkan perbandingan nilai CCR dan AUC untuk klasifikasi pada data *training*, sedangkan Tabel 4.16 menunjukkan perbandingan nilai CCR dan AUC untuk klasifikasi pada data *testing*.

Tabel 4.15 Perbandingan Nilai *Correct Clasification Rate* (CCR) dan *Area Under The Curve* (AUC) – Data *Training*

Metode	CCR (persen)	AUC (persen)
(1)	(2)	(3)
<i>Naïve Bayes</i>	68,0609	65,1798
<i>Hierarchical Naïve Bayes</i>	75,0578	72,2343
Regresi Logistik <i>Polytomous</i> (<i>enter</i>)	73,6893	68,8875
Regresi Logistik <i>Polytomous</i> (<i>stepwise</i>)	73,1688	67,6746
<i>Naïve Bayes</i> dengan Input Regresi Logistik <i>Polytomous</i> (<i>stepwise</i>)	69,0440	65,1529

Pada data *training*, metode *Hierarchical Naïve Bayes* mempunyai CCR dan AUC yang tertinggi, pada data *testing*, metode Regresi Logistik *Polytomous* (*enter*) mempunyai CCR dan AUC yang tertinggi.

Tabel 4.16 Perbandingan Nilai *Correct Clasification Rate* (CCR) dan *Area Under The Curve* (AUC) – data *Testing*

Metode	CCR (persen)	AUC (persen)
(1)	(2)	(3)
<i>Naïve Bayes</i>	66,8403	64,5369
<i>Hierarchical Naïve Bayes</i>	70,1389	68,3821
Regresi Logistik <i>Polytomous</i> (<i>enter</i>)	76,2153	71,3496
Regresi Logistik <i>Polytomous</i> (<i>stepwise</i>)	74,1319	68,3486
<i>Naïve Bayes</i> dengan Input Regresi Logistik <i>Polytomous</i> (<i>stepwise</i>)	68,2292	65,6655

Secara ukuran akurasi model kelima model tersebut bersaing. Namun secara komputasi, pengolahan klasifikasi untuk NB maupun HNB jauh lebih cepat dibandingkan Regresi Logistik *Polytomous*. Hal ini dapat ditunjukkan pada Tabel 4.17. Namun, dalam penelitian ini pemenuhan asumsi CI pada pembentukan variabel laten HNB belum dapat dipenuhi. Oleh karena itu hasil dari Regresi Logistik *Polytomous* lebih reliabel dari HNB.

Tabel 4.17 Perbandingan Lama Waktu Proses *Running Syntax*(detik)

Metode	User	Time	Elapse
(1)	(2)	(3)	(4)
<i>Naïve Bayes</i>	0,34	0,07	0,58
<i>Hierarchical Naïve Bayes</i>	1,78	0,22	4,90
Regresi Logistik <i>Polytomous</i> (<i>enter</i>)	653,30	141,80	796,39
Regresi Logistik <i>Polytomous</i> (<i>stepwise</i>)	599,38	81,97	681,25
<i>Naïve Bayes</i> dengan Input Regresi Logistik <i>Polytomous</i> (<i>stepwise</i>)	599,84	82,25	682,11

Perbandingan lama waktu proses *running syntax* pada Tabel 4.17 menggunakan *package (proc.time)* dari perangkat lunak R versi 4.3.2. *Package (proc.time)* menentukan berapa lama waktu nyata dan waktu *Central Processing Unit* (CPU) proses R yang sedang berjalan sudah dilakukan (dalam satuan detik). *User time* berhubungan dengan eksekusi dari *syntax*, *system time* berhubungan dengan sistem proses seperti membuka dan menutup *file*, sedangkan *elapse time*

adalah perbedaan waktu sejak mulai sampai proses berakhir. *Elastice time* berfungsi seperti *stopwatch*. Dengan spesifikasi processor Intel(R) Core(TM) i5-6200U CPU @2.30 GHz 2.4 GHz.dan RAM 4GB, dapat terlihat bahwa secara umum proses *running syntax* pada metode *Bayesian Network*, dalam hal ini NB dan HNB, jauh lebih cepat daripada metode Regresi Logistik *Polytomous*.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, Beberapa kesimpulan yang dapat diambil dan merujuk pada tujuan penelitian adalah sebagai berikut:

1. Dengan *Bayesian Network*, struktur model NB dibentuk dari sebelas variabel atribut, yaitu jenis kelamin, umur, status kawin, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, waktu tempuh, jumlah moda, biaya transportasi, dan PDRB kota tujuan serta satu variabel klasifikasi yaitu moda transportasi. Sedangkan struktur untuk model HNB dibentuk dari satu atribut independen yaitu jenis kelamin, tiga variabel laten dan satu variabel klasifikasi yaitu moda transportasi. Laten 1 dibentuk dari variabel jumlah moda dan biaya transportasi, laten 2 dibentuk dari variabel umur, status kawin, pendidikan dan kegiatan utama, serta laten 3 dibentuk dari variabel kota tujuan, jarak tempuh, waktu tempuh dan PDRB kota tujuan.
2. Regresi Logistik *Polytomous* dilakukan dengan menggunakan metode *enter* dan *stepwise*. Pada metode *stepwise* diperoleh delapan variabel yang signifikan yaitu jenis kelamin, umur, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, jumlah moda, dan biaya transportasi.
3. Pada penelitian ini, juga dilakukan klasifikasi NB dengan variabel input hasil Regresi Logistik *Polytomous* dilakukan dengan menggunakan metode *stepwise*
4. Dalam penelitian ini pemenuhan asumsi CI pada pembentukan variabel laten HNB belum dapat dipenuhi. Oleh karena itu hasil dari Regresi Logistik *Polytomous* lebih reliabel dari HNB sehingga model terbaik adalah Regresi Logistik *Polytomous (enter)*. Hal ini didukung oleh pencapaian nilai CCR dan AUC tertinggi pada data *testing* yang dicapai oleh Regresi Logistik *Polytomous (enter)*.

5.2 Saran

Berdasarkan hasil analisis serta kesimpulan di atas, maka dapat dirumuskan beberapa saran untuk penyempurnaan berkelanjutan dari hasil penelitian ini:

1. Perlu dikembangkan model *Hierarchical Naïve Bayes* yang mampu mengakomodir data-data berskala campuran (diskrit maupun kontinyu) atau secara umum mampu mengatasi data dengan distribusi yang beragam.
2. Pemenuhan asumsi untuk uji *Conditional Independence* belum dapat sepenuhnya dilakukan karena keterbatasan data. Diharapkan pada penelitian selanjutnya dapat dikembangkan metode yang dapat mengatasi hal tersebut.
3. Terdapat banyak frekuensi kosong untuk moda transportasi menurut beberapa karakteristik pada data penelitian ini. Hal ini menunjukkan bahwa sampel untuk beberapa moda transportasi sangat kecil. Oleh karena itu, dapat dikembangkan metode pengumpulan data untuk memperoleh jenis dan jumlah moda transportasi yang lebih representatif pada wilayah penelitian, misalnya dengan melakukan pencacahan pada terminal, halte busway, maupun stasiun kereta dengan pendekatan komuter bukan rumah tangga.
4. Berdasarkan hasil Regresi Logistik *Polytomous (stepwise)*, variabel yang mempengaruhi pemilihan moda adalah jenis kelamin, umur, pendidikan, kegiatan utama, kota tujuan, jarak tempuh, jumlah moda, dan biaya transportasi. Variabel-variabel inilah yang menjadi pertimbangan bagi komuter untuk memilih moda transportasi. Selanjutnya, NB dengan input Reglog *Polytomous (stepwise)*, didapatkan moda dengan peluang posterior tiga terbesar: sepeda motor, mobil, kendaraan umum dengan rute. Komuter memilih sepeda motor karena kepraktisan karena tidak berganti moda dan biaya yang rendah (sekitar Rp 6.001,00 - 12.000,00). Komuter memilih mobil karena kepraktisannya (tidak berganti moda) dan kenyamanannya. Kedua moda transportasi ini cenderung menuju ke arah DKI Jakarta. Oleh karena itu, saran untuk Pemerintah adalah menyediakan moda transportasi massal yang praktis, murah dan nyaman dengan tujuan lebih banyak ke arah DKI Jakarta pada pagi hari dan ke luar DKI Jakarta pada sore hari untuk mengakomodir kemauan dari pengguna motor dan mobil pribadi yang diharapkan dapat

menggeser peringkat ketiga terbanyak menjadi peringkat pertama terbanyak. Hal ini diharapkan bisa meningkatkan minat komuter memilih moda transportasi massal sehingga mengurangi kemacetan, kecelakaan dan tetap dapat menjaga kriteria moda transportasi yang diprioritaskan oleh dua teratas pengguna moda. Dalam penelitian ini jenis kelamin dan umur signifikan mempengaruhi pemilihan moda transportasi . Hal ini juga seharusnya menjadi pertimbangan Pemerintah untuk memberikan prioritas bagi wanita dan orang tua dalam transportasi massal.

DAFTAR PUSTAKA

- Agresti, A. (2002), *Categorical Data Analysis Second Edition*, John Wiley & Sons, Inc., New Jersey.
- Agresti, A. (2007), *An Introduction to Categorical Data Analysis Second Edition*, John Wiley & Sons, Inc., New Jersey.
- Ananta, A. dan Chotib (1996), *Mobilitas Penduduk dan Pembangunan Daerah: Analisis SUPAS 1995 (Indonesia)*, Lembaga Demografi Universitas Indonesia dan Kantor Menteri Negara Kependudukan/BKKBN, Jakarta.
- Ashari, A., Paryudi, I. dan Tjoa, A.M. (2013), "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool", *International Journal of Advanced Computer Science and Applications*, Vol. 4 No.11.
- Badan Pusat Statistik (BPS) (2014), *Statistik Komuter Jabodetabek - Hasil Survei Komuter Jabodetabek*, Badan Pusat Statistik, Jakarta.
- Ben-Gal, I. (2008), *Bayesian Networks. In Encyclopedia of Statistics in Quality and Reliability*, John Wiley & Sons, Ltd.
- Bielza, C. dan Larranaga, P. (2014), "Discrete Bayesian Network Classifiers: A Survey", *ACM Computing Surveys (CSUR)*, Vol. 47 (1).
- Cheng, J., Greiner, R., Kelly, J.B.D. dan Liu, W. (2002), "Learning Bayesian Network From Data: An Information-Theory Based Approach", *Artificial Intellegent*, Vol. 137, pp. 43-90.
- Dowdy, S., Weardon, S. dan Chilko, D. (2004), *Statistics for Research Third Edition*, John Wiley and Sons, Inc., New York.
- Draper, N.R. dan H., S. (1998), *Applied Regression Analysis, Third Edition*, John Wiley & Sons, New York.

- Friedman, N., Geiger, D. dan Goldszmidt, M. (1997), "Bayesian Network Classifiers", *Machine Learning*, Vol. 29, pp. 131–163.
- Gelman, A. dan Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, New York, Cambridge University Press.
- Giyarsih, S.R. (2001), "Gejala Urban Sprawl sebagai Pemicu Proses Densifikasi Permukiman di Daerah Pinggiran Kota (Urban Fringe Area) Kasus Pinggiran Kota Yogyakarta", *Jurnal Perencanaan Wilayah dan Kota* , Vol. 12 (1), pp. 40-45.
- Guntoro, D.W. (2016), *Analisis Faktor-Faktor yang Memengaruhi Penduduk Melakukan Migrasi Internal di Indonesia [skripsi]*, Yogyakarta, Universitas Negeri Yogyakarta.
- Haryono, S. (2017), *Bayesian Network untuk Klasifikasi Rumah Tangga Miskin di Aceh*, [tesis], ITS, Surabaya.
- Hosmer, D.W. dan Lemeshow, S. (2000), *Applied Logistic Regression* , John Wiley & Sons, Inc.
- Korb, K.B. dan Nicholson, A.E. (2010), *Bayesian Artificial Intelligence, Second Edition.*, CRC Press.
- Kusumaningrum, S.P. (2014), *Pengaruh Kondisi Individu dan Sosial Eomoni terhadap Keputusan Menjadi Commuter ke Kota Jakarta (Studi pada Tenaga Kerja Sektor Formal dari Bogor dan Depok yang Menggunakan Kereta Commuter Line) [skripsi]*, Universitas Brawijaya, Malang.
- Langseth, H. dan Nielsen, T.D. (2006), "Classification using Hierarchical Naive Bayes models", *Machine Learning*, Vol. 63(2), pp. 135-159.
- Lani, J. (2017), *Assumptions of Logistic Regression*. [Online] diakses dari: www.statisticssolutions.com/assumptions-of-logistic-regression/ [pada 22 Oktober 2017].
- Lee, E.S. (1966), "A Theory of Migration. Demography", *Demography*, Vol. 3, No. 1 , pp. 47-57.

- Li, S.Y., Roslan, S., Abdullah, M.C. dan Abdullah, H. (2015), "Commuter Families: Parental Readiness, Family Environment and Adolescent School Performance", *Procedia - Sosial and Behavioral Science* 172, pp. 686-692.
- Lovelace, R., Ballas, D. dan Watson, M. (2014), "A Spatial Microsimulation Approach for The Analysis of Commuter Patterns: From Individual to Regional Levels", *Journal of Transport Geography*, Vol. 34, pp. 282-296.
- Ma, T.Y. (2015), "Bayesian Networks for Multimodal Mode Choice Behavior Modelling: A Case Study for The Cross Border Workers of Luxembourg", *Transportation Research Procedia* 10, pp. 870-880.
- Mantra, I.B. (1981), *Population Movement in Wet Rice Communities*, Gajah Mada University Press, Yogyakarta.
- Mantra, I.B. (2000), *Demografi Umum. Edisi Kedua*, Pustaka Pelajar Offset, Yogyakarta.
- Mantra, I.B. dan Sunarto (1986), *Penelitian Kebijakan Kependudukan*, Pusat Penelitian Studi Kependudukan UGM, Yogyakarta.
- Mitchell, T. (2017), "Generative And Discriminative Classifiers: Naive Bayes and Logistic Regression", dalam *Machine Learning*, , Mc Graw Hill International Edition, pp. 1-17.
- Nagarajan, R., Scutari, M. dan Lebre, S. (2013), *Bayesian Network in R with Applications in Systems Biology*, Springer, New York.
- Sanjaya, Y.A. (2012), *Pembuatan Keputusan Migrasi dengan Pendekatan Regresi Hirarki di Indonesia [tesis]*, Universitas Padjajaran, Bandung.
- Tamin (2000), *Perencanaan dan Pemodelan Transportasi*, ITB, Bandung.
- Tian, Y., Shi, Y., Chen, X. dan Chen, W. (2011), "AUC Maximizing Support Vector Machines with Feature Selection", *Procedia Computer Science - International Conference on Computational Science, ICCS 2011*, Vol. 4 (2011), pp. 1691–1698.

- Walpole, R.E. (1995), *Pengantar Statistik Edisi ke-3 (Diterjemahkan oleh: Bambang Sumantri)*, Gramedia, Jakarta.
- Widaryatmo (2009), *Karakteristik pekerja pelaku mobilitas non permanen Indonesia 2007[tesis]*, Universitas Indonesia, Depok.
- Zhang, N.L., Nielsen, T.D. dan Jensen, F.V. (2003), "Latent Variable Discovery in Classification Models", *Artificial Intelligence in Medicine*, Vol. 30(3), pp. 283-299.
- Zhou, X., Obuchowski, N.A. dan McClish, D.K. (2002), *Statistical Methods in Diagnostic Medicine*, John Wiley & Sons, Inc.

LAMPIRAN

Lampiran 1. Moda Transportasi yang Digunakan

Kode	Moda	Persentase
(1)	(2)	(3)
1	Angkutan Perbatasan Terintegrasi Bus Transjakarta (APTB)	0,17
2	bus Transjakarta	2,13
3	kendaraan jemputan	2,05
4	kendaraan omprengan/sharing	0,31
5	kendaraan umum dengan rute	10,95
6	kendaraan umum dengan rute -- bus Transjakarta	0,26
7	kendaraan umum dengan rute -- kendaraan jemputan	0,26
8	kendaraan umum dengan rute -- kendaraan umum tanpa rute	0,19
9	kendaraan umum dengan rute -- kereta	2,59
10	kendaraan umum tanpa rute	0,21
11	kereta	1,08
12	kereta -- bus Transjakarta	0,23
13	mobil	12,60
14	mobil dinas	1,53
15	sepeda	0,68
16	sepeda motor	56,07
17	sepeda motor -- bus Transjakarta	0,21
18	sepeda motor -- kendaraan jemputan	0,78
19	sepeda motor -- kendaraan umum dengan rute	1,44
20	sepeda motor -- kendaraan umum dengan rute -- kereta	1,18
21	sepeda motor -- kereta	1,37
22	sepeda motor dinas	0,31
23	sepeda motor jemputan	0,49
24	sepeda motor ojek	1,18
25	sepeda motor ojek -- kendaraan umum dengan rute	1,01
26	sepeda motor ojek -- kereta	0,64
27	lainnya	0,10
Total		100,00

Lampiran 2. *Syntax* Pengambilan Sampel Data *Training - Testing*

```
library(haven)
dataku <- read_spss("D:/tesis/all.sav")

nrow(dataku)
nobs=nrow(dataku)
index=sample(nobs,ceiling((0.9*nobs)), replace=FALSE)
datatrain=dataku[index,]
datatest=dataku[-index,]
nrow(datatrain)
nrow(datatest)

write.table(datatrain, "D:/tesis/train.txt", sep="\t")
write.table(datatest, "D:/tesis/test.txt", sep="\t")
```


Lampiran 3. Persentase Nilai Harapan < 5 pada Tabel Kontingensi

	1	2	3	4	5	6	7	8	9	10	11
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1		20,74	3,70	9,26	25,00	9,26	20,74	21,11	29,01	24,44	14,20
2			20,74	22,72	34,63	13,70	30,96	31,41	31,60	32,59	21,98
3				10,49	32,41	9,26	20,74	21,85	32,10	23,70	14,81
4					34,26	10,49	27,65	27,16	33,33	28,89	16,05
5						12,96	29,63	28,52	34,88	27,78	21,60
6							24,81	25,19	33,33	27,78	32,10
7								32,59	35,31	29,33	22,72
8									35,06	29,78	22,72
9										21,23	15,23
10											23,21
11											

Keterangan: batas maksimal 20 persen

Lampiran 4. *Syntax Naïve Bayes*

```
#### membaca data
library(haven)
datatrain <- read_spss("D:/tesis/train.sav")
str(datatrain)

library(haven)
datatest <- read_spss("D:/tesis/test.sav")
str(datatest)

#### naive bayes
library(naivebayes)
nb <- naive_bayes(moda ~ moda27 ~ jk + cumur + kawin + pendidikan + kegut
+ kota + cjarak + clama + jmoda + cbiaya + cpdrb, data = datatrain)
nb

####prediksi data testing
predict_nb <- predict(nb, newdata=datatest)
write.table(predict_nb, "D:/tesis/predictnb.txt", sep="\t")

####Struktur
attach(datatrain)
library(phylopath)
models <- define_model_set(
  c(jk~moda27,
    kawin~moda27,
    pendidikan~moda27,
    kegut~moda27,
    kota~moda27,
    jmoda~moda27,
    cumur~moda27,
    cpdrb~moda27,
    cjarak~moda27,
    clama~moda27,
    cbiaya~moda27)
)
plot_model_set(models, algorithm = 'kk', text_size=4, edge_width=1)

####menghitung AUC
library(haven)
dataku <- read_spss("D:/tesis/hasiltest.sav")
aktual<-dataku$aktual
nb<-dataku$nb
library(rms)
rcorr.cens(nb,aktual)[1]
```

Lampiran 5. Output *Naïve Bayes*

```

===== Naive Bayes =====
Call:
naive_bayes.formula(formula = moda27 ~ jk + cumur + kawin + pendidikan +
  kegut + kota + cjarak + clama + jmoda + cbiaya + cpdrb, data = datatrain)

A priori probabilities:

      1      10      11      12      13
0.0015420200 0.0019275251 0.0094448728 0.0023130301 0.1268311488
      14      15      16      17      18
0.0146491904 0.0067463377 0.5587895143 0.0021202776 0.0079028527
      19      2      20      21      22
0.0150346955 0.0221665382 0.0113723978 0.0138781804 0.0030840401
      23      24      25      26      27
0.0053970702 0.0119506554 0.0100231303 0.0057825752 0.0009637625
      3      4      5      6      7
0.0213955281 0.0030840401 0.1098689283 0.0025057826 0.0026985351
      8      9
0.0019275251 0.0265998458

Tables:

jk      1      10      11      12      13      14      15
  1 0.5000000 0.6000000 0.5306122 0.6666667 0.7006079 0.6842105 0.6857143
  2 0.5000000 0.4000000 0.4693878 0.3333333 0.2993921 0.3157895 0.3142857

jk      16      17      18      19      2      20      21
  1 0.7961366 0.2727273 0.6585366 0.4230769 0.4260870 0.6440678 0.6666667
  2 0.2038634 0.7272727 0.3414634 0.5769231 0.5739130 0.3559322 0.3333333

jk      22      23      24      25      26      27      3
  1 0.9375000 0.2500000 0.3870968 0.3269231 0.3333333 0.6000000 0.4954955
  2 0.0625000 0.7500000 0.6129032 0.6730769 0.6666667 0.4000000 0.5045045

jk      4      5      6      7      8      9
  1 0.6250000 0.3964912 0.2307692 0.2142857 0.4000000 0.5289855
  2 0.3750000 0.6035088 0.7692308 0.7857143 0.6000000 0.4710145

cumur      1      10      11      12      13      14
  1 0.00000000 0.00000000 0.00000000 0.00000000 0.09422492 0.01315789
  2 0.12500000 0.10000000 0.32653061 0.58333333 0.08510638 0.02631579
  3 0.37500000 0.20000000 0.18367347 0.08333333 0.15957447 0.19736842
  4 0.25000000 0.40000000 0.24489796 0.16666667 0.28571429 0.34210526
  5 0.25000000 0.30000000 0.24489796 0.16666667 0.37537994 0.42105263

cumur      15      16      17      18      19      2
  1 0.22857143 0.04691273 0.00000000 0.00000000 0.03846154 0.00000000
  2 0.22857143 0.28527078 0.63636364 0.29268293 0.32051282 0.38260870
  3 0.08571429 0.29837875 0.27272727 0.46341463 0.25641026 0.29565217
  4 0.17142857 0.23283891 0.09090909 0.17073171 0.15384615 0.18260870
  5 0.28571429 0.13659883 0.00000000 0.07317073 0.23076923 0.13913043

cumur      20      21      22      23      24      25
  1 0.00000000 0.01388889 0.00000000 0.25000000 0.43548387 0.03846154
  2 0.22033898 0.22222222 0.06250000 0.32142857 0.32258065 0.30769231
  3 0.20338983 0.29166667 0.43750000 0.21428571 0.09677419 0.17307692
  4 0.32203390 0.26388889 0.37500000 0.10714286 0.04838710 0.21153846
  5 0.25423729 0.20833333 0.12500000 0.10714286 0.09677419 0.26923077

cumur      26      27      3      4      5      6
  1 0.00000000 0.40000000 0.28828829 0.06250000 0.14385965 0.00000000
  2 0.16666667 0.00000000 0.16216216 0.18750000 0.35087719 0.23076923
  3 0.36666667 0.00000000 0.12612613 0.31250000 0.17192982 0.30769231
  4 0.26666667 0.20000000 0.19819820 0.18750000 0.16315789 0.15384615
  5 0.20000000 0.40000000 0.22522523 0.25000000 0.17017544 0.30769231

```

Lanjutan Lampiran 5

cumur	7	8	9			
1	0.00000000	0.00000000	0.01449275			
2	0.21428571	0.20000000	0.27536232			
3	0.42857143	0.30000000	0.24637681			
4	0.21428571	0.20000000	0.21739130			
5	0.14285714	0.30000000	0.24637681			
kawin	1	10	11	12	13	14
1	0.3750000	0.1000000	0.3673469	0.6666667	0.2629179	0.1315789
2	0.6250000	0.9000000	0.6326531	0.3333333	0.7370821	0.8684211
kawin	15	16	17	18	19	2
1	0.4285714	0.4146257	0.7272727	0.2926829	0.4615385	0.6173913
2	0.5714286	0.5853743	0.2727273	0.7073171	0.5384615	0.3826087
kawin	20	21	22	23	24	25
1	0.2203390	0.3472222	0.1250000	0.6071429	0.7903226	0.5192308
2	0.7796610	0.6527778	0.8750000	0.3928571	0.2096774	0.4807692
kawin	26	27	3	4	5	6
1	0.4333333	0.4000000	0.4774775	0.3125000	0.5842105	0.3846154
2	0.5666667	0.6000000	0.5225225	0.6875000	0.4157895	0.6153846
kawin	7	8	9			
1	0.2857143	0.2000000	0.4130435			
2	0.7142857	0.8000000	0.5869565			
pendidikan	1	10	11	12	13	
1	0.00000000	0.20000000	0.16326531	0.00000000	0.14893617	
2	0.50000000	0.30000000	0.44897959	0.50000000	0.18237082	
3	0.50000000	0.50000000	0.38775510	0.50000000	0.66869301	
pendidikan	14	15	16	17	18	
1	0.03947368	0.57142857	0.20593308	0.00000000	0.04878049	
2	0.26315789	0.37142857	0.56122801	0.54545455	0.58536585	
3	0.69736842	0.05714286	0.23283891	0.45454545	0.36585366	
pendidikan	19	2	20	21	22	
1	0.16666667	0.10434783	0.03389831	0.09722222	0.12500000	
2	0.50000000	0.51304348	0.55932203	0.44444444	0.56250000	
3	0.33333333	0.38260870	0.40677966	0.45833333	0.31250000	
pendidikan	23	24	25	26	27	
1	0.42857143	0.66129032	0.19230769	0.16666667	0.80000000	
2	0.28571429	0.19354839	0.48076923	0.13333333	0.20000000	
3	0.28571429	0.14516129	0.32692308	0.70000000	0.00000000	
pendidikan	3	4	5	6	7	
1	0.46846847	0.50000000	0.39122807	0.07692308	0.28571429	
2	0.24324324	0.37500000	0.41578947	0.23076923	0.50000000	
3	0.28828829	0.12500000	0.19298246	0.69230769	0.21428571	
pendidikan	8	9				
1	0.30000000	0.16666667				
2	0.30000000	0.45652174				
3	0.40000000	0.37681159				
kegut	1	10	11	12	13	14
1	0.00000000	0.00000000	0.02040816	0.00000000	0.15349544	0.01315789
2	0.00000000	0.00000000	0.10204082	0.16666667	0.11702128	0.22368421
3	1.00000000	0.90000000	0.79591837	0.50000000	0.56382979	0.75000000
4	0.00000000	0.10000000	0.08163265	0.33333333	0.16565350	0.01315789
kegut	15	16	17	18	19	2
1	0.17142857	0.05657123	0.00000000	0.00000000	0.05128205	0.02608696
2	0.02857143	0.05001725	0.09090909	0.12195122	0.05128205	0.01739130
3	0.51428571	0.73301138	0.90909091	0.87804878	0.64102564	0.77391304
4	0.28571429	0.16040014	0.00000000	0.00000000	0.25641026	0.18260870

Lanjutan Lampiran 5

kegut	20	21	22	23	24	25
1	0.00000000	0.00000000	0.00000000	0.00000000	0.04838710	0.01923077
2	0.23728814	0.11111111	0.06250000	0.10714286	0.03225806	0.09615385
3	0.69491525	0.83333333	0.93750000	0.57142857	0.25806452	0.63461538
4	0.06779661	0.05555556	0.00000000	0.32142857	0.66129032	0.25000000
kegut	26	27	3	4	5	6
1	0.13333333	0.40000000	0.02702703	0.12500000	0.03333333	0.00000000
2	0.13333333	0.00000000	0.24324324	0.06250000	0.03684211	0.00000000
3	0.70000000	0.20000000	0.41441441	0.62500000	0.59824561	0.84615385
4	0.03333333	0.40000000	0.31531532	0.18750000	0.33157895	0.15384615
kegut	7	8	9			
1	0.07142857	0.20000000	0.05072464			
2	0.07142857	0.00000000	0.05072464			
3	0.85714286	0.60000000	0.79710145			
4	0.00000000	0.20000000	0.10144928			
kota	1	10	11	12	13	14
1	0.87500000	0.80000000	0.79591837	0.66666667	0.80851064	0.68421053
2	0.12500000	0.20000000	0.20408163	0.33333333	0.19148936	0.31578947
kota	15	16	17	18	19	2
1	0.65714286	0.70507071	1.00000000	0.26829268	0.62820513	1.00000000
2	0.34285714	0.29492929	0.00000000	0.73170732	0.37179487	0.00000000
kota	20	21	22	23	24	25
1	0.93220339	0.90277778	0.68750000	0.71428571	0.70967742	0.61538462
2	0.06779661	0.09722222	0.31250000	0.28571429	0.29032258	0.38461538
kota	26	27	3	4	5	6
1	0.96666667	0.80000000	0.62162162	0.68750000	0.68596491	0.84615385
2	0.03333333	0.20000000	0.37837838	0.31250000	0.31403509	0.15384615
kota	7	8	9			
1	0.64285714	0.80000000	0.83333333			
2	0.35714286	0.20000000	0.16666667			
cjarak	1	10	11	12	13	14
1	0.00000000	0.00000000	0.02040816	0.00000000	0.20060790	0.10526316
2	0.00000000	0.00000000	0.20408163	0.00000000	0.25531915	0.19736842
3	0.25000000	0.30000000	0.26530612	0.33333333	0.26443769	0.27631579
4	0.25000000	0.40000000	0.24489796	0.08333333	0.15349544	0.18421053
5	0.50000000	0.30000000	0.26530612	0.58333333	0.12613982	0.23684211
cjarak	15	16	17	18	19	2
1	0.74285714	0.26629872	0.18181818	0.00000000	0.11538462	0.19130435
2	0.17142857	0.31390135	0.18181818	0.12195122	0.19230769	0.40869565
3	0.00000000	0.22731977	0.27272727	0.39024390	0.26923077	0.22608696
4	0.05714286	0.12211107	0.36363636	0.26829268	0.25641026	0.14782609
5	0.02857143	0.07036909	0.00000000	0.21951220	0.16666667	0.02608696
cjarak	20	21	22	23	24	25
1	0.00000000	0.00000000	0.37500000	0.57142857	0.67741935	0.21153846
2	0.03389831	0.04166667	0.31250000	0.21428571	0.24193548	0.25000000
3	0.25423729	0.29166667	0.31250000	0.14285714	0.03225806	0.30769231
4	0.15254237	0.31944444	0.00000000	0.03571429	0.01612903	0.13461538
5	0.55932203	0.34722222	0.00000000	0.03571429	0.03225806	0.09615385
cjarak	26	27	3	4	5	6
1	0.03333333	1.00000000	0.25225225	0.25000000	0.39473684	0.15384615
2	0.06666667	0.00000000	0.26126126	0.12500000	0.27368421	0.23076923
3	0.36666667	0.00000000	0.16216216	0.12500000	0.15964912	0.23076923
4	0.26666667	0.00000000	0.16216216	0.18750000	0.09824561	0.15384615
5	0.26666667	0.00000000	0.16216216	0.31250000	0.07368421	0.23076923

Lanjutan Lampiran 5

cjarak	7	8	9			
1	0.28571429	0.20000000	0.02173913			
2	0.07142857	0.10000000	0.10869565			
3	0.42857143	0.30000000	0.28985507			
4	0.07142857	0.30000000	0.18115942			
5	0.14285714	0.10000000	0.39855072			
clama	1	10	11	12	13	14
1	0.00000000	0.00000000	0.00000000	0.00000000	0.07446809	0.02631579
2	0.00000000	0.00000000	0.28571429	0.08333333	0.24164134	0.19736842
3	0.12500000	0.30000000	0.44897959	0.00000000	0.30395137	0.34210526
4	0.37500000	0.10000000	0.18367347	0.33333333	0.19604863	0.30263158
5	0.50000000	0.60000000	0.08163265	0.58333333	0.18389058	0.13157895
clama	15	16	17	18	19	2
1	0.40000000	0.15867541	0.00000000	0.00000000	0.05128205	0.03478261
2	0.40000000	0.35494998	0.27272727	0.04878049	0.24358974	0.31304348
3	0.08571429	0.27457744	0.18181818	0.43902439	0.23076923	0.26956522
4	0.02857143	0.13418420	0.36363636	0.34146341	0.19230769	0.20869565
5	0.08571429	0.07761297	0.18181818	0.17073171	0.28205128	0.17391304
clama	20	21	22	23	24	25
1	0.00000000	0.00000000	0.12500000	0.39285714	0.43548387	0.00000000
2	0.05084746	0.13888889	0.50000000	0.32142857	0.43548387	0.21153846
3	0.33898305	0.31944444	0.25000000	0.14285714	0.06451613	0.26923077
4	0.25423729	0.37500000	0.06250000	0.07142857	0.03225806	0.19230769
5	0.35593220	0.16666667	0.06250000	0.07142857	0.03225806	0.32692308
clama	26	27	3	4	5	6
1	0.03333333	1.00000000	0.08108108	0.12500000	0.12807018	0.00000000
2	0.03333333	0.00000000	0.24324324	0.18750000	0.37192982	0.07692308
3	0.30000000	0.00000000	0.30630631	0.12500000	0.22807018	0.38461538
4	0.20000000	0.00000000	0.11711712	0.18750000	0.12631579	0.23076923
5	0.43333333	0.00000000	0.25225225	0.37500000	0.14561404	0.30769231
clama	7	8	9			
1	0.00000000	0.00000000	0.00000000			
2	0.28571429	0.20000000	0.14492754			
3	0.14285714	0.30000000	0.32608696			
4	0.21428571	0.30000000	0.29710145			
5	0.35714286	0.20000000	0.23188406			
jmoda	1	10	11	12	13	
1	0.0000000000	0.6000000000	0.4897959184	0.0000000000	0.9863221884	
2	0.6250000000	0.3000000000	0.2857142857	0.3333333333	0.0136778116	
3	0.3750000000	0.1000000000	0.2244897959	0.6666666667	0.0000000000	
jmoda	14	15	16	17	18	
1	0.8947368421	1.0000000000	0.9948258020	0.0000000000	0.0000000000	
2	0.1052631579	0.0000000000	0.0044843049	1.0000000000	1.0000000000	
3	0.0000000000	0.0000000000	0.0006898931	0.0000000000	0.0000000000	
jmoda	19	2	20	21	22	
1	0.0000000000	0.2260869565	0.0000000000	0.0000000000	1.0000000000	
2	0.7948717949	0.6347826087	0.0000000000	0.9027777778	0.0000000000	
3	0.2051282051	0.1391304348	1.0000000000	0.0972222222	0.0000000000	
jmoda	23	24	25	26	27	
1	0.9642857143	0.8870967742	0.0000000000	0.0000000000	1.0000000000	
2	0.0357142857	0.0967741935	0.6153846154	0.2666666667	0.0000000000	
3	0.0000000000	0.0161290323	0.3846153846	0.7333333333	0.0000000000	
jmoda	3	4	5	6	7	
1	0.8468468468	0.5625000000	0.4105263158	0.0000000000	0.0000000000	
2	0.1441441441	0.3750000000	0.4263157895	0.0000000000	1.0000000000	
3	0.0090090090	0.0625000000	0.1631578947	1.0000000000	0.0000000000	

Lanjutan Lampiran 5

jmoda	8	9
1	0.0000000000	0.0000000000
2	0.4000000000	0.6014492754
3	0.6000000000	0.3985507246

cbiaya	1	10	11	12	13	14
1	0.00000000	0.00000000	0.20408163	0.00000000	0.02279635	0.03947368
2	0.00000000	0.00000000	0.34693878	0.00000000	0.02583587	0.05263158
3	0.00000000	0.20000000	0.06122449	0.16666667	0.03799392	0.02631579
4	0.12500000	0.00000000	0.14285714	0.25000000	0.07750760	0.03947368
5	0.87500000	0.80000000	0.24489796	0.58333333	0.83586626	0.84210526

cbiaya	15	16	17	18	19	2
1	0.77142857	0.19868920	0.09090909	0.56097561	0.06410256	0.08695652
2	0.20000000	0.50051742	0.36363636	0.24390244	0.26923077	0.29565217
3	0.00000000	0.20386340	0.36363636	0.07317073	0.23076923	0.40869565
4	0.02857143	0.06760952	0.18181818	0.04878049	0.17948718	0.12173913
5	0.00000000	0.02932046	0.00000000	0.07317073	0.25641026	0.08695652

cbiaya	20	21	22	23	24	25
1	0.00000000	0.06944444	0.12500000	0.21428571	0.04838710	0.00000000
2	0.05084746	0.40277778	0.68750000	0.50000000	0.19354839	0.05769231
3	0.40677966	0.27777778	0.06250000	0.17857143	0.17741935	0.15384615
4	0.32203390	0.16666667	0.06250000	0.03571429	0.25806452	0.32692308
5	0.22033898	0.08333333	0.06250000	0.07142857	0.32258065	0.46153846

cbiaya	26	27	3	4	5	6
1	0.00000000	1.00000000	0.13513514	0.18750000	0.22807018	0.00000000
2	0.06666667	0.00000000	0.46846847	0.12500000	0.33157895	0.07692308
3	0.10000000	0.00000000	0.09909910	0.18750000	0.16666667	0.15384615
4	0.23333333	0.00000000	0.15315315	0.12500000	0.14912281	0.23076923
5	0.60000000	0.00000000	0.14414414	0.37500000	0.12456140	0.53846154

cbiaya	7	8	9
1	0.21428571	0.00000000	0.00000000
2	0.35714286	0.00000000	0.20289855
3	0.21428571	0.40000000	0.38405797
4	0.07142857	0.30000000	0.23188406
5	0.14285714	0.30000000	0.18115942

cpdrb	1	10	11	12	13	14
1	0.00000000	0.20000000	0.16326531	0.33333333	0.14437690	0.22368421
2	0.25000000	0.00000000	0.12244898	0.16666667	0.15653495	0.15789474
3	0.75000000	0.80000000	0.71428571	0.50000000	0.69908815	0.61842105

cpdrb	15	16	17	18	19	2
1	0.28571429	0.22731977	0.00000000	0.17073171	0.29487179	0.00000000
2	0.25714286	0.18178682	0.00000000	0.60975610	0.15384615	0.15652174
3	0.45714286	0.59089341	1.00000000	0.21951220	0.55128205	0.84347826

cpdrb	20	21	22	23	24	25
1	0.05084746	0.08333333	0.31250000	0.21428571	0.19354839	0.23076923
2	0.11864407	0.05555556	0.00000000	0.28571429	0.25806452	0.26923077
3	0.83050847	0.86111111	0.68750000	0.50000000	0.54838710	0.50000000

cpdrb	26	27	3	4	5	6
1	0.03333333	0.20000000	0.21621622	0.25000000	0.22807018	0.07692308
2	0.10000000	0.00000000	0.18918919	0.18750000	0.21403509	0.07692308
3	0.86666667	0.80000000	0.59459459	0.56250000	0.55789474	0.84615385

cpdrb	7	8	9
1	0.07142857	0.10000000	0.13043478
2	0.50000000	0.20000000	0.14492754
3	0.42857143	0.70000000	0.72463768

Lampiran 6. *Syntax Hierarchical Naïve Bayes*

```
##### memanggil data
library(haven)
dataku <- read_spss("D:/tesis/train.sav")

##### uji independensi
#### uji chi-square
#diambil variabel independennya saja
dites <-
cbind(datatrain$jk, datatrain$cumur, datatrain$kawin, datatrain$pendidikan, d
atrain$kegut, datatrain$kota, datatrain$jarak, datatrain$lama, datatrain$
jmoda, datatrain$biaya, datatrain$cpdrb)

##fungsi test chi square##
chit <- function (dites){
  a<- ncol(dites)
  hasil <- as.data.frame (matrix (a,a))
  for (i in 1:a){
    for (j in 1:a){
      if (j>i){
        b <- as.numeric(chisq.test(dites[,i],dites[,j])[3])
        hasil[i,j] <- b
        # p value nya aja yg dimasukin ke matriks
      }
    }
  }
  return (hasil)
}

##tes chisquare, hasil dalam bentuk matriks ##
chisquare <- chit(dites)
print(chisquare)

#### uji conditional independence
library(bnlearn)
angka_condi <- function (dites){
  a<- ncol(dites)
  hasil <- as.data.frame (matrix (a,a))
  for (i in 1:a){
    for (j in 1:a){
      if (i>j){
        b <- as.numeric(ci.test(dites[,i],dites[,j],dites[,1])[2])
        hasil[i,j] <- b
        # conditional chi square p-value nya yg diambil
      }
    }
  }
  return (hasil)
}

##pembentukan laten by conditional independence
hclustering <- function (datatrain){
  coba <- angka_condi(datatrain) #kalo pake chisquare ganti pake
angka_chi
  print("Matriks Conditional Independence")
  print(coba)
  coba<- as.dist(coba[-1,])
  hc<-hclust(coba)
  plot(hc)
}
```


Lanjutan Lampiran 6

```
##variabel laten
varlatenbaru <-function (kij,var_a,var_b,a,b){
  n <- nrow(cdatatrain)
  #a adalah vektor berisi kategori-kategori di variabel a
  #b adalah vektor berisi kategori-kategori di var b
  varbaru <- matrix(1,n)
  for (i in 1:n){
    for (j in 1:nrow(a)){
      if (var_a[i]== a[j]){
        x <- j
      }
    }
    for (k in 1:nrow(b)){
      if (var_b[i]== a[k]){
        y <- k
      }
    }

    varbaru[i] <- kij[x,y]
  }
}

#### naive bayes
library(naivebayes)
hnb <- naive_bayes(moda ~ moda27 ~ jk + laten1 + aten2 + laten3,
data=datatrain)
plot(hnb)

####prediksi data testing
predict_hnb<- predict(hnb, newdata=datatest)
write.table(predict_hnb, "D:/tesis/predicthnb.txt", sep="\t")

####Struktur
attach(datatrain)
library(phylopath)
models <- define_model_set(
  c(jk~moda27,
    laten1~moda27,
    laten2~moda27,
    laten3~moda27,
    jmoda~laten1,
    biaya~laten1,
    cumur~laten2,
    kawin~laten2,
    pendidikan~laten2,
    kegut~laten2,
    kota~laten3,
    clama~laten3,
    cjarak~laten3,
    cpdrb~laten3)
)
plot_model_set(models, algorithm = 'kk', text_size=4, edge_width=1)

####menghitung AUC
library(haven)
dataku <- read_spss("D:/tesis/hasiltest.sav")
aktual<-dataku$aktual
hnb<-dataku$hnb
library(rms)
rcorr.cens(hnb,aktual)[1]
```

Lampiran 7. Output *Hierarchical Naïve Bayes*

```

===== Naive Bayes =====
Call:
naive_bayes.formula(formula = moda27 ~ jk + laten1 + laten2 +
  laten3, data = datatrain)

A priori probabilities:

      1      10      11      12      13
0.0015420200 0.0019275251 0.0094448728 0.0023130301 0.1268311488
      14      15      16      17      18
0.0146491904 0.0067463377 0.5587895143 0.0021202776 0.0079028527
      19      2      20      21      22
0.0150346955 0.0221665382 0.0113723978 0.0138781804 0.0030840401
      23      24      25      26      27
0.0053970702 0.0119506554 0.0100231303 0.0057825752 0.0009637625
      3      4      5      6      7
0.0213955281 0.0030840401 0.1098689283 0.0025057826 0.0026985351
      8      9
0.0019275251 0.0265998458

Tables:

jk      1      10      11      12      13      14      15
1 0.5000000 0.6000000 0.5306122 0.6666667 0.7006079 0.6842105 0.6857143
2 0.5000000 0.4000000 0.4693878 0.3333333 0.2993921 0.3157895 0.3142857

jk      16      17      18      19      2      20      21
1 0.7961366 0.2727273 0.6585366 0.4230769 0.4260870 0.6440678 0.6666667
2 0.2038634 0.7272727 0.3414634 0.5769231 0.5739130 0.3559322 0.3333333

jk      22      23      24      25      26      27      3
1 0.9375000 0.2500000 0.3870968 0.3269231 0.3333333 0.6000000 0.4954955
2 0.0625000 0.7500000 0.6129032 0.6730769 0.6666667 0.4000000 0.5045045

jk      4      5      6      7      8      9
1 0.6250000 0.3964912 0.2307692 0.2142857 0.4000000 0.5289855
2 0.3750000 0.6035088 0.7692308 0.7857143 0.6000000 0.4710145

laten1      1      10      11      12      13
1 0.0000000000 0.0000000000 0.2040816327 0.0000000000 0.0212765957
10 0.5000000000 0.3000000000 0.1020408163 0.1666666667 0.0091185410
12 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
13 0.0000000000 0.0000000000 0.0204081633 0.0000000000 0.0000000000
14 0.0000000000 0.0000000000 0.0816326531 0.2500000000 0.0000000000
15 0.3750000000 0.1000000000 0.1224489796 0.4166666667 0.0000000000
2 0.0000000000 0.0000000000 0.2448979592 0.0000000000 0.0243161094
3 0.0000000000 0.2000000000 0.0000000000 0.0000000000 0.0379939210
4 0.0000000000 0.0000000000 0.0204081633 0.0000000000 0.0759878419
5 0.0000000000 0.4000000000 0.0204081633 0.0000000000 0.8267477204
6 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015197568
7 0.0000000000 0.0000000000 0.1020408163 0.0000000000 0.0015197568
8 0.0000000000 0.0000000000 0.0408163265 0.1666666667 0.0000000000
9 0.1250000000 0.0000000000 0.0408163265 0.0000000000 0.0015197568

laten1      14      15      16      17      18
1 0.0131578947 0.7714285714 0.1983442566 0.0000000000 0.0000000000
10 0.0394736842 0.0000000000 0.0003449465 0.0000000000 0.0731707317
12 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
13 0.0000000000 0.0000000000 0.0006898931 0.0000000000 0.0000000000
14 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
15 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
2 0.0263157895 0.2000000000 0.4981027941 0.0000000000 0.0000000000
3 0.0131578947 0.0000000000 0.2021386685 0.0000000000 0.0000000000
4 0.0394736842 0.0285714286 0.0672645740 0.0000000000 0.0000000000

```

Lanjutan Lampiran 7

5	0.8026315789	0.0000000000	0.0289755088	0.0000000000	0.0000000000
6	0.0263157895	0.0000000000	0.0003449465	0.0909090909	0.5609756098
7	0.0263157895	0.0000000000	0.0024146257	0.3636363636	0.2439024390
8	0.0131578947	0.0000000000	0.0010348396	0.3636363636	0.0731707317
9	0.0000000000	0.0000000000	0.0003449465	0.1818181818	0.0487804878
laten1	19	2	20	21	22
1	0.0000000000	0.0695652174	0.0000000000	0.0000000000	0.1250000000
10	0.1410256410	0.0434782609	0.0000000000	0.0694444444	0.0000000000
12	0.0256410256	0.0000000000	0.0508474576	0.0138888889	0.0000000000
13	0.0384615385	0.0086956522	0.4067796610	0.0277777778	0.0000000000
14	0.0256410256	0.0869565217	0.3220338983	0.0416666667	0.0000000000
15	0.1153846154	0.0434782609	0.2203389831	0.0138888889	0.0000000000
2	0.0000000000	0.1391304348	0.0000000000	0.0000000000	0.6875000000
3	0.0000000000	0.0173913043	0.0000000000	0.0000000000	0.0625000000
4	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0625000000
5	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0625000000
6	0.0641025641	0.0173913043	0.0000000000	0.0694444444	0.0000000000
7	0.2435897436	0.1565217391	0.0000000000	0.3888888889	0.0000000000
8	0.1923076923	0.3826086957	0.0000000000	0.2500000000	0.0000000000
9	0.1538461538	0.0347826087	0.0000000000	0.1250000000	0.0000000000
laten1	23	24	25	26	27
1	0.2142857143	0.0483870968	0.0000000000	0.0000000000	1.0000000000
10	0.0000000000	0.0483870968	0.2115384615	0.2333333333	0.0000000000
12	0.0000000000	0.0000000000	0.0000000000	0.0666666667	0.0000000000
13	0.0000000000	0.0000000000	0.0384615385	0.0666666667	0.0000000000
14	0.0000000000	0.0000000000	0.0961538462	0.2333333333	0.0000000000
15	0.0000000000	0.0161290323	0.2500000000	0.3666666667	0.0000000000
2	0.4642857143	0.1935483871	0.0000000000	0.0000000000	0.0000000000
3	0.1785714286	0.1774193548	0.0000000000	0.0000000000	0.0000000000
4	0.0357142857	0.2096774194	0.0000000000	0.0000000000	0.0000000000
5	0.0714285714	0.2580645161	0.0000000000	0.0000000000	0.0000000000
6	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
7	0.0357142857	0.0000000000	0.0576923077	0.0000000000	0.0000000000
8	0.0000000000	0.0000000000	0.1153846154	0.0333333333	0.0000000000
9	0.0000000000	0.0483870968	0.2307692308	0.0000000000	0.0000000000
laten1	3	4	5	6	7
1	0.1351351351	0.1250000000	0.2105263158	0.0000000000	0.0000000000
10	0.0630630631	0.1875000000	0.0473684211	0.0000000000	0.1428571429
12	0.0000000000	0.0000000000	0.0175438596	0.0769230769	0.0000000000
13	0.0000000000	0.0625000000	0.0333333333	0.1538461538	0.0000000000
14	0.0000000000	0.0000000000	0.0561403509	0.2307692308	0.0000000000
15	0.0090090090	0.0000000000	0.0561403509	0.5384615385	0.0000000000
2	0.4324324324	0.1250000000	0.1157894737	0.0000000000	0.0000000000
3	0.0900900901	0.0000000000	0.0263157895	0.0000000000	0.0000000000
4	0.1171171171	0.1250000000	0.0368421053	0.0000000000	0.0000000000
5	0.0720720721	0.1875000000	0.0210526316	0.0000000000	0.0000000000
6	0.0000000000	0.0625000000	0.0175438596	0.0000000000	0.2142857143
7	0.0360360360	0.0000000000	0.1982456140	0.0000000000	0.3571428571
8	0.0090090090	0.1250000000	0.1070175439	0.0000000000	0.2142857143
9	0.0360360360	0.0000000000	0.0561403509	0.0000000000	0.0714285714
laten1	8	9			
1	0.0000000000	0.0000000000			
10	0.1000000000	0.0724637681			
12	0.0000000000	0.0144927536			
13	0.2000000000	0.1449275362			
14	0.2000000000	0.1304347826			
15	0.2000000000	0.1086956522			
2	0.0000000000	0.0000000000			
3	0.0000000000	0.0000000000			
4	0.0000000000	0.0000000000			
5	0.0000000000	0.0000000000			
6	0.0000000000	0.0000000000			
7	0.0000000000	0.1884057971			
8	0.2000000000	0.2391304348			
9	0.1000000000	0.1014492754			

Lanjutan Lampiran 7

laten2	1	10	11	12	13
1	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
10	0.1250000000	0.0000000000	0.0612244898	0.0000000000	0.0045592705
11	0.0000000000	0.1000000000	0.0408163265	0.3333333333	0.0319148936
12	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0000000000	0.0833333333	0.0000000000
14	0.0000000000	0.0000000000	0.0816326531	0.1666666667	0.0167173252
15	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0030395137
16	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
17	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
18	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0942249240
20	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0204081633	0.0000000000	0.0000000000
22	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0015197568
23	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0015197568
24	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
25	0.0000000000	0.0000000000	0.0204081633	0.0000000000	0.0000000000
26	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
27	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
28	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0015197568
29	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
3	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
30	0.0000000000	0.0000000000	0.0204081633	0.0000000000	0.0000000000
31	0.0000000000	0.0000000000	0.0408163265	0.0000000000	0.0030395137
32	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
33	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0015197568
34	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0015197568
35	0.2500000000	0.0000000000	0.0000000000	0.0000000000	0.0516717325
36	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0030395137
37	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0045592705
38	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
39	0.0000000000	0.2000000000	0.0204081633	0.0000000000	0.0015197568
4	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
40	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0045592705
41	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0015197568
42	0.0000000000	0.0000000000	0.0612244898	0.0000000000	0.0060790274
laten2	14	15	16	17	18
1	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
10	0.0000000000	0.0571428571	0.1010693343	0.4545454545	0.1707317073
11	0.0000000000	0.0571428571	0.0758882373	0.0000000000	0.0000000000
12	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
14	0.0131578947	0.0571428571	0.0307002415	0.1818181818	0.0731707317
15	0.0000000000	0.0000000000	0.0051741980	0.0000000000	0.0000000000
16	0.0000000000	0.0000000000	0.0006898931	0.0000000000	0.0000000000
17	0.0000000000	0.0000000000	0.0041393584	0.0000000000	0.0000000000
18	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0006898931	0.0000000000	0.0000000000
2	0.0131578947	0.2285714286	0.0465677820	0.0000000000	0.0000000000
20	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
21	0.0131578947	0.0285714286	0.0162124871	0.0000000000	0.0487804878
22	0.0000000000	0.0000000000	0.0006898931	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
24	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
25	0.0000000000	0.0000000000	0.0041393584	0.0000000000	0.0000000000
26	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
27	0.0000000000	0.0000000000	0.0010348396	0.0000000000	0.0000000000
28	0.0000000000	0.0000000000	0.0068989307	0.0000000000	0.0000000000
29	0.0000000000	0.0000000000	0.0027595723	0.0000000000	0.0000000000
3	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
30	0.0000000000	0.0000000000	0.0013797861	0.0000000000	0.0000000000
31	0.0000000000	0.0000000000	0.0431183167	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0031045188	0.0000000000	0.0000000000
33	0.0131578947	0.0000000000	0.0010348396	0.0000000000	0.0000000000
34	0.0000000000	0.0000000000	0.0024146257	0.0000000000	0.0000000000
35	0.0526315789	0.0000000000	0.0334598137	0.0909090909	0.0243902439

Lanjutan Lampiran 7

36	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
37	0.0000000000	0.0000000000	0.0041393584	0.0000000000	0.0000000000
38	0.0000000000	0.0000000000	0.0006898931	0.0000000000	0.0000000000
39	0.0000000000	0.0285714286	0.0251810969	0.0000000000	0.0487804878
4	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
40	0.0000000000	0.0000000000	0.0051741980	0.0000000000	0.0000000000
41	0.0000000000	0.0000000000	0.0034494653	0.0000000000	0.0487804878
42	0.0263157895	0.0571428571	0.1124525699	0.0000000000	0.2439024390
laten2	19	2	20	21	22
1	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
10	0.0769230769	0.0956521739	0.0338983051	0.0694444444	0.0625000000
11	0.1538461538	0.1565217391	0.0677966102	0.0000000000	0.0000000000
12	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0169491525	0.0138888889	0.0000000000
14	0.0000000000	0.0869565217	0.0508474576	0.0555555556	0.0000000000
15	0.0256410256	0.0173913043	0.0000000000	0.0138888889	0.0000000000
16	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
17	0.0000000000	0.0086956522	0.0000000000	0.0000000000	0.0000000000
18	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2	0.0384615385	0.0000000000	0.0000000000	0.0138888889	0.0000000000
20	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0338983051	0.0000000000	0.0000000000
22	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
24	0.0000000000	0.0000000000	0.0169491525	0.0000000000	0.0000000000
25	0.0128205128	0.0000000000	0.0000000000	0.0000000000	0.0000000000
26	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
27	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
28	0.0000000000	0.0086956522	0.0000000000	0.0000000000	0.0625000000
29	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
3	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
30	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
31	0.0384615385	0.0695652174	0.0000000000	0.0555555556	0.0000000000
32	0.0000000000	0.0000000000	0.0000000000	0.0138888889	0.0000000000
33	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
34	0.0000000000	0.0000000000	0.0169491525	0.0000000000	0.0000000000
35	0.0641025641	0.1043478261	0.0169491525	0.0416666667	0.0000000000
36	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
37	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
38	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
39	0.0000000000	0.0173913043	0.0000000000	0.0000000000	0.0625000000
4	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
40	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
41	0.0128205128	0.0000000000	0.0169491525	0.0000000000	0.0000000000
42	0.0769230769	0.0434782609	0.0847457627	0.1111111111	0.1875000000
laten2	23	24	25	26	27
1	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
10	0.1071428571	0.0483870968	0.0384615385	0.0000000000	0.0000000000
11	0.0000000000	0.0645161290	0.1346153846	0.0333333333	0.0000000000
12	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
14	0.1071428571	0.0483870968	0.0384615385	0.1000000000	0.0000000000
15	0.0000000000	0.0161290323	0.0000000000	0.0000000000	0.0000000000
16	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
17	0.0357142857	0.0000000000	0.0000000000	0.0000000000	0.0000000000
18	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2	0.2500000000	0.4354838710	0.0384615385	0.0000000000	0.4000000000
20	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
22	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
24	0.0000000000	0.0000000000	0.0192307692	0.0000000000	0.0000000000
25	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
26	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
27	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000

Lanjutan Lampiran 7

28	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
29	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
3	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
30	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
31	0.0000000000	0.0161290323	0.0384615385	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
33	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
34	0.0000000000	0.0000000000	0.0000000000	0.0333333333	0.0000000000
35	0.0714285714	0.0161290323	0.0576923077	0.1666666667	0.0000000000
36	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
37	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
38	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
39	0.0714285714	0.0322580645	0.0192307692	0.0000000000	0.0000000000
4	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
40	0.0000000000	0.0161290323	0.0000000000	0.0000000000	0.0000000000
41	0.0000000000	0.0000000000	0.0000000000	0.0333333333	0.0000000000
42	0.0714285714	0.0161290323	0.0000000000	0.0000000000	0.0000000000
laten2	3	4	5	6	7
1	0.0090090090	0.0000000000	0.0000000000	0.0000000000	0.0000000000
10	0.0630630631	0.0000000000	0.0912280702	0.0769230769	0.0000000000
11	0.0000000000	0.0625000000	0.0859649123	0.0769230769	0.0000000000
12	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
14	0.0180180180	0.0000000000	0.0403508772	0.0000000000	0.0000000000
15	0.0000000000	0.0000000000	0.0017543860	0.0769230769	0.0000000000
16	0.0090090090	0.0000000000	0.0000000000	0.0000000000	0.0000000000
17	0.0090090090	0.0000000000	0.0052631579	0.0000000000	0.0000000000
18	0.0000000000	0.0000000000	0.0052631579	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2	0.2792792793	0.0625000000	0.1403508772	0.0000000000	0.0000000000
20	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0122807018	0.0000000000	0.0714285714
22	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
24	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
25	0.0090090090	0.0000000000	0.0052631579	0.0000000000	0.0000000000
26	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
27	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
28	0.0000000000	0.0000000000	0.0070175439	0.0000000000	0.0000000000
29	0.0090090090	0.0000000000	0.0017543860	0.0000000000	0.0000000000
3	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
30	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
31	0.0090090090	0.0625000000	0.0333333333	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0035087719	0.0000000000	0.0000000000
33	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0714285714
34	0.0090090090	0.0000000000	0.0017543860	0.0000000000	0.0000000000
35	0.0180180180	0.0000000000	0.0438596491	0.1538461538	0.0000000000
36	0.0000000000	0.0000000000	0.0052631579	0.0000000000	0.0000000000
37	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
38	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
39	0.0270270270	0.1875000000	0.0140350877	0.0000000000	0.1428571429
4	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
40	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
41	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
42	0.0180180180	0.0625000000	0.0456140351	0.0000000000	0.2142857143
laten2	8	9			
1	0.0000000000	0.0000000000			
10	0.0000000000	0.0724637681			
11	0.2000000000	0.0724637681			
12	0.0000000000	0.0000000000			
13	0.0000000000	0.0000000000			
14	0.0000000000	0.0797101449			
15	0.0000000000	0.0072463768			
16	0.0000000000	0.0000000000			
17	0.0000000000	0.0072463768			
18	0.0000000000	0.0000000000			
19	0.0000000000	0.0000000000			
2	0.0000000000	0.0144927536			

Lanjutan Lampiran 7

```

20 0.0000000000 0.0000000000
21 0.0000000000 0.0072463768
22 0.0000000000 0.0000000000
23 0.0000000000 0.0000000000
24 0.0000000000 0.0000000000
25 0.0000000000 0.0072463768
26 0.0000000000 0.0000000000
27 0.0000000000 0.0000000000
28 0.0000000000 0.0144927536
29 0.0000000000 0.0000000000
3 0.0000000000 0.0000000000
30 0.0000000000 0.0000000000
31 0.0000000000 0.0362318841
32 0.0000000000 0.0000000000
33 0.0000000000 0.0000000000
34 0.0000000000 0.0000000000
35 0.0000000000 0.0362318841
36 0.0000000000 0.0000000000
37 0.0000000000 0.0000000000
38 0.0000000000 0.0000000000
39 0.1000000000 0.0072463768
4 0.0000000000 0.0000000000
40 0.0000000000 0.0000000000
41 0.0000000000 0.0000000000
42 0.1000000000 0.0797101449
[ reached getOption("max.print") -- omitted 40 rows ]

laten3      1      10      11      12      13
1 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0075987842
10 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015197568
11 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
12 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0045592705
13 0.0000000000 0.0000000000 0.0204081633 0.0000000000 0.0106382979
14 0.0000000000 0.0000000000 0.0816326531 0.0000000000 0.0668693009
15 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0136778116
16 0.0000000000 0.0000000000 0.0612244898 0.0000000000 0.0699088146
17 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0030395137
18 0.0000000000 0.0000000000 0.0204081633 0.0000000000 0.0334346505
19 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
2 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0319148936
20 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0136778116
21 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0045592705
22 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015197568
23 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0045592705
24 0.0000000000 0.0000000000 0.0816326531 0.0833333333 0.0197568389
25 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0227963526
26 0.0000000000 0.1000000000 0.0816326531 0.0000000000 0.0577507599
27 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0045592705
28 0.2500000000 0.0000000000 0.0408163265 0.0000000000 0.0653495441
29 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0060790274
3 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0075987842
30 0.0000000000 0.2000000000 0.0000000000 0.0833333333 0.0547112462
31 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
32 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015197568
33 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000
34 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0030395137
35 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0030395137
36 0.0000000000 0.2000000000 0.1224489796 0.0000000000 0.0379939210
37 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015197568
38 0.0000000000 0.0000000000 0.0204081633 0.0000000000 0.0243161094
39 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0045592705
4 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0668693009
40 0.2500000000 0.2000000000 0.0204081633 0.0833333333 0.0455927052
41 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0030395137
42 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0015197568

```

Lanjutan Lampiran 7

laten3	14	15	16	17	18
1	0.0000000000	0.1142857143	0.0169023801	0.0000000000	0.0000000000
10	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
11	0.0000000000	0.0000000000	0.0034494653	0.0000000000	0.0000000000
12	0.0000000000	0.0000000000	0.0113832356	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0186271128	0.0000000000	0.0000000000
14	0.0394736842	0.0285714286	0.0955501897	0.0000000000	0.0000000000
15	0.0131578947	0.0000000000	0.0110382891	0.0000000000	0.0000000000
16	0.0789473684	0.0285714286	0.0645050017	0.0909090909	0.0000000000
17	0.0000000000	0.0285714286	0.0020696792	0.0000000000	0.0000000000
18	0.0263157895	0.0000000000	0.0148327009	0.0909090909	0.0000000000
19	0.0000000000	0.0000000000	0.0013797861	0.0000000000	0.0000000000
2	0.0263157895	0.1428571429	0.0579510176	0.0000000000	0.0000000000
20	0.0000000000	0.0000000000	0.0041393584	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0013797861	0.0000000000	0.0000000000
22	0.0000000000	0.0000000000	0.0034494653	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0055191445	0.0000000000	0.0000000000
24	0.0131578947	0.0000000000	0.0300103484	0.0000000000	0.0000000000
25	0.0131578947	0.0000000000	0.0051741980	0.0000000000	0.0000000000
26	0.1052631579	0.0000000000	0.0669196275	0.0909090909	0.0487804878
27	0.0000000000	0.0000000000	0.0075888237	0.0000000000	0.0000000000
28	0.0789473684	0.0000000000	0.0348395999	0.0909090909	0.0731707317
29	0.0131578947	0.0000000000	0.0027595723	0.0000000000	0.0243902439
3	0.0000000000	0.0000000000	0.0120731287	0.0000000000	0.0000000000
30	0.0131578947	0.0000000000	0.0155225940	0.0909090909	0.0000000000
31	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
33	0.0131578947	0.0285714286	0.0017247327	0.0000000000	0.0000000000
34	0.0263157895	0.0000000000	0.0120731287	0.1818181818	0.0000000000
35	0.0000000000	0.0000000000	0.0027595723	0.0000000000	0.0243902439
36	0.0263157895	0.0000000000	0.0210417385	0.0000000000	0.0243902439
37	0.0000000000	0.0000000000	0.0058640911	0.0000000000	0.0000000000
38	0.0394736842	0.0000000000	0.0255260435	0.0909090909	0.0243902439
39	0.0000000000	0.0285714286	0.0034494653	0.0000000000	0.0000000000
4	0.0263157895	0.2000000000	0.0576060711	0.0909090909	0.0000000000
40	0.0131578947	0.0000000000	0.0179372197	0.0909090909	0.0000000000
41	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
42	0.0000000000	0.0000000000	0.0003449465	0.0000000000	0.0000000000
laten3	19	2	20	21	22
1	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
10	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
11	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
12	0.0000000000	0.0173913043	0.0000000000	0.0000000000	0.0000000000
13	0.0000000000	0.0173913043	0.0000000000	0.0000000000	0.0000000000
14	0.0384615385	0.0869565217	0.0000000000	0.0138888889	0.1250000000
15	0.0128205128	0.0173913043	0.0000000000	0.0000000000	0.0000000000
16	0.0128205128	0.0956521739	0.0338983051	0.0138888889	0.1250000000
17	0.0000000000	0.0260869565	0.0000000000	0.0000000000	0.0000000000
18	0.0128205128	0.0956521739	0.0000000000	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2	0.0512820513	0.0086956522	0.0000000000	0.0000000000	0.0625000000
20	0.0128205128	0.0521739130	0.0000000000	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
22	0.0000000000	0.0086956522	0.0000000000	0.0000000000	0.0000000000
23	0.0000000000	0.0086956522	0.0000000000	0.0138888889	0.0000000000
24	0.0256410256	0.0173913043	0.0169491525	0.0138888889	0.0625000000
25	0.0128205128	0.0086956522	0.0000000000	0.0000000000	0.0000000000
26	0.0769230769	0.0695652174	0.1355932203	0.0694444444	0.0625000000
27	0.0000000000	0.0173913043	0.0169491525	0.0000000000	0.0000000000
28	0.0256410256	0.0434782609	0.0338983051	0.1111111111	0.0625000000
29	0.0128205128	0.0086956522	0.0000000000	0.0000000000	0.0000000000
3	0.0000000000	0.0260869565	0.0000000000	0.0000000000	0.0000000000
30	0.0384615385	0.0434782609	0.0338983051	0.0277777778	0.0000000000
31	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
33	0.0128205128	0.0000000000	0.0000000000	0.0000000000	0.0000000000
34	0.0128205128	0.0608695652	0.0000000000	0.0555555556	0.0000000000
35	0.0128205128	0.0000000000	0.0169491525	0.0000000000	0.0000000000

Lanjutan Lampiran 7

36	0.0128205128	0.0173913043	0.0169491525	0.0833333333	0.0000000000
37	0.0000000000	0.0086956522	0.0169491525	0.0000000000	0.0000000000
38	0.0641025641	0.0086956522	0.0169491525	0.1388888889	0.0000000000
39	0.0000000000	0.0086956522	0.0000000000	0.0000000000	0.0000000000
4	0.0256410256	0.0956521739	0.0000000000	0.0000000000	0.1875000000
40	0.0256410256	0.0434782609	0.0677966102	0.0416666667	0.0000000000
41	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
42	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
laten3	23	24	25	26	27
1	0.0357142857	0.0645161290	0.0000000000	0.0000000000	0.0000000000
10	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
11	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
12	0.0357142857	0.0322580645	0.0000000000	0.0333333333	0.0000000000
13	0.0714285714	0.0483870968	0.0000000000	0.0000000000	0.0000000000
14	0.0357142857	0.0645161290	0.0000000000	0.0000000000	0.0000000000
15	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
16	0.0000000000	0.0161290323	0.0384615385	0.0333333333	0.0000000000
17	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
18	0.0000000000	0.0000000000	0.0384615385	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0192307692	0.0000000000	0.0000000000
2	0.1071428571	0.1935483871	0.0000000000	0.0000000000	0.8000000000
20	0.0000000000	0.0000000000	0.0384615385	0.0000000000	0.0000000000
21	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
22	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
24	0.0357142857	0.0000000000	0.0192307692	0.0000000000	0.0000000000
25	0.0357142857	0.0000000000	0.0000000000	0.0000000000	0.0000000000
26	0.0357142857	0.0000000000	0.0384615385	0.1333333333	0.0000000000
27	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
28	0.0357142857	0.0161290323	0.0384615385	0.0666666667	0.0000000000
29	0.0000000000	0.0161290323	0.0576923077	0.0333333333	0.0000000000
3	0.0000000000	0.0322580645	0.0384615385	0.0000000000	0.0000000000
30	0.0000000000	0.0000000000	0.0769230769	0.1333333333	0.0000000000
31	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
33	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
34	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
35	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
36	0.0000000000	0.0161290323	0.0192307692	0.1000000000	0.0000000000
37	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
38	0.0000000000	0.0000000000	0.0192307692	0.1000000000	0.0000000000
39	0.0357142857	0.0000000000	0.0000000000	0.0000000000	0.0000000000
4	0.1428571429	0.1935483871	0.0961538462	0.0333333333	0.0000000000
40	0.0000000000	0.0000000000	0.0384615385	0.0666666667	0.0000000000
41	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
42	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
laten3	3	4	5	6	7
1	0.0180180180	0.1250000000	0.0105263158	0.0000000000	0.0000000000
10	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
11	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
12	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
13	0.0000000000	0.0000000000	0.0087719298	0.0000000000	0.0714285714
14	0.0720720721	0.0000000000	0.0771929825	0.0000000000	0.0000000000
15	0.0000000000	0.0000000000	0.0070175439	0.0000000000	0.0000000000
16	0.0720720721	0.0000000000	0.0649122807	0.1538461538	0.0000000000
17	0.0000000000	0.0000000000	0.0070175439	0.0000000000	0.0000000000
18	0.0360360360	0.0000000000	0.0210526316	0.0000000000	0.0000000000
19	0.0000000000	0.0000000000	0.0052631579	0.0000000000	0.0000000000
2	0.0360360360	0.0000000000	0.0526315789	0.0000000000	0.0000000000
20	0.0360360360	0.0000000000	0.0087719298	0.0769230769	0.0000000000
21	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
22	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
23	0.0000000000	0.0000000000	0.0070175439	0.0000000000	0.0714285714
24	0.0180180180	0.0000000000	0.0017543860	0.0000000000	0.0000000000
25	0.0000000000	0.0000000000	0.0087719298	0.0000000000	0.0000000000
26	0.0270270270	0.0000000000	0.0263157895	0.0769230769	0.0000000000
27	0.0000000000	0.0000000000	0.0105263158	0.0000000000	0.0000000000

Lanjutan Lampiran 7

28	0.0180180180	0.0000000000	0.0333333333	0.1538461538	0.0000000000
29	0.0000000000	0.0000000000	0.0087719298	0.0000000000	0.0000000000
3	0.0000000000	0.0000000000	0.0263157895	0.0000000000	0.0000000000
30	0.0540540541	0.1250000000	0.0245614035	0.0000000000	0.1428571429
31	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
32	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
33	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
34	0.0000000000	0.0000000000	0.0122807018	0.0000000000	0.0000000000
35	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
36	0.0360360360	0.0000000000	0.0070175439	0.0000000000	0.0000000000
37	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
38	0.0270270270	0.1250000000	0.0175438596	0.0000000000	0.0000000000
39	0.0090090090	0.0000000000	0.0052631579	0.0000000000	0.0000000000
4	0.0450450450	0.0625000000	0.1087719298	0.0769230769	0.1428571429
40	0.0360360360	0.0625000000	0.0280701754	0.0769230769	0.0000000000
41	0.0000000000	0.0000000000	0.0017543860	0.0000000000	0.0000000000
42	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
laten3	8	9			
1	0.0000000000	0.0000000000			
10	0.0000000000	0.0000000000			
11	0.0000000000	0.0000000000			
12	0.0000000000	0.0000000000			
13	0.0000000000	0.0000000000			
14	0.0000000000	0.0000000000			
15	0.0000000000	0.0144927536			
16	0.0000000000	0.0579710145			
17	0.0000000000	0.0000000000			
18	0.1000000000	0.0144927536			
19	0.0000000000	0.0000000000			
2	0.0000000000	0.0000000000			
20	0.0000000000	0.0072463768			
21	0.0000000000	0.0000000000			
22	0.0000000000	0.0000000000			
23	0.0000000000	0.0000000000			
24	0.0000000000	0.0289855072			
25	0.0000000000	0.0000000000			
26	0.1000000000	0.0724637681			
27	0.0000000000	0.0072463768			
28	0.0000000000	0.0869565217			
29	0.0000000000	0.0000000000			
3	0.1000000000	0.0000000000			
30	0.2000000000	0.0652173913			
31	0.0000000000	0.0000000000			
32	0.0000000000	0.0000000000			
33	0.0000000000	0.0000000000			
34	0.0000000000	0.0217391304			
35	0.0000000000	0.0144927536			
36	0.1000000000	0.0217391304			
37	0.0000000000	0.0217391304			
38	0.1000000000	0.0289855072			
39	0.0000000000	0.0072463768			
4	0.0000000000	0.0072463768			
40	0.0000000000	0.0217391304			
41	0.0000000000	0.0000000000			
42	0.0000000000	0.0000000000			
[reached getOption("max.print") -- omitted 61 rows]					

Lampiran 8. *Syntax* Regresi Logistik *Polytomous*

```
#### membaca data
library(haven)
datatrain<- read_spss("D:/tesis/train.sav")

library(haven)
datatest <- read_spss("D:/tesis/test.sav")

####regresi polytomous enter
library(nnet)
M <-multinom(~ moda27 ~ jk + cumur + kawin + pendidikan + kegut + kota +
cjarak + clama + jmoda + cbiaya + cpdrb, data = datatrain,
             MaxNWts=1e5)
summary(M,cor=FALSE, Wald=TRUE)
r=predict(M, datatest)
write.table(r, "D:/tesis/reglogenter.txt", sep="\t")

####regresi polytomous stepwise
stepws=step(M,direction="both")
summary(stepws,cor=FALSE, Wald=TRUE)
s=predict(stepws, datatest)
write.table(s, "D:/tesis/reglogstepwise.txt", sep="\t")

#### AUC
library(haven)
dataku <- read_spss("D:/tesis/hasiltest.sav")
aktual<-dataku$aktual
rgenter<-dataku$rgenter
rgstepwise<-dataku$rgstepwise

library(rms)
rcorr.cens(rgenter,aktual)[1]
rcorr.cens(rgstepwise,aktual)[1]
```

Lampiran 9. Output Regresi Logistik *Polytomous*

a. Metode *Enter*

Call:

```
multinom(formula = moda27 ~ jk + cumur + kawin + pendidikan +
  kegut + kota + cjarak + clama + jmoda + cbiaya + cpdrb, data = datatrain,
  MaxNWts = 1e+05)
```

Coefficients:

	(Intercept)	jk2	cumur2	cumur3	cumur4	
10	-4.073205987	-0.01007526	1.1501424	0.5831775	1.67874237	
11	-1.725164938	-0.21036596	4.3224404	2.3866383	3.25256788	
12	0.294140730	-1.33457861	0.9408253	-1.7395955	-0.01776415	
13	10.085091244	-0.77064358	-0.8723307	-1.3693771	-0.06584946	
14	7.048259798	-0.64321424	-3.0474322	-3.2414660	-2.00569575	
15	12.467176826	-0.90683481	-0.4874155	-3.3208229	-2.08320005	
16	12.513600930	-1.51392898	1.0622981	-0.2068009	0.03883846	
17	-3.287915912	0.08761831	1.1023939	-1.5547135	-1.58210889	
18	-6.349514336	-1.15682290	-0.6848628	-2.2070508	-2.24550399	
19	0.965739949	0.20817278	0.9543520	-0.1056662	0.32928912	
2	0.192202850	-0.24506609	6.9968289	6.1108125	6.87672357	
20	-4.337133444	-0.72574361	-0.6321223	-2.5255466	-1.96686866	
21	-7.793370049	-0.86431976	0.1034188	-1.2277207	-0.83440694	
22	0.003055171	-2.82671498	0.6491340	0.3951963	0.48795551	
23	4.895954333	0.75208511	-0.2305631	-2.0254873	-1.96994388	
24	7.340861648	0.28544616	0.2557165	-1.0029670	-0.80168879	
25	-7.336465103	0.61681806	1.3168660	0.4471279	1.80639046	
26	2.401611340	0.70692632	2.9380186	2.4361152	2.85457631	
27	8.055964457	0.50555403	-2.6904864	-2.1361886	0.95575395	
3	6.569523535	0.02024169	-1.8644646	-3.9715218	-2.93174875	
4	6.187793062	-0.09510396	2.0109053	2.0835204	2.01931078	
5	9.796078991	0.24930349	0.7061178	-0.5636419	0.28576064	
6	-9.233920533	0.76955943	3.4328578	2.4150983	2.12253542	
7	-2.807500466	2.07321659	1.5169494	-0.1199025	-0.22538622	
8	-5.442008064	0.41484318	3.4733217	2.5816309	2.08379915	
9	-2.551420205	-0.13238701	0.4476615	-0.8391244	-0.32722177	
	cumur5	kawin2	pendidikan2	pendidikan3	kegut2	
10	1.489126231	3.1507661	-1.63577202	-0.414098546	-0.25064836	
11	3.370120896	1.4204719	-1.18110180	0.275996283	1.52859893	
12	-0.145614700	-0.1408651	1.84537747	3.291008417	4.20371528	
13	-0.001345756	1.3687741	-1.15594336	1.685862715	-0.75853967	
14	-2.077204706	1.4257272	-0.07914100	2.106576035	2.56304516	
15	-1.670676963	2.1652621	-0.81995299	-0.562176174	-0.96639115	
16	-0.399822713	1.5442308	-0.46255411	0.700110535	0.32540271	
17	-4.574373547	0.7628825	1.25242882	2.605960418	1.36411858	
18	-3.549331293	1.8172945	-0.83870801	1.597567537	5.84705627	
19	0.646929832	1.1026882	-1.04551654	0.129637187	-0.61679483	
2	6.775615363	0.1647812	-0.68637718	0.583725583	-1.42317389	
20	-2.266891711	1.9245871	0.81697491	1.822705221	7.13577664	
21	-1.058846292	1.1919631	-0.85710159	0.881795124	4.78548066	
22	-0.422023426	2.6429889	-0.09208518	1.214042386	5.90021851	
23	-1.770727825	1.3891952	-1.65095165	0.079625277	5.51743954	
24	-0.331444435	0.6866519	-1.99554769	-0.749390197	0.33258472	
25	2.092286567	0.2013202	-1.90241141	-1.087922151	1.41282758	
26	2.502670194	0.4871492	-2.77817947	-0.474989779	-1.00809657	
27	1.672733332	0.1268774	-0.89774361	-2.716446917	-3.95503671	
3	-3.168169943	1.4932304	-2.08575471	-0.547998718	2.90886892	
4	1.879289057	1.3503484	-2.39424573	-2.165292304	0.17177064	
5	0.523153102	0.6451100	-1.55284766	-0.676512297	0.04407557	
6	3.453946825	1.7974015	-1.85663029	1.312571290	-2.27712050	
7	-0.116541361	1.5702743	-2.73902335	-2.478890259	-0.91609229	
8	2.470844501	4.2227378	-2.65496084	-0.900971905	-5.53083067	
9	-0.124682388	0.7915747	-1.30181063	0.007251598	-0.41095070	
	kegut3	kegut4	kota2	cjarak2	cjarak3	cjarak4
10	1.5000176	6.44829265	0.62288392	-0.24153517	1.99092143	2.29927138
11	-0.9893536	1.20800256	0.57105123	3.11857497	2.59750666	3.09655289
12	0.8909058	4.12852060	-2.34559691	-0.25426385	1.68007642	-0.01055099
13	-3.8088318	1.40090218	-0.02147741	0.82246150	-0.90089831	-1.12365225
14	-0.7713412	0.26743176	1.30081476	1.26736861	-0.19280702	-0.30847507

Lanjutan Lampiran 9a

15	-3.3844721	-1.42095923	0.12047749	0.37374720	-5.38380858	-1.06019051
16	-2.4252474	0.66357652	0.74561066	1.51419036	0.34119308	0.23156293
17	-1.1935788	-4.50983285	-0.64386201	0.30588450	-0.01056203	0.65415685
18	2.4679712	-1.24718520	3.88840620	2.49354375	2.43352672	2.44160005
19	-2.5478150	0.75599990	1.79995459	1.50662755	0.57467506	0.64018210
2	-2.5229820	0.69889311	-5.41727162	1.69302555	-0.08157649	-0.21192506
20	3.7305408	5.95022626	0.01235746	2.65318954	3.37034586	3.48343658
21	2.3734252	4.02623574	0.12426674	4.52115834	5.32096020	5.85221212
22	3.6475888	1.49378775	1.23338929	0.88142695	0.05827683	-4.93151550
23	1.6991368	3.74674609	0.55097488	0.78504307	-0.79066520	-1.51343557
24	-3.0431340	0.94909801	0.66615999	0.48351826	-2.74985027	-2.60608347
25	-1.2228735	2.64429844	0.58984393	-0.01057742	-1.62389700	-2.64910161
26	-3.6034707	-1.90682260	-2.19271591	1.29583148	1.41980887	1.45668608
27	-4.9093107	-2.27433272	2.96114284	0.10292249	-0.77325963	-0.46026151
3	-1.9184553	0.18910348	3.81254652	1.25187408	-0.37460080	-0.06742939
4	-2.7846399	1.10092243	0.58986073	0.69306600	-0.86465580	-0.33622248
5	-2.1121285	0.89392459	1.40360936	0.93361972	-0.76996004	-1.03420803
6	-0.7269487	3.72275176	4.69017140	-4.19834768	-5.66323589	-5.32865629
7	-3.5065373	-7.62322094	1.42688701	-0.70733150	-0.46646761	-2.21906428
8	-4.5733689	1.33236938	2.45121697	-1.90331634	-1.92686749	-1.70771198
9	-2.2393848	-0.09296143	0.32690474	2.17695628	2.08902936	1.90477854
	cjarak5	clama2	clama3	clama4	clama5	jmoda2
10	0.9993934	-1.25836248	0.83477652	-1.38549634	0.65855518	-3.3456923
11	3.0007213	4.81173820	3.80863434	1.85256125	1.20906181	-1.9679567
12	1.3721919	0.85262684	-3.06444834	-0.17775598	0.17513329	0.5973969
13	-2.3899245	-0.64226686	-1.75652170	-2.55341769	-2.00766318	-7.0780590
14	-0.9280674	0.78020863	-0.32250100	-1.11901469	-1.51857999	-4.4581213
15	-2.6588538	0.61348885	-0.73698266	-0.97632943	0.74094212	-9.2308571
16	-0.4968870	-0.01245927	-1.25954886	-2.18670355	-2.00159488	-7.5508315
17	-4.2607908	2.79974671	1.52233946	1.25151903	0.91014239	3.7711215
18	1.8643319	1.78967646	3.99603297	2.80952049	2.31591449	4.4986880
19	-0.3572030	0.51746566	-0.58450725	-1.73363297	-1.00594798	6.1496504
2	-2.2384506	1.04507829	-0.06174161	-1.03257333	-0.29095551	-0.6921850
20	4.0078176	-1.01883306	-2.01101672	-3.54855224	-3.21409831	-3.0348280
21	5.7875157	3.55218957	3.10604675	1.96193875	1.27882420	4.5581409
22	-5.7601473	0.33283459	-1.40192287	-2.56558980	-1.39080147	-7.1148433
23	-1.7350633	-0.51941335	-1.96716430	-2.73819741	-2.13656027	-5.1846899
24	-3.0021386	-1.06706047	-3.50788012	-4.16315519	-3.70407453	-4.1741468
25	-4.0645325	5.05696094	4.28946110	3.56172629	4.28478182	3.0424626
26	0.4670870	-2.44341011	-2.32212564	-3.94756920	-3.25312610	2.5460664
27	-2.2687152	-4.01922842	-3.79426082	-3.85781804	-4.64327134	-4.1558937
3	-0.4375624	1.28465071	1.13588364	-0.01271845	1.31370043	-3.7191400
4	-0.8036855	0.53933460	-0.57366584	-0.85821108	0.09436111	-2.3177529
5	-1.9666910	1.03256272	-0.01750974	-1.03773057	-0.35748744	-1.2787120
6	-5.8788292	0.71478936	1.32813976	-0.29087416	0.01674836	-3.4514613
7	-1.4105968	7.07990537	5.56588041	5.81000686	6.76359021	4.9779865
8	-4.0779651	2.69532717	2.25589328	1.43324700	0.96283303	1.1847918
9	2.2660466	3.32526157	2.61564081	1.36889949	1.24694765	3.6252232
	jmoda3	cbiaya2	cbiaya3	cbiaya4	cbiaya5	cpdrb2
10	-3.6214250	0.24085556	4.2359417	-1.13890878	1.4425153	-5.0739818
11	-0.9327220	0.41986508	-1.0649043	-1.68358309	-3.9933591	-1.1490171
12	1.7935154	-0.74523266	2.0931231	0.65502140	-1.6555478	-4.8553509
13	-14.5586137	2.22923535	4.3386472	3.60835265	3.9898721	-0.4873168
14	-12.7157786	1.02335799	1.4465018	0.29681288	0.7555487	-0.5346376
15	-5.4630296	0.04015621	-4.4696777	-2.32005622	-8.2801691	-0.3284201
16	-22.8612306	1.91897353	2.3043771	-0.43712250	-3.5133973	-0.3382782
17	-5.7623617	-1.22600232	-1.1193473	-3.00975166	-9.3387754	-1.3726017
18	-1.9524354	-1.95974268	-3.4821141	-4.80934323	-7.1564195	1.0921741
19	5.9582388	0.11800236	0.1329085	-1.30736885	-3.7772135	-0.6721884
2	-0.7312773	0.92982109	1.7175493	-0.81147169	-3.6621939	1.1784837
20	5.7623611	0.96346910	1.4324170	-1.12368761	-4.7561968	-1.2336346
21	3.8935512	0.30390252	-0.2928008	-2.31484304	-5.9114880	-1.9515986
22	-5.2318908	2.66283549	1.7507539	0.03644297	-2.0407857	-5.2154849
23	-7.0574479	2.95478084	3.6890893	0.38759810	-0.8068630	0.4669244
24	-4.0283019	4.14991073	6.2166858	5.56954734	4.5764792	1.0608237
25	3.2403700	3.43616086	4.8992837	4.71824396	2.6840811	0.6553834
26	4.4866994	3.09408065	3.0083565	1.89290549	-0.1433739	-2.5027068
27	-2.0942195	-2.09671522	-0.6863071	-2.35233910	-4.8377192	-2.6908044

Lanjutan Lampiran 9a

3	-5.8334158	2.60174847	1.9105260	0.86146386	-1.6754390	0.6407556
4	-3.6746068	0.69266304	1.7465864	0.03376820	-1.3315211	-0.5732874
5	-0.7912682	0.94881515	0.9562999	-0.37796856	-3.0647437	0.1357375
6	6.6182006	3.10402954	1.3344526	0.87426004	-0.8965098	0.8648804
7	-1.3701718	-1.97267897	-2.3573785	-4.34110957	-6.3394751	1.8726841
8	3.0053408	0.04886791	4.4649015	3.14405883	0.4883554	1.4777848
9	4.3458659	4.09984634	4.4947231	2.41907596	-0.8986631	-1.0498303
cpdrb3						
10	-1.03365707					
11	-0.68938821					
12	-4.70381265					
13	-1.20243679					
14	-0.81184216					
15	-1.25092689					
16	-0.68091288					
17	2.89446630					
18	0.86439553					
19	-0.20536530					
2	1.12929717					
20	-0.32079663					
21	-0.39241425					
22	-0.54197599					
23	-0.70289901					
24	-0.23955473					
25	0.87324805					
26	-1.71197397					
27	2.78327106					
3	2.08373689					
4	-0.81172442					
5	-0.05332453					
6	5.92460746					
7	0.52670369					
8	2.24646860					
9	-0.57736902					

Residual Deviance: 8846.387
AIC: 10354.39

b. Metode Stepwise

Call:
multinom(formula = moda27 ~ jk + cumur + pendidikan + kegut +
kota + cjarak + jmoda + cbiaya, data = datatrain, MaxNWts = 1e+05)

Coefficients:

	(Intercept)	jk2	cumur2	cumur3	cumur4	cumur5
10	-6.16012953	-0.15773914	1.8706308	2.8752276	4.64455523	4.15961198
11	1.10985797	-0.22031251	5.9339555	4.2517470	5.69348941	5.62576988
12	-4.85882816	-1.38317342	0.3065452	-2.5624878	-1.07490277	-1.41530408
13	11.49988076	-0.86149836	-1.4324536	-1.6470333	0.13298394	0.02826185
14	9.02953268	-0.73561035	-3.5151828	-3.2987216	-1.50250889	-1.78594099
15	14.16906695	-0.90106859	-0.9395239	-2.9942032	-1.09276506	-0.76786477
16	14.33297517	-1.53652713	0.4730145	-0.3719329	0.50380248	-0.08962806
17	-0.07616781	-0.09712111	1.1057972	-1.4806959	-1.11974482	-5.85690673
18	-6.27753936	-1.03993144	-2.3746448	-3.3987924	-2.84322074	-4.05786183
19	0.64848961	0.11388796	0.2670994	-0.5734131	0.25205288	0.39687466
2	1.88946601	-0.28967806	8.9042621	7.7897541	8.68054262	8.34483003
20	-7.23200987	-0.78119668	-2.1763433	-3.4829816	-2.42233214	-3.12236411
21	-8.53374448	-0.94586414	0.2685357	-0.9217152	0.03493558	-0.44832194
22	1.26632299	-2.78789380	0.8719107	1.3478671	2.36913298	1.33445518
23	6.00191631	0.68955450	-0.9044977	-2.3058694	-1.76644281	-1.81010622
24	10.05590897	0.04107015	-0.4177187	-1.8941392	-1.33921327	-0.94854911
25	-6.86656778	0.57522532	0.7973126	-0.2020933	1.30420907	1.29888637
26	1.24333074	0.62785998	2.4545652	2.0516418	2.65714058	1.99623270

Lanjutan Lampiran 9b

27	11.64937824	0.22413288	-3.7151421	-3.7917460	0.85405525	1.52289908
3	11.68938423	-0.06738912	-2.1199072	-3.8132890	-2.14951984	-2.46303582
4	7.79910969	-0.11847019	1.5962450	2.0937434	2.62504282	2.32130597
5	12.48347034	0.16582188	0.2204222	-1.0968945	0.15238584	0.16689918
6	-5.10358950	1.12092278	3.8229390	3.6251108	3.54333196	5.16495350
7	2.44495532	1.62430359	1.5646288	0.9143261	1.57217899	1.20803272
8	-1.48483323	0.38267002	4.6527563	5.0443138	5.58308912	5.74942355
9	-3.88703553	-0.15760789	0.1909754	-1.1100821	-0.18653169	-0.29134727
	pendidikan2	pendidikan3	kegut2	kegut3	kegut4	kota2
10	-2.0595497	-0.571812808	-1.44324924	2.5381975	7.7270828	1.2623920
11	-1.2667920	0.085056719	2.24152480	-1.0791547	1.2182965	1.1774012
12	-2.5001332	4.274288454	5.18589202	0.9592471	4.7201215	1.8436157
13	-1.1688484	1.568032436	0.01453048	-3.6454428	1.3796733	1.1925848
14	-0.2409019	1.862223080	3.37695638	-0.5933864	0.2510467	2.1793198
15	-0.9420058	-0.738909370	0.05229963	-3.2683102	-1.5907297	1.4024726
16	-0.5331314	0.576142060	1.08014009	-2.3495474	0.6906567	1.5674094
17	2.6013102	4.004114568	2.57737372	-1.1321789	-4.1379893	-4.4782587
18	-0.5486445	1.330715398	5.96805520	2.2937499	-1.9224426	3.5154576
19	-1.1582209	0.035316258	-0.04302417	-2.6698445	0.7764436	2.0328393
2	-0.6555581	0.657010334	-0.87371169	-2.5443776	0.8244126	-8.6863942
20	0.6075323	1.506421133	8.57984985	4.1923647	6.2275565	0.2350008
21	-0.9100518	0.841391542	5.73387843	2.5357774	4.6553469	0.1872857
22	-0.1896627	1.113641738	7.25913789	4.0505934	2.5242973	1.8886891
23	-1.7831296	-0.095210448	7.18426690	2.6607081	4.5498314	1.4039413
24	-2.1117129	-0.618789377	0.72838824	-3.2261740	0.8094333	1.2688433
25	-1.8481977	-1.052887849	1.75842521	-1.3505229	2.4237303	2.4799292
26	-3.0458050	-0.610191029	-0.57575985	-3.8580464	-1.5076844	-0.6499282
27	-1.1171834	-2.946999770	-5.70345196	-4.6661166	-0.8532351	0.7738417
3	-2.0846331	-0.587449843	3.56304363	-1.9389007	0.1497849	2.0627431
4	-2.4133295	-2.244751004	1.02639026	-2.6917752	1.2961097	1.4175880
5	-1.5594862	-0.711133004	0.75812461	-2.0115342	1.0409894	1.6384073
6	-1.4230752	2.029325569	-3.05001198	0.5871993	5.5430693	-0.3791342
7	-2.1637308	-1.696736383	-0.14831411	-3.3191665	-8.0855477	2.1521627
8	-2.6565056	-1.157459212	-7.91260058	-4.1394923	0.3631601	1.0529915
9	-1.3405032	-0.004149994	0.07607245	-2.4058447	-0.1045791	0.8077590
	cjarak2	cjarak3	cjarak4	cjarak5	jmoda2	jmoda3
10	0.19612975	2.86063406	3.1664086	1.90328537	-4.5143296	-5.765857
11	4.86196840	3.50025074	3.6932081	3.22848850	-3.3148744	-3.082099
12	-0.54150029	2.24453395	0.4587589	1.79143952	0.8693834	1.403083
13	0.94041941	-1.40309143	-1.8260153	-2.99914079	-7.9504049	-31.434883
14	1.51058778	-0.69165503	-1.1262754	-1.80617068	-5.7392730	-15.625477
15	0.78925311	-6.60972930	-1.0566699	-2.03439951	-10.9487528	-8.115739
16	1.63091713	-0.27803954	-0.5944943	-1.41890494	-8.2083419	-9.229607
17	1.08870022	0.07208390	0.6492770	-5.63907972	3.9399628	-2.773087
18	6.66944266	6.04939416	5.7466443	5.03895672	8.2676236	-2.532385
19	1.83022495	0.23023965	0.0902596	-1.02719867	7.9366049	7.177985
2	2.04875711	-0.32204088	-0.6673219	-2.81335191	-1.8857665	-2.519096
20	3.92399344	4.15734071	4.0494187	4.41123853	-5.5726943	7.526579
21	6.86395182	6.79035092	7.2162646	6.80473182	6.7748481	5.456518
22	0.93499718	-0.77703236	-7.2705188	-8.48695028	-8.9319573	-9.240591
23	0.70215555	-1.45816257	-2.3198607	-2.53802306	-6.2610292	-11.348050
24	0.14696084	-3.95719686	-4.1419316	-4.35320915	-5.1727452	-6.147923
25	0.60882186	-1.42778945	-2.5414794	-4.02860999	6.9789979	6.786432
26	1.27471496	0.60636011	0.5639719	-0.66048520	1.6522212	2.951270
27	-2.16638181	-3.12237415	-3.1976124	-6.58853607	-5.0938943	-3.198458
3	2.04821080	-0.04715467	0.3820451	0.00107343	-4.8628375	-7.331659
4	1.02310004	-0.89169688	-0.3347109	-0.81271933	-3.4899334	-5.277305
5	1.32251980	-1.02873146	-1.4221562	-2.44580552	-2.3940217	-2.403357
6	-3.52018128	-6.06032120	-6.5344944	-6.64420141	-4.7486182	5.537292
7	0.01700368	0.16060916	-1.8324826	-1.16746558	5.8714489	-1.187084
8	-0.74662237	-1.71597628	-1.6503566	-4.09972078	0.8780406	1.974206
9	2.75403547	1.85518758	1.4316717	1.65845740	4.8651230	4.994725
	cbiaya2	cbiaya3	cbiaya4	cbiaya5		
10	1.78598984	6.08498867	-1.91901563	2.4944732		
11	0.77429437	-0.90147042	-2.37025477	-5.0258128		
12	-0.48970393	2.84436668	0.86187422	-1.7702173		
13	2.24804296	3.85547411	2.55346377	2.4207507		
14	1.47604013	1.77257923	0.01397481	0.1911876		
15	0.37349212	-4.18660464	-2.46933700	-10.7787513		

Lanjutan Lampiran 9b

16	2.10115867	2.21351692	-1.01038759	-4.4046170
17	-0.04241394	-0.22596289	-2.95843362	-11.6840705
18	-1.53292266	-3.03356678	-5.02804803	-7.6896850
19	0.50412852	0.25494920	-1.77649187	-4.4744880
2	1.30547824	1.86888512	-1.18368733	-4.2937633
20	1.07696311	1.34743496	-2.03795813	-5.9439835
21	0.60415193	-0.07496355	-2.79655536	-6.7573705
22	2.84609122	1.79628012	-0.35132727	-2.7000126
23	2.91923077	3.31694878	-0.37853917	-2.1640352
24	3.97056673	5.49215721	4.10572976	2.4973079
25	5.99894669	7.32388982	6.59629279	4.3676871
26	3.55908940	3.25066388	1.61586573	-0.6925609
27	-2.88218806	-2.22712326	-4.46537481	-8.5143168
3	3.12334822	2.34604769	0.75009981	-2.1223489
4	0.98079372	2.12502002	-0.10351732	-1.8846818
5	1.37004248	1.18755630	-0.72741054	-3.6988703
6	4.58822133	2.70961253	1.84002886	-0.1055850
7	-0.66233560	-0.88373529	-3.53412527	-5.7302104
8	0.67839997	6.08240428	3.94376743	1.1155296
9	7.90696710	8.17157235	5.40890243	1.7428400

Residual Deviance: 9134.293

AIC: 10278.29

Lampiran 10. *Syntax Naïve Bayes dengan Input Regresi Logistik Polytomous*

```
#### membaca data
library(haven)
datatrain <- read_spss("D:/tesis/train99.sav")

library(haven)
datatest <- read_spss("D:/tesis/test.sav")

####regresi polytomous stepwise
stepws=step(M,direction="both")
summary(stepws,cor=FALSE, Wald=TRUE))

####NB
library(naivebayes)
rnb <- naive_bayes(modat27 ~ jk + umur + pendidikan +
  kegut + kota + cjarak + jmoda + cbiaya, data = datatrain)
predict(rnb)
d=predict(rnb, newdata= datatest)
write.table(d, "D:/tesis/stepwise_nb.txt", sep="\t")

####Struktur
attach(datatrain)
library(phylopath)
models <- define_model_set(
  c(jk~modat27,
    umur~modat27,
    pendidikan~modat27,
    kegut~modat27,
    kota~modat27,
    cjarak~modat27,
    jmoda~modat27,
    cbiaya~modat27)
)
plot_model_set(models, algorithm = 'kk', text_size=4, edge_width=1)

#### AUC
aktual<-dataku$aktual
rgnb<-dataku$rgnb

library(rms)
rcorr.cens(rgnb,aktual)[1]
```

Lampiran 11. Output *Naïve Bayes* dengan Input Regresi Logistik *Polytomous*

```

===== Naive Bayes =====

Call:
naive_bayes.formula(formula = moda27 ~ jk + cumur + pendidikan +
  kegut + kota + cjarak + jmoda + cbiaya, data = datatrain)

A priori probabilities:

      1      10      11      12      13
0.0015420200 0.0019275251 0.0094448728 0.0023130301 0.1268311488
      14      15      16      17      18
0.0146491904 0.0067463377 0.5587895143 0.0021202776 0.0079028527
      19      2      20      21      22
0.0150346955 0.0221665382 0.0113723978 0.0138781804 0.0030840401
      23      24      25      26      27
0.0053970702 0.0119506554 0.0100231303 0.0057825752 0.0009637625
      3      4      5      6      7
0.0213955281 0.0030840401 0.1098689283 0.0025057826 0.0026985351
      8      9
0.0019275251 0.0265998458

Tables:

jk      1      10      11      12      13      14      15
  1 0.5000000 0.6000000 0.5306122 0.6666667 0.7006079 0.6842105 0.6857143
  2 0.5000000 0.4000000 0.4693878 0.3333333 0.2993921 0.3157895 0.3142857

jk      16      17      18      19      2      20      21
  1 0.7961366 0.2727273 0.6585366 0.4230769 0.4260870 0.6440678 0.6666667
  2 0.2038634 0.7272727 0.3414634 0.5769231 0.5739130 0.3559322 0.3333333

jk      22      23      24      25      26      27      3
  1 0.9375000 0.2500000 0.3870968 0.3269231 0.3333333 0.6000000 0.4954955
  2 0.0625000 0.7500000 0.6129032 0.6730769 0.6666667 0.4000000 0.5045045

jk      4      5      6      7      8      9
  1 0.6250000 0.3964912 0.2307692 0.2142857 0.4000000 0.5289855
  2 0.3750000 0.6035088 0.7692308 0.7857143 0.6000000 0.4710145

cumur      1      10      11      12      13      14
  1 0.00000000 0.00000000 0.00000000 0.00000000 0.09422492 0.01315789
  2 0.12500000 0.10000000 0.32653061 0.58333333 0.08510638 0.02631579
  3 0.37500000 0.20000000 0.18367347 0.08333333 0.15957447 0.19736842
  4 0.25000000 0.40000000 0.24489796 0.16666667 0.28571429 0.34210526
  5 0.25000000 0.30000000 0.24489796 0.16666667 0.37537994 0.42105263

cumur      15      16      17      18      19      2
  1 0.22857143 0.04691273 0.00000000 0.00000000 0.03846154 0.00000000
  2 0.22857143 0.28527078 0.63636364 0.29268293 0.32051282 0.38260870
  3 0.08571429 0.29837875 0.27272727 0.46341463 0.25641026 0.29565217
  4 0.17142857 0.23283891 0.09090909 0.17073171 0.15384615 0.18260870
  5 0.28571429 0.13659883 0.00000000 0.07317073 0.23076923 0.13913043

cumur      20      21      22      23      24      25
  1 0.00000000 0.01388889 0.00000000 0.25000000 0.43548387 0.03846154
  2 0.22033898 0.22222222 0.06250000 0.32142857 0.32258065 0.30769231
  3 0.20338983 0.29166667 0.43750000 0.21428571 0.09677419 0.17307692
  4 0.32203390 0.26388889 0.37500000 0.10714286 0.04838710 0.21153846
  5 0.25423729 0.20833333 0.12500000 0.10714286 0.09677419 0.26923077

cumur      26      27      3      4      5      6
  1 0.00000000 0.40000000 0.28828829 0.06250000 0.14385965 0.00000000
  2 0.16666667 0.00000000 0.16216216 0.18750000 0.35087719 0.23076923
  3 0.36666667 0.00000000 0.12612613 0.31250000 0.17192982 0.30769231
  4 0.26666667 0.20000000 0.19819820 0.18750000 0.16315789 0.15384615
  5 0.20000000 0.40000000 0.22522523 0.25000000 0.17017544 0.30769231

```

Lanjutan Lampiran 11

cumur	7	8	9			
1	0.00000000	0.00000000	0.01449275			
2	0.21428571	0.20000000	0.27536232			
3	0.42857143	0.30000000	0.24637681			
4	0.21428571	0.20000000	0.21739130			
5	0.14285714	0.30000000	0.24637681			
pendidikan	1	10	11	12	13	
1	0.00000000	0.20000000	0.16326531	0.00000000	0.14893617	
2	0.50000000	0.30000000	0.44897959	0.50000000	0.18237082	
3	0.50000000	0.50000000	0.38775510	0.50000000	0.66869301	
pendidikan	14	15	16	17	18	
1	0.03947368	0.57142857	0.20593308	0.00000000	0.04878049	
2	0.26315789	0.37142857	0.56122801	0.54545455	0.58536585	
3	0.69736842	0.05714286	0.23283891	0.45454545	0.36585366	
pendidikan	19	2	20	21	22	
1	0.16666667	0.10434783	0.03389831	0.09722222	0.12500000	
2	0.50000000	0.51304348	0.55932203	0.44444444	0.56250000	
3	0.33333333	0.38260870	0.40677966	0.45833333	0.31250000	
pendidikan	23	24	25	26	27	
1	0.42857143	0.66129032	0.19230769	0.16666667	0.80000000	
2	0.28571429	0.19354839	0.48076923	0.13333333	0.20000000	
3	0.28571429	0.14516129	0.32692308	0.70000000	0.00000000	
pendidikan	3	4	5	6	7	
1	0.46846847	0.50000000	0.39122807	0.07692308	0.28571429	
2	0.24324324	0.37500000	0.41578947	0.23076923	0.50000000	
3	0.28828829	0.12500000	0.19298246	0.69230769	0.21428571	
pendidikan	8	9				
1	0.30000000	0.16666667				
2	0.30000000	0.45652174				
3	0.40000000	0.37681159				
kegut	1	10	11	12	13	14
1	0.00000000	0.00000000	0.02040816	0.00000000	0.15349544	0.01315789
2	0.00000000	0.00000000	0.10204082	0.16666667	0.11702128	0.22368421
3	1.00000000	0.90000000	0.79591837	0.50000000	0.56382979	0.75000000
4	0.00000000	0.10000000	0.08163265	0.33333333	0.16565350	0.01315789
kegut	15	16	17	18	19	2
1	0.17142857	0.05657123	0.00000000	0.00000000	0.05128205	0.02608696
2	0.02857143	0.05001725	0.09090909	0.12195122	0.05128205	0.01739130
3	0.51428571	0.73301138	0.90909091	0.87804878	0.64102564	0.77391304
4	0.28571429	0.16040014	0.00000000	0.00000000	0.25641026	0.18260870
kegut	20	21	22	23	24	25
1	0.00000000	0.00000000	0.00000000	0.00000000	0.04838710	0.01923077
2	0.23728814	0.11111111	0.06250000	0.10714286	0.03225806	0.09615385
3	0.69491525	0.83333333	0.93750000	0.57142857	0.25806452	0.63461538
4	0.06779661	0.05555556	0.00000000	0.32142857	0.66129032	0.25000000
kegut	26	27	3	4	5	6
1	0.13333333	0.40000000	0.02702703	0.12500000	0.03333333	0.00000000
2	0.13333333	0.00000000	0.24324324	0.06250000	0.03684211	0.00000000
3	0.70000000	0.20000000	0.41441441	0.62500000	0.59824561	0.84615385
4	0.03333333	0.40000000	0.31531532	0.18750000	0.33157895	0.15384615
kegut	7	8	9			
1	0.07142857	0.20000000	0.05072464			
2	0.07142857	0.00000000	0.05072464			
3	0.85714286	0.60000000	0.79710145			
4	0.00000000	0.20000000	0.10144928			

Lanjutan Lampiran 11

kota	1	10	11	12	13	14
1	0.87500000	0.80000000	0.79591837	0.66666667	0.80851064	0.68421053
2	0.12500000	0.20000000	0.20408163	0.33333333	0.19148936	0.31578947
kota	15	16	17	18	19	2
1	0.65714286	0.70507071	1.00000000	0.26829268	0.62820513	1.00000000
2	0.34285714	0.29492929	0.00000000	0.73170732	0.37179487	0.00000000
kota	20	21	22	23	24	25
1	0.93220339	0.90277778	0.68750000	0.71428571	0.70967742	0.61538462
2	0.06779661	0.09722222	0.31250000	0.28571429	0.29032258	0.38461538
kota	26	27	3	4	5	6
1	0.96666667	0.80000000	0.62162162	0.68750000	0.68596491	0.84615385
2	0.03333333	0.20000000	0.37837838	0.31250000	0.31403509	0.15384615
kota	7	8	9			
1	0.64285714	0.80000000	0.83333333			
2	0.35714286	0.20000000	0.16666667			
cjarak	1	10	11	12	13	14
1	0.00000000	0.00000000	0.02040816	0.00000000	0.20060790	0.10526316
2	0.00000000	0.00000000	0.20408163	0.00000000	0.25531915	0.19736842
3	0.25000000	0.30000000	0.26530612	0.33333333	0.26443769	0.27631579
4	0.25000000	0.40000000	0.24489796	0.08333333	0.15349544	0.18421053
5	0.50000000	0.30000000	0.26530612	0.58333333	0.12613982	0.23684211
cjarak	15	16	17	18	19	2
1	0.74285714	0.26629872	0.18181818	0.00000000	0.11538462	0.19130435
2	0.17142857	0.31390135	0.18181818	0.12195122	0.19230769	0.40869565
3	0.00000000	0.22731977	0.27272727	0.39024390	0.26923077	0.22608696
4	0.05714286	0.12211107	0.36363636	0.26829268	0.25641026	0.14782609
5	0.02857143	0.07036909	0.00000000	0.21951220	0.16666667	0.02608696
cjarak	20	21	22	23	24	25
1	0.00000000	0.00000000	0.37500000	0.57142857	0.67741935	0.21153846
2	0.03389831	0.04166667	0.31250000	0.21428571	0.24193548	0.25000000
3	0.25423729	0.29166667	0.31250000	0.14285714	0.03225806	0.30769231
4	0.15254237	0.31944444	0.00000000	0.03571429	0.01612903	0.13461538
5	0.55932203	0.34722222	0.00000000	0.03571429	0.03225806	0.09615385
cjarak	26	27	3	4	5	6
1	0.03333333	1.00000000	0.25225225	0.25000000	0.39473684	0.15384615
2	0.06666667	0.00000000	0.26126126	0.12500000	0.27368421	0.23076923
3	0.36666667	0.00000000	0.16216216	0.12500000	0.15964912	0.23076923
4	0.26666667	0.00000000	0.16216216	0.18750000	0.09824561	0.15384615
5	0.26666667	0.00000000	0.16216216	0.31250000	0.07368421	0.23076923
cjarak	7	8	9			
1	0.28571429	0.20000000	0.02173913			
2	0.07142857	0.10000000	0.10869565			
3	0.42857143	0.30000000	0.28985507			
4	0.07142857	0.30000000	0.18115942			
5	0.14285714	0.10000000	0.39855072			
jmoda	1	10	11	12	13	
1	0.0000000000	0.6000000000	0.4897959184	0.0000000000	0.9863221884	
2	0.6250000000	0.3000000000	0.2857142857	0.3333333333	0.0136778116	
3	0.3750000000	0.1000000000	0.2244897959	0.6666666667	0.0000000000	
jmoda	14	15	16	17	18	
1	0.8947368421	1.0000000000	0.9948258020	0.0000000000	0.0000000000	
2	0.1052631579	0.0000000000	0.0044843049	1.0000000000	1.0000000000	
3	0.0000000000	0.0000000000	0.0006898931	0.0000000000	0.0000000000	
jmoda	19	2	20	21	22	
1	0.0000000000	0.2260869565	0.0000000000	0.0000000000	1.0000000000	
2	0.7948717949	0.6347826087	0.0000000000	0.9027777778	0.0000000000	
3	0.2051282051	0.1391304348	1.0000000000	0.0972222222	0.0000000000	

Lanjutan Lampiran 11

jmoda	23	24	25	26	27	
1	0.9642857143	0.8870967742	0.0000000000	0.0000000000	1.0000000000	
2	0.0357142857	0.0967741935	0.6153846154	0.2666666667	0.0000000000	
3	0.0000000000	0.0161290323	0.3846153846	0.7333333333	0.0000000000	
jmoda	3	4	5	6	7	
1	0.8468468468	0.5625000000	0.4105263158	0.0000000000	0.0000000000	
2	0.1441441441	0.3750000000	0.4263157895	0.0000000000	1.0000000000	
3	0.0090090090	0.0625000000	0.1631578947	1.0000000000	0.0000000000	
jmoda	8	9				
1	0.0000000000	0.0000000000				
2	0.4000000000	0.6014492754				
3	0.6000000000	0.3985507246				
cbiaya	1	10	11	12	13	14
1	0.00000000	0.00000000	0.20408163	0.00000000	0.02279635	0.03947368
2	0.00000000	0.00000000	0.34693878	0.00000000	0.02583587	0.05263158
3	0.00000000	0.20000000	0.06122449	0.16666667	0.03799392	0.02631579
4	0.12500000	0.00000000	0.14285714	0.25000000	0.07750760	0.03947368
5	0.87500000	0.80000000	0.24489796	0.58333333	0.83586626	0.84210526
cbiaya	15	16	17	18	19	2
1	0.77142857	0.19868920	0.09090909	0.56097561	0.06410256	0.08695652
2	0.20000000	0.50051742	0.36363636	0.24390244	0.26923077	0.29565217
3	0.00000000	0.20386340	0.36363636	0.07317073	0.23076923	0.40869565
4	0.02857143	0.06760952	0.18181818	0.04878049	0.17948718	0.12173913
5	0.00000000	0.02932046	0.00000000	0.07317073	0.25641026	0.08695652
cbiaya	20	21	22	23	24	25
1	0.00000000	0.06944444	0.12500000	0.21428571	0.04838710	0.00000000
2	0.05084746	0.40277778	0.68750000	0.50000000	0.19354839	0.05769231
3	0.40677966	0.27777778	0.06250000	0.17857143	0.17741935	0.15384615
4	0.32203390	0.16666667	0.06250000	0.03571429	0.25806452	0.32692308
5	0.22033898	0.08333333	0.06250000	0.07142857	0.32258065	0.46153846
cbiaya	26	27	3	4	5	6
1	0.00000000	1.00000000	0.13513514	0.18750000	0.22807018	0.00000000
2	0.06666667	0.00000000	0.46846847	0.12500000	0.33157895	0.07692308
3	0.10000000	0.00000000	0.09909910	0.18750000	0.16666667	0.15384615
4	0.23333333	0.00000000	0.15315315	0.12500000	0.14912281	0.23076923
5	0.60000000	0.00000000	0.14414414	0.37500000	0.12456140	0.53846154
cbiaya	7	8	9			
1	0.21428571	0.00000000	0.00000000			
2	0.35714286	0.00000000	0.20289855			
3	0.21428571	0.40000000	0.38405797			
4	0.07142857	0.30000000	0.23188406			
5	0.14285714	0.30000000	0.18115942			

BIOGRAFI PENULIS



Penulis dilahirkan di Magelang, Jawa Tengah pada tanggal 4 Mei 1988 merupakan anak pertama dari dua bersaudara pasangan Bapak Bambang Supriyanto dan Ibu Sri Istati. Penulis menyelesaikan pendidikan di SDN Bumirejo II (1994-2000), SMPN 1 Kota Magelang (1990-2003), SMAN 1 Kota Magelang (2003-2006) dan Sekolah Tinggi Ilmu Statistik Jurusan Statistika Peminatan Sosial dan Kependudukan (2006-2010). Setelah lulus DIV, penulis ditempatkan pada Direktorat Diseminasi Statistik, Badan Pusat Statistik. Pada bulan Juli 2017, penulis diberikan kesempatan untuk melanjutkan pendidikan S2 di Institut Teknologi Sepuluh Nopember (ITS) Surabaya Jurusan Statistika dengan beasiswa APBN BPS. Jika ingin mendiskusikan lebih lanjut mengenai tesis ini, penulis dapat dihubungi melalui alamat email ratihkd@bps.go.id.

Surabaya, 27 Januari 2018

Ratih Kusuma Dewi