KLASIFIKASI JANGKA WAKTU MENDAPATKAN KERJA DI DKI JAKARTA DENGAN METODE *BOOTSTRAP AGGREGATING* REGRESI LOGISTIK ORDINAL

Skripsi

Disusun untuk melengkapi syarat-syarat

guna memperoleh gelar Sarjana Sains

SHARAH ANNISA

3125121986

PROGRAM STUDI MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI JAKARTA

2017

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

KLASIFIKASI JANGKA WAKTU MENDAPATKAN KERJA DI DKI JAKARTA DENGAN METODE *BOOTSTRAP AGGREGATING* REGRESI LOGISTIK ORDINAL

Nama : Sharah Annisa

No. Registrasi : 3125121986

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: Prof. Dr. Suyono, M.Si.	.....	.....
	NIP. 19671218 199303 1 005		
Wakil Penanggung Jawab			
Wakil Dekan I	: Dr. Muktiningsih, M.Si.	.....	.....
	NIP. 19640511 198903 2 001		
Ketua	: Dr. Lukita Ambarwati, S.Pd, M.Si.	.....	.....
	NIP. 19721026 200112 2 001		
Sekretaris	: Dr. Eti Dwi Wiraningsih, S.Pd, M.Si.	.....	.....
	NIP. 19810203 200604 2 001		
Penguji Ahli	: Ir. Fariani Hermin, M.T.	.....	.....
	NIP. 19600211 198703 2 001		
Pembimbing I	: Dra. Widyanti Rahayu, M.Si.	.....	.....
	NIP. 19661103 200112 2 001		
Pembimbing II	: Vera Maya Santi, M.Si.	.....	.....
	NIP. 19790531 200501 2 006		

Dinyatakan lulus ujian skripsi tanggal: 10 Agustus 2017

ABSTRACT

SHARAH ANNISA, 3125121986. Classification of Time Phase to Getting a Job in DKI Jakarta with Method of Bootstrap Aggregating Ordinal Logistic Regression. Thesis. Faculty of Mathematics and Natural Science, Jakarta State University. 2017.

Ordinal logistic regression is a method to know the relationship between the independent variables with categorical dependent variables. This method also is used as a classification process. However, this classification method provides unstable parameter estimation, so to obtain a stable parameter in ordinal logistic regression model used bootstrap aggregating (bagging). The purpose of this study is to find some factors that affect the time phase to getting a job in DKI Jakarta and to count classification accuracy. The result of this study indicates that the variables of age, marriage status, last education level, and last salary have an effect on the time phase to getting a job in DKI Jakarta with value of classification accuracy in ordinal logistic regression is 57.33 %. While the value of classification accuracy in bagging ordinal logistic regression is 61.33 %. It is mean bagging can increase the value of classification accuracy.

Keywords *: ordinal logistic regression, bootstrap aggregating (bagging), classification accuracy, time phase to getting a job.*

ABSTRAK

SHARAH ANNISA, 3125121986. Klasifikasi Jangka Waktu Mendapatkan Kerja di DKI Jakarta dengan Metode *Bootstrap Aggregating* Regresi Logistik Ordinal. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2017.

Regresi logistik ordinal merupakan metode untuk mengetahui hubungan antara peubah penjelas dengan peubah respon yang bersifat kategorik bertingkat. Metode ini juga dapat digunakan sebagai proses pengklasifikasian. Namun metode pengklasifikasian ini memberikan pendugaan parameter yang tidak stabil, untuk memperoleh parameter yang stabil pada model regresi logistik ordinal digunakan metode *bootstrap aggregating (bagging)*. Ketepatan klasifikasi regresi logistik juga dapat ditingkatkan dengan metode *bootstrap aggregating (bagging)*. Tujuan dari penelitian ini adalah untuk mengetahui faktor-faktor apa saja yang berpengaruh terhadap jangka waktu mendapat kerja di DKI Jakarta dan menghitung ketepatan klasifikasi. Hasil penelitian ini menunjukkan bahwa peubah umur, status pernikahan, tingkat pendidikan terakhir, dan penghasilan sebelumnya berpengaruh terhadap jangka waktu mendapat kerja di DKI Jakarta dengan nilai ketepatan klasifikasi regresi logistik ordinal yaitu sebesar 57.33 %. Sedangkan nilai ketepatan klasifikasi *bagging* regresi logistik ordinal yaitu sebesar 61.33 %. Hal ini berarti *bagging* dapat meningkatkan nilai ketepatan klasifikasi.

Kata kunci : regresi logistik ordinal, *bootstrap aggregating (bagging)*, ketepatan klasifikasi, jangka waktu mendapatkan kerja.

PERSEMBAHANKU...

" Dan orang yang bersungguh-sungguh untuk (mencari keridhaan) Kami, Kami akan tunjukkan kepada mereka jalan-jalan Kami. Dan sungguh ALLAH beserta orang-orang yang berbuat baik. (QS. Al - 'Ankabut : 69) "

" Tenag, Resapi, Tuntaskan "

-M. Afif Makarim, S.Pd.

Skripsi ini kupersembahkan untuk Papa, Mama, dan kedua Kakakku. *"Terima kasih atas dukungan, do'a, serta kasih sayang kalian".*

KATA PENGANTAR

Puji syukur kepada Allah Azza wa jalla atas berkat rahmat, hidayah, dan limpahan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Klasifikasi Jangka Waktu Mendapatkan Kerja di DKI Jakarta dengan Metode *Bootstrap Aggregating* Regresi Logistik Ordinal" yang merupakan salah satu syarat dalam memperoleh gelar Sarjana Program Studi Matematika Universitas Negeri Jakarta. Sholawat serta salam penulis sampaikan kepada manusia pilihan Allah yang menjadi tauladan sepanjang masa dan sempurna budi pekertinya, Nabi Muhammad SAW, beserta keluarga, sahabat, dan umatnya.

Terselesaikannya skripsi ini tidak terlepas dari adanya bantuan berbagai pihak yang telah banyak memberikan arahan, bantuan pemikiran, doa, semangat dan lain sebagainya. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan banyak terima kasih kepada:

1. Mama dan Papa yang senantiasa memberikan perhatian, selalu memberikan do'a, dukungan semangat yang luar biasa, nasihat serta motivasi dengan penuh cinta dan kasih sayang yang tulus sehingga penulis dapat menyelesaikan penulisan skripsi ini.

2. Ibu Dra. Widyanti Rahayu, M.Si., selaku Dosen Pembimbing I dan Ibu Vera Maya Santi, M.Si., selaku Dosen Pembimbing II, yang telah meluangkan waktu dan dengan sabar memberikan pengarahan, bimbingan, saran-saran serta dukungannya sehingga skripsi ini dapat terselesaikan dengan baik.

3. Ibu Dr. Lukita Ambarwati, S.Pd, M.Si., selaku Koordinator Program Studi Matematika FMIPA UNJ yang telah banyak membantu penulis,

Ibu Ir. Fariani Hermin M.T, selaku Pembimbing Akademik atas segala bimbingan dan kerja sama Ibu selama perkuliahan, dan seluruh Bapak/Ibu dosen atas pengajarannya yang telah diberikan, serta karyawan/karyawati FMIPA UNJ yang telah memberikan informasi yang penulis butuhkan dalam menyelesaikan skripsi.

4. Kedua kakak penulis, Mba Gita, dan Mba Asa, terima kasih atas segala perhatian, kasih sayang, dan motivasi serta doanya sehingga penulis dapat menyelesaikan penelitian ini.

5. Timah, Icha, Mella, Mei, Ela, Irma, Ziezie, Lusi, dan Leny, teman tercinta tersayang yang tak henti memberikan dukungan semangat, dan cinta. Terima kasih banyak telah menjadi bagian dari motivator yang luar biasa kepada penulis. Sukses terus ya kalian!

6. Teman-teman seperjuangan, teman-teman satu bimbingan yang sering mengingatkan skripsi, yang telah banyak memberikan semangat, saran dan nasihat kepada penulis dalam proses pengerjaan skripsi. Serta Teman-teman Matematika Murni 2012 yang telah memberikan banyak pengalaman, canda, tawa serta ilmu selama masa perkuliahan. Sungguh penulis sangat senang sekali bisa menjadi salah satu bagian dari kalian yang luar biasa.

7. Muhammad, Afif Makarim, S.Pd., Darisman, S.Sos., dan Andre Barbarosa, S.Psi., yang selalu perhatian dan pengertian, yang paling cerewet bertanya kabar skripsi, terima kasih telah banyak memberikan dukungan semangat, saran, motivasi dan tentunya yang selalu ngajak jalan dan main. Terima kasih, semoga kebaikan kalian dibalas oleh Allah Azza wa jalla.

8. Semua pihak yang tidak bisa penulis tulis satu persatu. Terima kasih atas dukungan dan motivasinya. Jazakumullah khoir.

Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan karena sesungguhnya kesempurnaan hanya milik Allah Azza wa jalla. Oleh karena itu, kritik serta saran yang membangun sangat penulis harapkan untuk kesempurnaan skripsi ini. Akhirnya penulis berharap semoga skripsi ini dapat bermanfaat, terutama bagi penulis sendiri serta bagi yang membacanya.

Jakarta, Agustus 2017

Sharah Annisa

DAFTAR ISI

ABSTRACT i

ABSTRAK ii

KATA PENGANTAR iv

DAFTAR ISI viii

DAFTAR TABEL ix

DAFTAR GAMBAR x

I PENDAHULUAN 1

1.1 Latar Belakang Masalah 1

1.2 Perumusan Masalah 3

1.3 Pembatasan Masalah 4

1.4 Tujuan Penulisan 4

1.5 Manfaat Penulisan 4

1.6 Metode Penelitian 5

II LANDASAN TEORI 6

2.1 Jangka Waktu Mencari kerja 6

2.2 Regresi Logistik 7

2.2.1 Analisis Regresi 7

2.2.2 Regresi Logistik 10

2.3 Regresi Logistik Ordinal 13

2.3.1 Estimasi Parameter 16

2.3.2 Pengujian Parameter 16

2.3.3 Uji Kesesuaian Model 18
2.4 Interpretasi Model 19
2.5 Ketepatan Klasifikasi 20
2.5.1 Tahapan Penulisan 21

III PEMBAHASAN **23**

3.1 Pengantar *Bootstrap Aggregating* (*Bagging*) 23
3.2 Metode *Bagging* 24
3.3 Ketepatan Klasifikasi *Bagging* 25
3.4 Contoh Kasus 28
3.4.1 Sumber Data 28
3.4.2 Hasil dan Pembahasan 29

IV PENUTUP **38**

4.1 Kesimpulan 38
4.2 Saran 39

DAFTAR PUSTAKA **40**

LAMPIRAN-LAMPIRAN **42**

DAFTAR TABEL

2.1 Tabel Ketepatan Klasifikasi Regresi Logitik 20

3.1 Tabel Regresi Logistik 31

3.2 Hasil *Odds Ratio* 33

3.3 Ketepatan Klasifikasi Regresi Logistik Ordinal 35

3.4 Hasil Klasifikasi *Bagging* Regresi Logistik Ordinal 36

3.5 Hasil *Odds Ratio* 36

4.1 Ketepatan Klasifikasi Regresi Logistik Ordinal 38

4.2 Data SAKERNAS 43

4.3 Data Training 46

4.4 Data Testing 48

DAFTAR GAMBAR

2.1 Diagram Alur 22

3.1 Diagram Alur *Bagging* 27

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Bekerja merupakan aktivitas yang harus dilakukan masyarakat agar dapat menjalankan roda kehidupan untuk memenuhi kebutuhan pokok sehari-hari dari penghasilan kerja tersebut, karena pada dasarnya kebutuhan pokok manusia adalah kebutuhan yang harus segera dipenuhi dan tidak bisa ditunda, seperti kebutuhan untuk makan dan tempat tinggal. Bekerja menurut Badan Pusat Statistik (BPS) adalah kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus) dalam seminggu yang lalu. Hasil pendapatan itu yang dijadikan untuk memenuhi kebutuhan masyarakat.

Saat ini proses menuju dunia kerja tidak mudah, banyak persaingan dan persyaratan yang diminta oleh suatu perusahaan terhadap calon tenaga kerja. Tiap tenaga kerja atau calon pekerja tentu ingin mendapat pekerjaan dengan cepat setelah menyelesaikan sekolah atau masa studi perguruan tinggi karena tidak ingin terlalu lama menganggur, namun keberuntungan dalam mendapat pekerjaan akan berbeda beda pada tiap calon pekerja. Tingkat jangka waktu dalam mencari pekerjaan sangat bervariasi, ada yang hanya dalam waktu singkat yaitu beberapa bulan setelah lulus Sekolah atau Perguruan Tinggi langsung mendapat perkerjaan, sebaliknya ada yang sampai beberapa tahun baru mendapat perkerjaan sehingga membuat calon pekerja stress karena menganggur terlalu lama.

Melihat semakin banyak calon pekerja dan persaingan dalam mendapat pekerjaan juga semakin ketat, maka perlu dilakukan identifikasi karakteristik calon pekerja dengan cara memperhatikan faktor apa saja yang dapat mempengaruhi jangka waktu mendapat kerja, sehingga dapat menjadi acuan bagi calon pekerja dalam menentukan strategi untuk mendapat pekerjaan yang sesuai dengan cepat. Faktor-faktor tersebut merupakan data kontinu dan kategori yang nantinya akan digunakan sebagai peubah penjelas dalam penelitian ini. Adapun peubah respon berupa kategori bertingkat. Kategori bertingkat adalah data kategori yang memiliki urutan peringkat, mengandung tingkatan, atau tidak setara, biasa disebut dengan data ordinal, contoh Indeks Prestasi Kumulatif mahasiswa dan stadium kanker.

Pada statistika, metode yang dapat digunakan untuk memetakan tingkat jangka waktu masyarakat dalam mendapat kerja yang sesuai dengan faktor-faktor yang mempengaruhinya adalah dengan menggunakan metode klasifikasi. Pengklasifikasian merupakan salah satu metode statistik dalam pengelompokan suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari, seperti pengklasifikasian data pada bidang akademik dalam kasus klasifikasi program studi mahasiswa baru, pada bidang kesehatan dalam kasus klasifikasi status gizi buruk.

Terdapat beberapa metode klasifikasi, seperti analisis klaster, analisis faktor, analisis diskriminan, dan analisis regresi. Dalam penelitian ini, data yang digunakan adalah peubah penjelas yang bersifat kontinu dan kategorik. Kemudian data peubah respon berskala kategorik bertingkat. Berdasarkan data tersebut, metode yang cocok untuk mengatasi masalah pengklasifikasian ini adalah metode analisis regresi logistik. Selain dikenal untuk mendeskripsikan hubungan antara suatu peubah respon dan satu atau lebih peubah penjelas, analisis regresi logistik juga dapat digunakan dalam pengklasifikasian. Namun

metode pengklasifikasian dengan menggunakan regresi logistik akan memberikan pendugaan parameter yang tidak stabil, artinya jika terdapat perubahan pada data *set* maka akan menyebabkan perubahan yang signifikan pada model (Breiman, 1994). Maka untuk memperoleh parameter yang stabil pada model regresi logistik digunakan metode yang memberikan tingkat akurasi prediksi lebih baik yaitu *Bootstrap Aggregating* (*bagging*). Metode *bagging* ini diusulkan oleh Breiman (1994) dengan tujuan yaitu untuk memperbaiki stabilitas dan meningkatkan ketepatan dalam klasifikasi.

Penelitian sebelumnya yang berkaitan dengan *bagging* regresi logistik yaitu Suryaningrum (2011) mengklasifikasi kesejahteraan rumah tangga di kota Malang dengan pendekatan *bagging* regresi logistik biner, sedangkan Reza (2015) menggunakan pendekatan *bagging* regresi logistik multinomial untuk menentukan ketepatan klasifikasi pemilihan kontrasepsi di kota Semarang. Oleh karena itu, pada skripsi ini akan dibahas tentang metode pendekatan *bagging* regresi logistik ordinal sebagai peningkatan ketepatan klasifikasi regresi logistik ordinal pada kasus klasifikasi jangka waktu mendapat kerja di DKI Jakarta.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka beberapa masalah yang akan dibahas adalah sebagai berikut:

1. Bagaimana nilai klasifikasi regresi logistik ordinal pada jangka waktu mendapat kerja?

2. Bagaimana keakuratan prediksi ketepatan klasifikasi dengan *bagging* regresi logistik ordinal?

1.3 Pembatasan Masalah

Batasan masalah dalam penelitian ini adalah data yang digunakan, data Survey Angkatan Kerja Nasional Di DKI Jakarta bulan Agustus 2014 dengan tiga kategori peubah respon yaitu waktu kurang dari 12 bulan, antara 12 bulan sampai 24 bulan, dan lebih dari 24 bulan.

1.4 Tujuan Penulisan

Adapun tujuan yang ingin dicapai dalam penulisan adalah sebagai berikut:

1. Memperoleh nilai klasifikasi regresi logistik ordinal pada jangka waktu mendapatkan kerja.
2. Mengkaji keakuratan prediksi ketepatan klasifikasi dengan *bagging* regresi logistik ordinal.

1.5 Manfaat Penulisan

Manfaat yang diharapkan dari penulisan ini adalah:

1. Dapat memperkaya wawasan tentang metode analisis regresi logistik ordinal dan *bagging* regresi logistik ordinal serta aplikasinya.
2. Diharapkan hasil skripsi ini dapat dijadikan bahan referensi serta pembanding dalam penelitian selanjutnya yang relevan dan lebih mendalam.
3. Diharapkan pula dapat memberikan informasi yang berguna bagi semua pihak yang terkait dengan studi kasus yang diteliti.

1.6 Metode Penelitian

Metode penulisan skripsi ini merupakan kajian teoritis berdasarkan buku-buku dan jurnal yang berkaitan tentang permasalahan yang ada dalam penulisan seperti Agresti (2007) dan Breiman (1994). Pembahasan dan pengaplikasian terhadap data yang diberikan merupakan hasil pemikiran yang logis dan sistematis.

BAB II

LANDASAN TEORI

Pada bab ini, akan dibahas tentang analisis regresi logistik, analisis regresi logistik ordinal beserta estimasi parameter, pengujian parameter, uji kesesuaian model, interpretasi model, dan ketepatan klasifikasi. Sebagai awalan, akan dijelaskan beberapa faktor yang diduga mempengaruhi tingkat jangka waktu mendapat pekerjaan.

2.1 Jangka Waktu Mencari kerja

Bekerja adalah kegiatan yang harus dilakukan masyarakat untuk memenuhi kebutuhan sehari-hari. Proses menuju dunia kerja saat ini tidak mudah, banyak persaingan dan persyaratan yang diminta oleh perusahaan terhadap calon tenaga kerja. Mendapatkan pekerjaan menjadi mudah apabila calon tenaga kerja memiliki kompetensi atau kualifikasi yang sesuai dengan kebutuhan perusahaan. Perusahaan hanya akan memilih tenaga kerja yang benar-benar memiliki kualifikasi yang diinginkan, terdapat faktor internal dan eksternal yang mempengaruhi keberhasilan tenaga kerja dalam mendapatkan pekerjaan misal diantaranya faktor internal yaitu usia dan jenis kelamin.

Faktor internal lainnya yang mempengaruhi seperti sikap *(attitude)*, mentalitas, keahlian *(skill)* yang merupakan kelemahan-kelemahan yang berasal dari dalam diri calon tenaga kerja yang seringkali tidak disadari oleh calon tenaga kerja tersebut. Disamping faktor internal, terdapat faktor eksternal yang mungkin mempengaruhi tingkat jangka waktu calon kerja dalam mendapat

pekerjaan, yaitu tingkat pendidikan dan penghasilan.

Umur produktif biasanya akan mendapatkan peluang kerja lebih cepat dibanding umur yang telah melewati usia produktif. Tingkat pendidikan juga umumnya berpengaruh positif terhadap lama mencari kerja karena semakin tinggi pendidikan seseorang maka waktu memperoleh pekerjaan semakin cepat. Selain itu, pendapatan juga akan sangat menentukan bagi pencari kerja terdidik untuk memperoleh pekerjaan karena pendapatan yang sesuai utnuk memenuhi kebutuhan hidup bisa menjadi salah satu parameter lamanya seorang untuk memperoleh pekerjaan. Selain itu, faktor lain yang mungkin mempengaruhi tingkat jangka waktu calon kerja dalam mendapat pekerjaan yaitu status keluarga, status pernikahan, dan jenis perkerjaan yang diinginkan. Penelitian kali ini, jangka waktu mendapat kerja digunakan sebagai peubah respon, yang terbagi menjadi tiga waktu, yaitu kurang dari 12 bulan, antara 12 bulan sampai 24 bulan, dan lebih dari 24 bulan.

2.2 Regresi Logistik

2.2.1 Analisis Regresi

Regresi linier berganda adalah suatu prosedur untuk mendapatkan hubungan matematis dalam bentuk persamaan antara dua atau lebih peubah penjelas yang dihubungkan dengan satu peubah respon. Jika hanya ada satu peubah penjelas maka disebut regresi linier sederhana. Untuk memberikan gambaran tentang permasalahan tersebut biasanya digunakan model regresi yaitu model regresi linier berganda. Kemudian untuk menaksir koefisien parameter, salah satu cara mendapatkan penaksir yang baik adalah dengan menggunakan metode maksimum likelihood. Metode ini berguna untuk menentukan parameter yang memaksimalkan kemungkinan dari data sampel.

Analisis regresi linier memiliki empat asumsi klasik yang harus terpenuhi. Pengujian asumsi klasik untuk menguji kelayakan model regresi dan kelayakan peubah penjelas. Tujuannya ialah agar dapat menghasilkan nilai parameter yang baik sehingga hasil penelitian dapat lebih diandalkan. Terdapat empat uji asumsi klasik yang harus dilakukan terhadap suatu model regresi tersebut, yaitu uji normalitas, uji autokorelasi, uji multikolinieritas, dan uji heterokedastisitas. Uji normalitas yaitu untuk menguji apakah peubah berdistribusi normal atau tidak. Uji autokorelasi untuk mengetahui ada atau tidaknya penyimpangan korelasi yang terjadi antara residual pada satu pengamatan dengan pengamatan lain pada model regresi. Uji multikolinieritas adalah untuk melihat ada atau tidaknya hubungan linier atau korelasi yang tinggi antar peubah penjelas dalam model regresi. Terakhir uji heterokedastisitas bertujuan untuk menguji apakah dalam regresi varian dari residual tidak sama untuk satu pengamatan ke pengamatan yang lain. Kemudian syarat yang harus dipenuhi dari keempat asumsi tersebut yaitu berdistribusi normal, tidak ada autokorelasi, tidak ada multikolinieritas, dan tidak ada heterokedastisitas.

Setelah keempat asumsi terpenuhi maka dilakukan uji serentak dan uji parsial untuk mengetahui apakah peubah penjelas berpengaruh baik secara bersama-sama ataupun secara parsial terhadap peubah respon.

Uji Serentak (Uji F)

Uji F digunakan untuk mengetahui apakah seluruh peubah penjelas secara bersama-sama berpengaruh terhadap peubah respon. Nilai F yang tinggi menunjukkan pengaruh seluruh peubah penjelas signifikan.

Langkah langkah pengujiannya sebagai berikut :

- Menentukan hipotesis :

  $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ artinya secara bersama-sama tidak terda-

pat pengaruh yang signifikan antara peubah penjelas terhadap peubah respon.

H_1 : paling tidak ada satu $\beta_i \neq 0$ dengan $i = 1, 2, \ldots, k$ artinya secara bersama-sama terdapat pengaruh yang signifikan antara peubah penjelas terhadap peubah respon.

- Menentukan derajat kepercayaan α.
- Menentukan F hitung dan F tabel :

  Jika $F_{hit} > F_{tabel}$, maka tolak H_0

  Jika $F_{hit} \leq F_{tabel}$, maka terima H_0
- Kesimpulan

Uji Parsial (Uji t-student)

Uji t-student digunakan untuk menguji signifikansi dari peubah penjelas satu demi satu. Jadi untuk setiap koefisien regresi nilai t-student selalu dihitung. Uji t-student dilakukan dengan melihat nilai signifikansi masing-masing peubah penjelas terhadap peubah respon dengan asumsi bahwa peubah lain dianggap konstan. Jika semua peubah penjelas signifikan pada pengujian t-student, biasanya pada uji F akan diperoleh hasil yang signifikan.

Langkah langkah pengujiannya sebagai berikut :

- Menentukan hipotesis :

  H_0 : $\beta_i = 0$ artinya peubah penjelas ke-i tidak berpengaruh signifikan terhadap peubah respon secara parsial.

  H_1 : $\beta_i \neq 0$ dengan $i = 1, 2, \ldots, k$ artinya peubah penjelas ke-i berpengaruh signifikan terhadap peubah respon secara parsial.
- Menentukan derajat kepercayaan α

- Menentukan t hitung dan t tabel :

  Jika signifikansi $t_{hit} > t_{tabel}$ maka tolak H_0

  Jika signifikansi $t_{hit} \leq t_{tabel}$ maka terima H_0

- Kesimpulan

2.2.2 Regresi Logistik

Metode regresi merupakan analisis data yang mendeskripsikan hubungan antara suatu peubah respon dan satu atau lebih peubah penjelas atau prediktor (Hosmer & Lemeshow, 2000). Analisis regresi linier digunakan jika pengaruh peubah penjelas terhadap peubah respon bersifat linier dan peubah respon bertipe kontinu sedangkan peubah penjelas dapat bertipe kontinu atau kategorik. Adapun jika peubah respon bertipe kategorik maka penggunaan regresi linier ini untuk menjelaskan pengaruh peubah penjelas terhadap peubah respon menjadi tidak valid lagi. Analisis regresi logistik adalah salah satu metode yang dapat digunakan untuk mencari hubungan peubah respon yang bersifat dikotomus (berskala nominal atau ordinal dengan dua kategori) atau bersifat polikotomus (mempunyai skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih peubah penjelas. Peubah penjelas yang digunakan sebagai pembentuk model terhadap peubah respon dapat bersifat kontinu atau kategorik (Agresti, 2007).

Regresi logistik tidak memerlukan asumsi normalitas, autokorelasi, dan heterokedastisitas, dikarenakan peubah respon bersifat kategorik, sehingga residualnya, tidak memerlukan ketiga pengujian tersebut. Selain itu, beberapa asumsi regresi logistik yaitu:

1. Regresi logistik tidak membutuhkan hubungan linier antara peubah respon dengan peubah penjelas.

2. Peubah respon harus bersifat kategorik (dikotomi atau polikotomi).

3. Asumsi homokedastisitas tidak diperlukan.

4. Kategori dalam peubah penjelas harus terpisah satu sama lain atau bersifat eksklusif.

5. Regresi logistik dapat menyeleksi hubungan karena menggunakan pendekatan non linier log transformasi untuk memprediksi odds ratio.

Adapun fungsi logistik adalah sebagai berikut:

$$f(x) = \frac{1}{1+e^{-x}}, -\infty < x < \infty \tag{2.1}$$

Dengan melihat hasil kemungkinan nilai $f(x)$ yang berkisar antara 0 dan 1, ini menunjukkan bahwa regresi logistik menggambarkan probabilitas terjadinya suatu kejadian.

Menurut Hosmer dan Lemeshow (2000), dalam persamaan regresi logistik, notasi $\pi(x) = E(Y|x)$ digunakan untuk mempresentasikan mean dari Y jika diberikan nilai x yang juga merupakan bentuk taksiran fungsi peluang dan bentuk khusus dari model regresi logistik yang dinyatakan dalam persamaan sebagai berikut.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{pi})} \tag{2.2}$$

dimana:

$\pi(x)$ = Nilai peluang peubah respon.

x_{pi} = Nilai amatan dari peubah penjelas ke-p.

p = Banyaknya peubah penjelas yang diamati.

β_p = Parameter (koefisien regresi) untuk peubah penjelas ke-p.

Jika model dari persamaan (2.2) ditransformasi dengan transformasi logit dari $\pi(x)$. Maka hasilnya merupakan bentuk penyederhanaan dari model regresi logistik dan dapat ditulis sebagai berikut.

$$\begin{aligned} g(\mathrm{x}) &= \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \\ &= \beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{pi} \end{aligned} \tag{2.3}$$

Dengan $g(\mathrm{x}) = \text{logit}\ \pi(x)$.

Estimasi Parameter

Pada model regresi logistik metode estimasi yang digunakan untuk menaksir nilai $\boldsymbol{\beta}$ adalah Metode *Maksimum Likelihood Estimation* (MLE). Prinsip metode ini adalah mengestimasi nilai parameter $\boldsymbol{\beta}$ yang terdapat dalam persamaan regresi logistik dengan memaksimumkan fungsi likelihood. Bentuk umum dari fungsi likelihood untuk sampel dengan n observasi independen, (y_i, x_i) dimana $i = 1, 2, \ldots, n$ adalah sebagai berikut (Hosmer & Lemenshow, 2000).

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \tag{2.4}$$

Selanjutnya didapatkan fungsi log likelihood sebagai berikut.

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \ln[\pi(x_i)] + (1-y_i)\ln(1-\pi(x_i)) \tag{2.5}$$

Hasil parameter yang dilakukan melalui metode MLE adalah dengan mencari turunan fungsi log likelihood terhadap parameter yang akan diestimasi selanjutnya disamadengankan nol. Hasil turunan pertama dari fungsi log likelihood secara umum sebagai berikut (Hosmer & Lemenshow, 2000).

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{kj}} = \sum_{i=1}^{n} x_{ki}(y_{ji} - \pi_{ji}) = 0$$

Langkah selanjutnya adalah mencari turunan kedua dari fungsi *ln-likelihood* terhadap parameter yang akan diestimasi.

$$\frac{\partial^2 l\left(\boldsymbol{\beta}\right)}{\partial\beta_{kj}\beta_{k'j}} = -\sum_{i=1}^{n} x_{k'i}x_{ki}\pi_{ji}\left(1-\pi_{ji}\right) = 0$$

Nilai parameter estimasi yang dihasilkan diduga bersifat tidak linier, jadi penyeleseaiannya tidak dapat dilakukan secara langsung maka dapat menggunakan metode numerik untuk memperoleh estimasi parameternya yaitu dengan metode iterasi *Newton-Raphson*. Metode ini dilakukan dengan iterasi yang menggunakan rumus sebagai berikut (Agresti, 2007).

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \left(\mathbf{H}^{(t)}\right)^{-1}\mathbf{q}^{(t)}$$

dimana:

$\mathbf{H}^t$ adalah matriks turunan kedua fungsi log likelihood dan $\mathbf{q}^t$ adalah matriks turunan pertama fungsi log likelihood.

$$\mathbf{q}^{(t)} = \begin{pmatrix} \frac{\partial l(\boldsymbol{\beta})}{\partial\beta_{01}} & \frac{\partial l(\boldsymbol{\beta})}{\partial\beta_{02}} & \cdots & \frac{\partial l(\boldsymbol{\beta})}{\partial\beta} \end{pmatrix}^T$$

$$\left(\mathbf{H}^{(t)}\right) = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\beta})}{\partial{\beta_{01}}^2} & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_{01}\partial\beta_{02}} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_{01}\partial\beta} \\ \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_{01}\partial\beta_{02}} & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial{\beta_{02}}^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_{02}\partial\beta} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_{01}\partial\beta} & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_{02}\partial\beta} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta^2} \end{pmatrix}^T$$

dengan banyaknya iterasi $t = 0, 1, 2, \ldots$. Iterasi *Newton-Raphson* akan berhenti apabila $\left\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\right\| \leq \varepsilon$

2.3 Regresi Logistik Ordinal

Regresi logistik ordinal merupakan metode analisis yang digunakan untuk mengetahui hubungan antara peubah respon dengan peubah penjelas dimana

peubah responnya memiliki lebih dari dua kategori dan dalam setiap kategori memiliki tingkatan atau berskala ordinal (Hosmer & Lemenshow, 2000).

Model yang digunakan dalam regresi logistik ordinal adalah model logit atau yang disebut dengan istilah *cummulative logit models*. Menurut Agresti (2007), peluang kumulatif $P\left(Y \leq j|\mathrm{x}_i\right)$ dapat didefinisikan sebagai berikut:

$$P\left(Y \leq j|\mathrm{x}_i\right) = \frac{\exp\left(\beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}{1 + \exp\left(\beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)} \tag{2.6}$$

dimana:

$\mathrm{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ merupakan nilai pengamatan ke-i dari setiap variabel p peubah penjelas dimana $i = (1, 2, \ldots, n)$.

β_k = Faktor koefisien yang tidak diketahui yang bersesuian dengan x_i .

$k = (1, 2, \ldots, p)$ banyaknya peubah penjelas

$j = 1, 2, \ldots, J-1$ dimana J banyaknya peubah respon

Adapun peluang kumulatif yang dapat dibentuk jika terdapat J kategori respon adalah sebanyak $J-1$ peluang kumulatif. Misal terdapat tiga kategori respon, maka peluang kumulatif yang didapat bisa ditulis sebagai berikut:

$$P\left(Y \leq 1|\mathrm{x}_i\right) = \frac{\exp\left(\beta_{01} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}{1 + \exp\left(\beta_{01} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}$$

$$P\left(Y \leq 2|\mathrm{x}_i\right) = \frac{\exp\left(\beta_{02} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}{1 + \exp\left(\beta_{02} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}$$

Model kumulatif logit merupakan model didapat dengan membandingkan peluang kumulatif yaitu perbandingan peluang yang kurang dari atau sama dengan kategori respon ke-j pada peubah penjelas p yang dinyatakan dalam

vektor x_i yaitu $P\left(Y \leq j|\mathrm{x}_i\right)$, dengan probabilitas yang lebih besar dari kategori respon ke-j, $P\left(Y > j|\mathrm{x}_i\right)$ (Hosmer & Lemenshow, 2000). Sesuai dengan definisi *cummulative logit models*, maka didapatkan model logit regresi ordinal atau *cummulative logit models* adalah sebagai berikut:

$$\begin{aligned}
\text{Logit } P\left(Y \leq j|\mathrm{x}_i\right) &= \ln\left[\frac{P\left(Y \leq j|\mathrm{x}_i\right)}{P\left(Y > j|\mathrm{x}_i\right)}\right] \\
&= \ln\left[\frac{P\left(Y \leq j|\mathrm{x}_i\right)}{1 - P\left(Y \leq j|\mathrm{x}_i\right)}\right] \\
&= \ln\left[\frac{\frac{\exp\left(\beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}{1+\exp\left(\beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}}{1 - \frac{\exp\left(\beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}{1+\exp\left(\beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik}\right)}}\right] \\
&= \beta_{0j} + \sum_{k=1}^{p} \beta_k \mathrm{x}_{ik} \qquad (2.7)
\end{aligned}$$

Jika $\pi_j\left(\mathrm{x}_i\right) = P\left(Y = j|\mathrm{x}_i\right)$ menyatakan peluang kategori respon ke- j pada peubah penjelas p yang dinyatakan dalam vektor x_i dan $P\left(Y \leq j|\mathrm{x}_i\right)$ menyatakan peluang kumulatif pada peubah penjelas p yang dinyatakan dalam vektor x_i maka nilai $P\left(Y \leq j|\mathrm{x}_i\right)$ diperoleh dengan persaman berikut:

$$P\left(Y \leq j|\mathrm{x}_i\right) = \pi_1\left(\mathrm{x}_i\right) + \pi_2\left(\mathrm{x}_i\right) + \ldots + \pi_j\left(\mathrm{x}_i\right) \qquad (2.8)$$

Sehingga dapat diperoleh nilai peluang masing-masing kategori peubah respon sebagai berikut:

$$\begin{aligned}
P\left(Y = 1|\mathrm{x}_i\right) &= \pi_1\left(\mathrm{x}_i\right) = P\left(Y \leq 1|\mathrm{x}_i\right) \\
P\left(Y = 2|\mathrm{x}_i\right) &= \pi_2\left(\mathrm{x}_i\right) = P\left(Y \leq 2|\mathrm{x}_i\right) - \pi_1\left(\mathrm{x}_i\right) \\
&\vdots \\
P\left(Y = J|\mathrm{x}_i\right) &= \pi_J\left(\mathrm{x}_i\right) = 1 - P\left(Y \leq J-1|\mathrm{x}_i\right) \qquad (2.9)
\end{aligned}$$

Nilai $\pi_j\left(\mathrm{x}_i\right)$ pada persamaan diatas dijadikan pedoman dalam pengklasifikasian. Suatu pengamatan akan diklasifikasikan pada respon kategori j berdasarkan nilai $\pi_j\left(\mathrm{x}_i\right)$ yang terbesar.

2.3.1 Estimasi Parameter

Pada penelitian ini metode estimasi yang digunakan untuk menaksir nilai $\boldsymbol{\beta}$ adalah Metode *Maksimum Likelihood Estimation* (MLE). Prinsip metode ini adalah mengestimasi nilai parameter $\boldsymbol{\beta}$ yang terdapat dalam persamaan regresi logistik dengan memaksimumkan fungsi likelihood. Bentuk umum dari likelihood untuk sampel dengan n observasi adalah sebagai berikut (Hosmer & Lemenshow, 2000).

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_1(\mathrm{x}_i)^{y_{1i}} \pi_2(\mathrm{x}_i)^{y_{2i}} \ldots \pi_J(\mathrm{x}_i)^{y_{Ji}} \tag{2.10}$$

Selanjutnya didapatkan fungsi log likelihood sebagai berikut.

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_{1i} \ln\left[\pi_1(\mathrm{x}_i)\right] + y_{2i} \ln\left[\pi_2(\mathrm{x}_i)\right] + \ldots + y_{Ji} \ln\left[\pi_J(\mathrm{x}_i)\right] \tag{2.11}$$

Langkah selanjutnya adalah mencari turunan pertama dan turunan kedua fungsi log likelihood terhadap parameter yang akan diestimasi dan disama dengankan nol. Nilai parameter yang dihasilkan bersifat tidak linier sehingga diperlukan metode numerik untuk memperoleh estimasi parameternya yaitu dengan metode iterasi *Newton-Raphson*. Hal ini serupa pada regresi logistik biner yang telah dijelaskan.

2.3.2 Pengujian Parameter

Setelah mendapatkan parameter maka selanjutnya adalah menguji signifikansi dari parameter yang telah diestimasi tersebut secara serentak dan parsial. Maksud dari pengujian parameter ini adalah untuk mengetahui apakah peubah penjelas yang terdapat dalam model berpengaruh atau tidak terhadap peubah responnya. Pengujian dilakukan terhadap koefisien β dari model yang telah diperoleh. Dalam model regresi logistik terdapat dua jenis pengujian yaitu pengujian secara serentak dan parsial.

1. Uji Serentak

   Uji serentak dilakukan untuk mengetahui keberartian koefisien β terhadap peubah respon secara keseluruhan atau serentak. Jika parameter yang diuji signifikan maka dapat dikatakan model yang dibentuk sesuai untuk memodelkan peubah respon. Hipotesis yang digunakan adalah sebagai berikut.

   $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$

   H_1 : paling sedikit terdapat satu $\beta_k \neq 0$ dengan $k = 1, 2, \ldots, p$

   Statistik uji yang digunakan adalah uji G atau *likelihood ratio test.*

   $$G = -2 \ln \left[\frac{L_0}{L_1} \right] \tag{2.12}$$

   dimana:

   L_0 = likelihood untuk model yang tidak mengandung peubah penjelas yang akan diuji signifikansinya.

   L_1 = likelihood untuk model yang mengandung peubah penjelas yang akan diuji signifikansinya.

   Daerah penolakan H_0 adalah jika $G > \chi^2_{(v,\alpha)}$ atau $p - value < \alpha$ dengan v adalah derajat bebas banyaknya peubah penjelas.

2. Uji Parsial

   Uji parsial digunakan untuk menguji pengaruh setiap parameter β secara individual terhadap respon. Statistik uji yang digunakan adalah uji *Wald* (Hosmer & Lemenshow, 2000). Berdasarkan hasil dari uji *Wald* maka akan diketahui apakah suatu peubah penjelas layak atau tidak masuk kedalam model. Hipotesis yang digunakan adalah sebagai berikut.

   $H_0 : \beta_k = 0$

   $H_1 : \beta_k \neq 0$ dengan $k = 1, 2, \ldots, p$ p adalah jumlah penjelas dalam model.

Statistik uji *Wald* adalah sebagai berikut.

$$W = \frac{\widehat{\beta}_k}{SE\left(\widehat{\beta}_k\right)} \tag{2.13}$$

dengan $SE\left(\widehat{\beta}_k\right)$ adalah taksiran standar *error* parameter. Daerah penolakan H_0 adalah jika $p - value < \alpha$ atau $|W| > Z_{\alpha/2}$

2.3.3 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui apakah suatu model yang terbentuk sudah sesuai untuk digunakan atau tidak. Statistik uji yang digunakan adalah uji (Hosmer & Lemenshow, 2000) dengan hipotesis sebagai berikut .

H_0: model sesuai (tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

H_1: model tidak sesuai (ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

Statistik uji *Chi-square* adalah sebagai berikut.

$$\widehat{C} = \sum_{k=1}^{g} \frac{\left(O_k - n'_k \overline{\pi_k}\right)^2}{n'_k \overline{\pi_k}\left(1 - \overline{\pi_k}\right)} \tag{2.14}$$

dimana:

$O_k = \sum_{j=1}^{C_k} y_j$ adalah jumlah nilai peubah respon pada kelompok ke-k.

$\overline{\pi_k} = \sum_{j=1}^{C_k} \frac{m_j \widehat{\pi_j}}{n'_k}$ adalah rata-rata taksiran probabilitas dimana m_j banyaknya subjek pada c_k kombinasi peubah penjelas.

g = jumlah kelompok.

n'_k = jumlah subjek dalam kelompok ke-k.

Daerah penolakan H_0 adalah jika $p - value > \alpha$ atau $\widehat{C} < \chi^2_{(g-2)}$.

2.4 Interpretasi Model

Interpretasi dalam regresi logistik menggunakan nilai *odds ratio* yang menunjukkan perbandingan tingkat kecenderungan dari kategori yang ada dalam satu peubah penjelas. Interpretasi model meliputi menentukan hubungan fungsional antara peubah respon dan peubah penjelas serta mendefinisikan unit perubahan peubah respon yang disebabkan oleh peubah penjelas. *Odds ratio* untuk $Y \leq j$ terhadap $Y > j$ yang dihitung pada dua nilai (misal x=a dan x=b) adalah sebagai berikut.

$$\begin{aligned}
\psi(a,b) &= \frac{\left[\frac{P(Y\leq j|x=a)}{P(Y>j|x=a)}\right]}{\left[\frac{P(Y\leq j|x=b)}{P(Y>j|x=b)}\right]} \\
&= \frac{\exp(\beta_{0j}+\beta_i(a))}{\exp(\beta_{0j}+\beta_i(b))} \\
&= \exp\{(\beta_{0j}+\beta_i(a))-(\beta_{0j}+\beta_i(b))\} \\
&= \exp[\beta_i(a-b)] \qquad (2.15)
\end{aligned}$$

Sehingga jika $a-b=1$ maka $\psi = \exp(\beta_i)$

Persaman (2.15) dapat diinterpretasikan bahwa peluang suatu respon memiliki kategori lebih kecil atau sama dengan kategori ke-j dibandingkan dengan respon yang memiliki kategori lebih besar dari kategori ke-j pada $x=a$ adalah sebesar $\exp(\beta_i)$ kali dibandingkan pada $x=b$.

Nilai *odds ratio* digunakan untuk menunjukkan kecenderungan hubungan suatu peubah penjelas terhadap peubah respon. Jika nilai $\psi = 1$ maka itu menunjukkan jika resiko dua kejadian adalah sama yang berarti $x=a$ memiliki resiko yang sama dengan $x=b$ untuk menghasilkan $Y \leq j$ dibandingkan $Y > j$. Jika $1 < \psi < \infty$ berarti $x=a$ memiliki ψ kali lebih tinggi dibanding $x=b$ begitu pula untuk $0 < \psi < 1$. Untuk peubah penjelas kontinu maka diperlukan unit perubahan sebesar c, sehingga *odds ratio* diperoleh dengan $\exp(c\beta_i)$ (Hosmer & Lemenshow, 2000).

2.5 Ketepatan Klasifikasi

Evaluasi prosedur ketepatan klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi (Johnson & Wichern, 1992). Pengukuran yang biasa dipakai untuk melihat kesalahan klasifikasi adalah *apparent error rate* (APER). Nilai APER menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi (Johnson & Wichern, 1992). Nilai APER ini akan lebih mudah dihitung jika disajikan dalam bentuk tabel. Berikut ini tabel APER:

Tabel 2.1: Tabel Ketepatan Klasifikasi Regresi Logitik

Kelompok Pengamatan	Kelompok Prediksi		
	y_1	y_2	y_3
y_1	n_{11}	n_{12}	n_{13}
y_2	n_{21}	n_{22}	n_{23}
y_3	n_{31}	n_{32}	n_{33}

dimana:

n_{11} = Jumlah subjek dari y_1 yang tepat diklasifikasikan sebagai y_1

n_{22} = Jumlah subjek dari y_2 yang tepat diklasifikasikan sebagai y_2

n_{33} = Jumlah subjek dari y_3 yang tepat diklasifikasikan sebagai y_3

n_{12} = Jumlah subjek dari y_1 yang salah diklasifikasikan sebagai y_2

n_{13} = Jumlah subjek dari y_1 yang salah diklasifikasikan sebagai y_3

n_{21} = Jumlah subjek dari y_2 yang salah diklasifikasikan sebagai y_1

n_{23} = Jumlah subjek dari y_2 yang salah diklasifikasikan sebagai y_3

n_{31} = Jumlah subjek dari y_3 yang salah diklasifikasikan sebagai y_1

n_{32} = Jumlah subjek dari y_3 yang salah diklasifikasikan sebagai y_2

Maka rumus APER adalah sebagai berikut:

$$\text{APER}= \left(\frac{n_{12} + n_{13} + n_{21} + n_{23} + n_{31} + n_{32}}{n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} + n_{31} + n_{32} + n_{33}} \right) \tag{2.16}$$

Nilai APER dalam %. Maka nilai ketepatan klasifikasi adalah

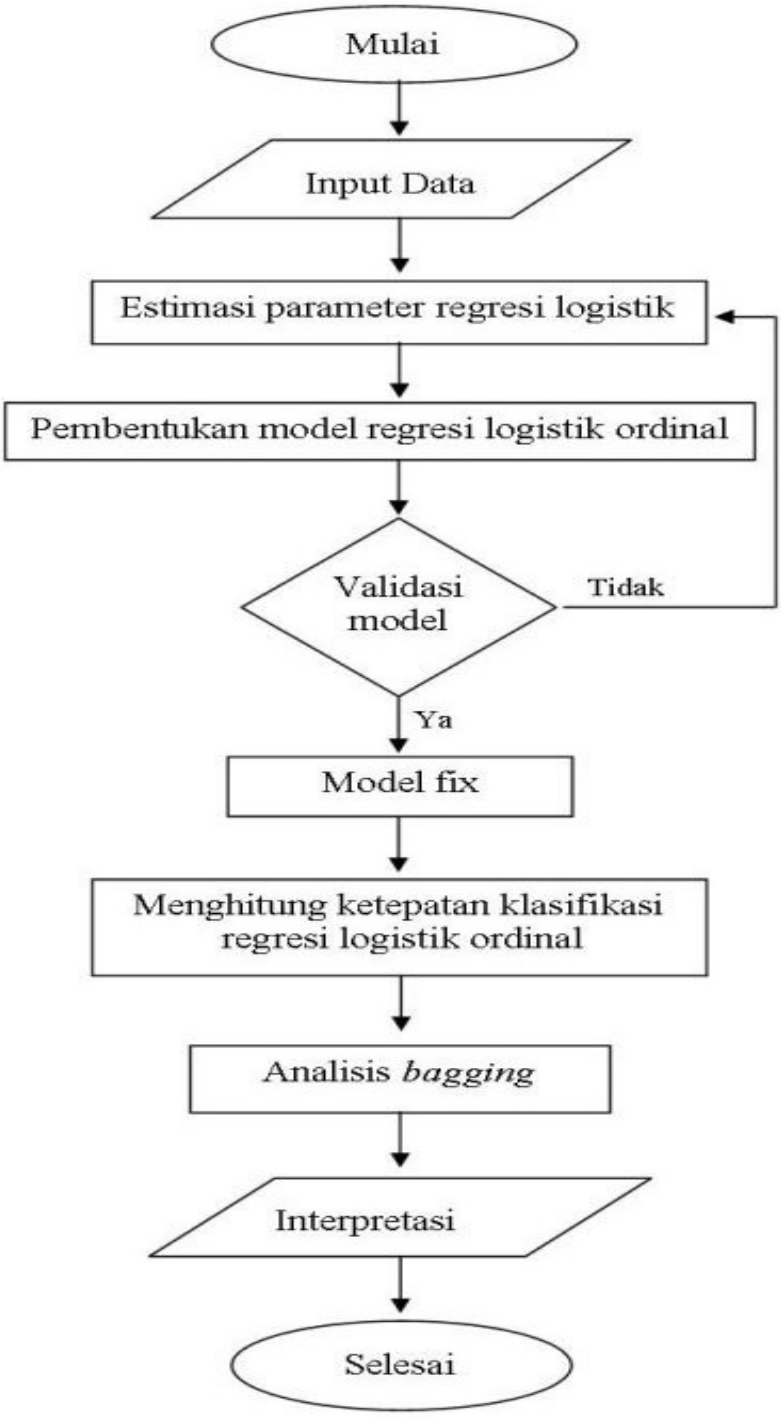
$$\text{Ketepatan Klasifikasi} = 1 - \text{APER}. \tag{2.17}$$

Tidak ada kriteria nilai ketepatan klasifikasi itu dapat dikatakan baik. Sehingga berapapun rentang nilai dari ketepatan klasifikasi, tetap dapat ditingkatkan lagi dengan menggunakan metode yang dapat meningkatkan ketepatan klasifikasi yaitu metode *bootstrap aggregating* (*bagging*). Seperti pada penelitian sebelumnya oleh Suryaningrum. Nilai ketepatan klasifikasi yang didapat oleh Suryaningrum tanpa *bagging* sudah cukup tinggi yaitu 97,8%, setelah dilakukan *bagging* maka terjadi peningkatan nilai ketepatan klasifikasi menjadi 98%.

2.5.1 Tahapan Penulisan

Proses dalam penulisan ini melalui beberapa tahapan untuk melakukan analisis data yaitu sebagai berikut:

1. Melakukan Penginputan data. Membagi data menjadi data *training* dan data *testing*.
2. Melakukan estimasi parameter regresi logistik ordinal dengan MLE.
3. Melakukan pengujian secara serentak dan secara parsial.
4. Melalukan uji kesesuaian model dengan uji *Chi-square*.
5. Mendapatkan peubah penjelas yang signifikan berpengaruh pada model regresi logistik ordinal serta interpretasi model dengan *odds ratio*
6. Menghitung ketepatan klasifikasi pada regresi logistik ordinal.
7. Melakukan analisa *bagging* regresi logistik ordinal.
8. Interpretasi dari masing-masing metode dan memberikan kesimpulan.



Gambar 2.1: Diagram Alur

BAB III

PEMBAHASAN

3.1 Pengantar *Bootstrap Aggregating* (*Bagging*)

Metode *Bootstrap Aggregating* (*bagging*) pertama kali digunakan oleh Breiman (1994). *Bagging* digunakan untuk memperbaiki stabilitas dan kekuatan prediksi dengan mereduksi variansi dari suatu prediktor pada metode klasifikasi dan regresi yang penggunannya tidak dibatasi hanya untuk memperbaiki estimator.

Bagging bekerja dengan baik pada metode klasifikasi yang menghasilkan parameter yang tidak stabil, yang mana perubahan kecil pada data *set* akan menghasilkan perubahan besar pada model yang diperoleh. Beberapa metode yang tidak stabil adalah *neural networks*, regresi, dan pohon regresi *(regression tree)*. Sedangkan contoh metode yang stabil adalah *k-nearest neighboor* (Breiman, 1994). Oleh karena itu, peningkatan ketepatan klasifikasi dengan *bagging* memberikan hasil yang baik jika model yang dihasilkan dari perulangan berbeda satu sama lain.

Bagging sendiri merujuk pada pendekatan *bootstrap* yang mana dari *bootstrap* pada data nantinya akan didapat suatu parameter, dimana setiap parameter akan diaggregat dengan melalukan rata-rata untuk mendapatkan model baru. Pendekatan *bootstrap* yang seperti ini dinamakan *bagging*. *Bootstrap* sendiri adalah metode berbasis resampling data sampel dengan syarat pengembalian pada datanya dalam menyelesaikan statistik ukuran suatu sampel dengan harapan sampel tersebut mewakili data populai sebenarnya. *Bootstrap*

memungkinakan untuk melakukan inferensi statistik tanpa membuat asumsi distribusi yang kuat dan tidak memerlukan formulasi analitis untuk distribusi sampling suatu estimator. Sebagai pengganti, *bootstrap* menggunakan distribusi empiris untuk mengestimasi distribusi sampling. Jadi jika penyelesaian analitik tidak mungkin dilakukan dimana anggapan (suatu distribusi, misalnya kenormalan data) tidak dipenuhi maka dengan menggunakan *bootstrap* masih dapat dilakukan suatu inferensi.

Bagging prediktor adalah metode untuk membangkitkan *multiple versions* dari prediktor dan menggunakannya untuk mendapatkan kumpulan prediktor. *Multiple versions* dibentuk dengan replikasi *bootstrap* dari sebuah data percobaan (Breiman, 1994). *Bagging* regresi logistik bekerja dengan cara mengambil n bagian contoh dari data asli secara acak dengan pengembalian sebanyak B kali dan memodelkan regresi logistik untuk setiap bagian contoh pada setiap B pengulangan. Prediksi dari setiap pengamatan adalah rata-rata dari prediksi B.

3.2 Metode *Bagging*

Sebuah data D terdiri dari $\{(\mathrm{y_n}, \mathrm{x_n}), \mathrm{n}{=}1, \ldots, \mathrm{N}\}$ dimana y dapat berupa kelas label atau respon numerik. Jika input adalah x, maka y diprediksi dengan $\varphi(x, D)$ dimana $\varphi(x, D)$ adalah prediktor. Untuk mendapatkan prediktor yang lebih baik dibandingkan dengan prediktor tunggal maka dilakukan replikasi *bootstrap* $\{D_k\}$ yang kemudian disebut $\{\varphi(x, D_k)\}$ dimana $\{D_k\}$ adalah rangkaian data yang terdiri dari N observasi independen. Replikasi *bootstrap* dilakukan sebanyak B kali sehingga ada $\left\{D^B\right\}$ dari D dan dibentuk prediktor $\left\{\varphi(x, D)^B\right\}$. $\left\{D^B\right\}$ adalah resampling dengan pengembalian, dimana $\left\{D^B\right\}$ diambil secara acak, tetapi dengan pengembalian dari $\{D\}$. Setiap

$\{(\mathrm{y_n}, \mathrm{x_n})$ dapat muncul berulang kali karena proses pengembalian atau tidak muncul sama sekali dalam $\left\{D^B\right\}$.

Aggregate classifier atau metode klasifikasi agregat μ_A diberikan secara umum dalam:

$$\mu_A(\mathrm{y}) = E_F\left[\widehat{\mu}(\mathrm{y}, D_k)\right] \tag{3.1}$$

dimana ekspektasi sampel D_k berdasarkan distribusi probabilitas F (fungsi distribusi empirik). Breiman mengenalkan prosedur *bagging* dimana syarat $\mu_A(y)$ diperoleh dengan membangkitkan B dari D dengan metode *bootstrap*. Untuk setiap resampel *bootstrap* dapat dihitung:

$$\widehat{\mu}^{(*b)}\left(\mathrm{y}, {D_k}^{(*b)}\right), b = 1, \ldots, B \tag{3.2}$$

dan kemudin menaksir *classifier* sebagai berikut.

$$\widehat{\mu}_A = E_{\widehat{F}}\left[\widehat{\mu}(\mathrm{y}, {D_k}^*)\right] \tag{3.3}$$

3.3 Ketepatan Klasifikasi *Bagging*

Beberapa metode klasifikasi menaksir peluang $\widehat{p}(\mathrm{j}\,|\mathrm{x})$ suatu objek dengan prediksi vektor x termasuk dalam kelas j. Kemudian kelas yang sesuai dengan x ditaksir sebagai $max_j\ \widehat{p}(\mathrm{j}\,|\mathrm{x})$. Untuk metode seperti ini, ketepatan *bagging* dengan voting (peluang terbesar) adalah rata-rata $\widehat{p}(\mathrm{j}\,|\mathrm{x})$ dari semua replikasi *bootstrap* sehingga diperoleh $\widehat{p}_{\mathrm{B}}(\mathrm{j}\,|\mathrm{x})$ dan kemudian menggunakannya untuk menaksir kelas $max_j\ \widehat{p}(\mathrm{j}\,|\mathrm{x})$. Penaksir ini dihitung di setiap klasifikasi *bootstrap* yang dilakukan. Hasil kesalahan klasifikasi selalu identik dengan kesalahan klasifikasi di tiap replikasi *bootstrap*. Sejauh ini fakta-fakta mengindikasikan bahwa *bagging* menaksir lebih akurat daripada penaksir tunggal. Untuk menguji pernyataan ini dilakukan perbandingan kedua estimasi dengan

$\mathrm{p}^*(\mathrm{j}|\mathrm{x})$ dari nilai sebenarnya, perbandingan dilakukan dengan membandingkan prediksi eror dari hasil model tunggal adalah sebagai berikut.

$$e_s = |\widehat{p}\,(\mathrm{j}\,|\mathrm{x}) - \mathrm{p}^*(\mathrm{j}|\mathrm{x})\,| \tag{3.4}$$

dimana $\widehat{p}\,(\mathrm{j}\,|\mathrm{x})$ adalah penaksir peluang dari model tunggal dan $\mathrm{p}^*(\mathrm{j}|\mathrm{x})$ adalah peluang sebenarnya. Sedangkan prediksi eror untuk model hasil *bagging* adalah sebagai berikut.

$$e_B = |\widehat{p}_B\,(\mathrm{j}\,|\mathrm{x}) - \mathrm{p}^*(\mathrm{j}|\mathrm{x})\,| \tag{3.5}$$

dimana $p\,(\mathrm{j}\,|\mathrm{x})$ penaksir dari peluang pada setiap replikasi, sehingga prediksi eror dari model *bagging* merupakan hasil rata-rata prediksi eror pada setiap pengambilan sampel pada setiap B replikasi *bootstrap* (Breiman, 1994).

Tahapan *Bagging* Regresi Logistik

Selanjutnya tahapan-tahapan *bagging* pada regresi logistik adalah:

1. Mengambil sampel *bootstrap* sebanyak n dari data *set* dengan pengulangan sebanyak n. Pengambilan sampel sedemikian hingga setiap variabel agrregat dalam setiap observasi.

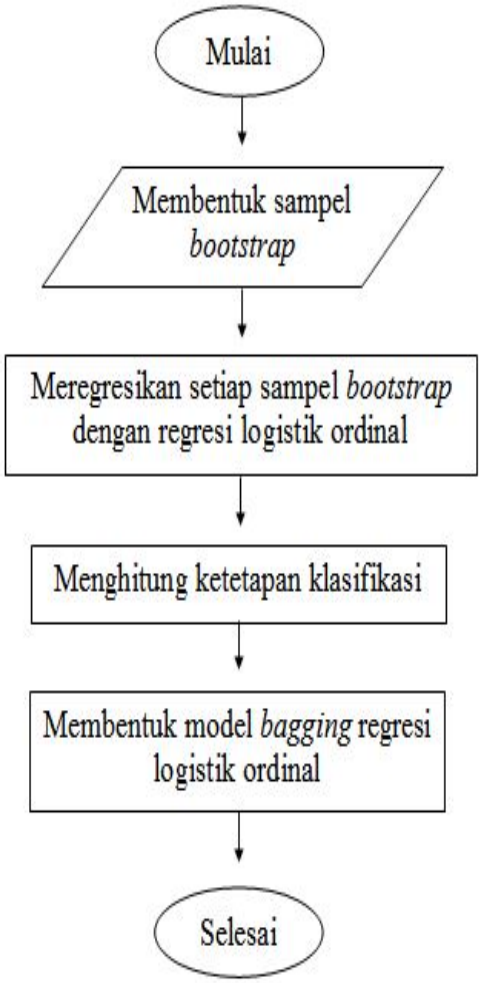
2. Memodelkan regresi logistik hasil sampel *bootstrap* D^B.

3. Menghitung ketepatan klasifikasi. Kesalahan klasifikasi pada langkah ini disebut e_B.

4. Mengulang langkah 1-3 sebanyak B kali replikasi *bootstrap*. (Replikasi yang digunakan sebanyak $50, 60, 70, 80, 90$, dan 100 kali replikasi)

5. Memperoleh ketepatan klasifikasi *bagging* yaitu rata-rata ketepatan klasifikasi setiap pengulangan sampai B. Sehingga kesalahan klasifikasi *bagging* untuk replikasi B kali adalah $\overline{e}_B$

6. Membentuk model *bagging* regresi logistik akhir.



Gambar 3.1: Diagram Alur *Bagging*

Jumlah Replikasi

Penentuan besarnya jumlah replikasi B sangat variatif, karena besar kecilnya B dapat memberikan hasil yang berbeda pada setiap tahapan analisis. Sutton (2005) merekomendasikan replikasi sebanyak 25 atau 50 kali, namun Hastie dkk (2001) menyatakan bahwa peningkatan akurasi akan terjadi jika banyaknya replikasi ditingkatkan dari 50 ke 100 kali, sedangkan jika banyak-

nya ditingkatkan hingga lebih dari 100 kali akan menghasilkan akurasi yang tidak lebih besar dari akurasi replikasi 100 kali. Sementara itu, Efron dan Tibshirani (1993) merekomendasikan nilai B yang kecil, misalnya 25 kali. Hasil penelitian sebelumnya oleh Suryaningrum, *bagging* dapat meningkatkan nilai ketepatan klasifikasi sebesar 0.2% dari 97.8% menjadi 98% dengan replikasi sebanyak 60 kali. Sedangakan penelitian Reza, dengan replikasi sebanyak 50 dapat meningkatan nilai ketepatan klasifikasi sebanyak 2% dari 49% menjadi 51%. Sehingga dalam penelitian ni digunakan replikasi sebanyak 50 sampai 100 kali replikasi.

3.4 Contoh Kasus

3.4.1 Sumber Data

Data yang digunakan dalam metode penelitian ini yaitu data SAKERNAS bulan Agustus 2014 di DKI Jakarta. Peubah respon yang digunakan ada tiga, yaitu berapa Lama mencari pekerjaan (Y)

1. kurang dari 12 bulan (Y_1)
2. 12 sampai kurang dari 24 bulan (Y_2)
3. 24 bulan atau lebih (Y_3)

Sedangkan peubah penjelas yang dipakai adalah sebagai berikut.

1. Hubungan dengan Kepala Rumah Tangga (KRT) (X_1)
    1. Kepala Rumah Tangga ($X_{1(1)}$)
    2. Istri/suami ($X_{1(2)}$)
    3. Anak ($X_{1(3)}$)
    4. Lainnya (cucu, menantu, orang tua/mertua, dll) ($X_{1(4)}$)
2. Jenis Kelamin (X_2)

    1. Laki-laki ($X_{2(1)}$)
    2. Perempuan ($X_{2(2)}$)
3. Umur (X_3)
4. Status Pernikahan (X_4)
    1. Belum nikah ($X_{4(1)}$)
    2. Nikah ($X_{4(2)}$)
    3. Lainnya ($X_{4(3)}$)
5. Pendidikan tertinggi yang ditamatkan (X_5)
    1. SD ($X_{5(1)}$)
    2. SMP ($X_{5(2)}$)
    3. SMA/SMK ($X_{5(3)}$)
    4. Universitas ($X_{5(4)}$)
6. Mengikuti pelatihan kursus (X_6)
    1. Ya ($X_{6(1)}$)
    2. Tidak ($X_{6(2)}$)
7. Penghasilan (dalam juta) (X_7)
8. Pekerjaan yang dicari (X_8)
    1. Pekerjaan purna waktu (Full Time) ($X_{8(1)}$)
    2. Pekerjaan paruh waktu (Part Time) ($X_{8(2)}$)

3.4.2 Hasil dan Pembahasan

Analisis regresi logistik ordinal pada contoh kasus ini, digunakan untuk mengetahui faktor apa saja yang mempengaruhi cepat atau lambatnya masya-

rakat dalam mendapat pekerjaan dan untuk mengukur nilai ketepatan klasifikasi. Peubah respon yang digunakan terdiri dari 3 kategori dan pada tiap kategori ada tingkatannya sehingga digunakan regresi logistik untuk menganalisisnya.

Data bersumber dari Badan Pusat Statistik (BPS) yaitu data SAKERNAS berjumlah 374 orang di DKI Jakarta pada bulan Agustus 2014. Karakteristik calon pekerja perlu dikaji guna mengetahui jangka waktu dalam mendapat kerja. Pembagian rentang waktu dibagi menjadi tiga yaitu pada kategori pertama sebanyak 107 orang, kategori kedua sebanyak 142 orang, dan kategori terakhir sebanyak 125 orang. Sehingga total seluruhunya adalah 374 orang.

Ada banyak faktor yang dapat mempengaruhi cepat lambat mendapat pekerjaan. Faktor yang diduga berpengaruh tersebut antara lain usia, jenis kelamin, status keluarga, status pernikahan, jenjang pendidikan terakhir, penghasilan sebelumnya, dan jenis pekerjaan yang diambil.

Pada analisis regresi logistik ordinal terdapat dua pengujian yaitu pengujian secara serentak dan pengujian secara simultan. Sebelumnya data dibagi dua terlebih dahulu yaitu data *training* yang digunakan untuk model dan data *testing* yang digunakan untuk ketepatan klasifikasi. Data dibagi menjadi 80 % untuk data *training* yaitu berjumlah 299 dan 20 % untuk data *testing* berjumlah 75. Selanjutnya dijelaskan hasil pengujian serentak dan parsial.

Pengujian Secara Serentak

Pengujian secara serentak ini dilakukan untuk mengetahui signifikansi peubah penjelas secara bersama-sama. Hipotesis yang digunakan adalah sebagai berikut.

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_8 = 0$

H_1 : paling sedikit terdapat satu $\beta_k \neq 0$ dengan $k = 1, 2, \ldots, 8$

Taraf signifikansi yang digunakan yaitu $\alpha = 10\%$ dengan statistik uji G yang mengikuti distribusi *Chi-square*. Berdasarkan hasil pengujian serentak,dapat diketauhi nilai $p - value = 0.000$ dan nilai $G = 126.106$. Hasil ini menandakan bahwa tolak H_0 karena $p - value < \alpha$ sehingga dapat dikatakan paling tidak ada satu peubah penjelas yang berpengaruh secara signifikan terhadap peubah respon.

Pengujian Secara Parsial

Pengujian secara parsial dilakukan karena pada pengujian serentak telah diindikasikan jika paling tidak ada satu peubah penjelas yang berpengaruh signifikan terhadap peubah respon. Hipotesis yang digunakan adalah sebagai berikut.

$H_0 : \beta_k = 0$

$H_1 : \beta_k \neq 0$ dengan $k = 1, 2, \ldots, 8$

Tabel 3.1: Tabel Regresi Logistik

Variabel	Coef	P-Value
$X_{1(2)}$	0.059	0.943
$X_{1(3)}$	0.439	0.410
$X_{1(4)}$	0.360	0.539
$X_{2(2)}$	0.095	0.723
X_3	-0.041	0.020*
$X_{4(2)}$	0.796	0.045*
$X_{4(3)}$	1.057	0.153
$X_{5(2)}$	0.894	0.214
$X_{5(3)}$	3.532	0.000*
$X_{5(4)}$	4.238	0.000*
$X_{6(2)}$	0.519	0.331
X_7	0.502	0.002*
$X_{8(2)}$	0.610	0.128
Konstanta 1	-4.474	0.000*
Konstanta 2	-2.342	0.000*

Statistik uji yang dilakukan adalah statistik uji Wald dengan tingkat keper-

cayaan atau tingkat signifikansi yang digunakan yaitu $\alpha = 10\%$. Berdasarkan hasil pengamatan peubah penjelas yang memiliki nilai *p-value* kurang dari $\alpha = 10\%$ berarti berpengaruh signifikan terhadap peubah respon adalah umur (X_3), status perkawinan (X_4), pendidikan terakhir (X_5), dan penghasilan sebelumnya (X_7).

Berdasarkan estimasi parameter yang telah didapatkan, maka fungsi logitnya dapat ditulis sebagai berikut.

$$\text{Logit1=-4.474-0.041}x_3\text{+0.796}x_{4(2)}\text{+3.532}x_{5(3)}\text{+4.238}x_{5(4)}\text{+0.502}x_7$$

$$\text{Logit2=-2.342-0.041}x_3\text{+0.796}x_{4(2)}\text{+3.532}x_{5(3)}\text{+4.238}x_{5(4)}\text{+0.502}x_7$$

Setelah didapatkan fungsi logit, selanjutnya dapat ditentukan nilai peluang kumulatif kategori respon dan peluang dari masing-masing kategori respon. Sebagai contoh dapat dihitung peluang usia calon pekerja 26 tahun, belum menikah, lulusan SMA/SMK, dan belum memiliki penghasilan sebelumnya pada setiap kategori respon.

$$\text{Logit1=-4.474-0.041}\,(26)\,\text{+0.796}\,(0)\,\text{+3.532}\,(0)\,\text{+4.238}\,(1)\,\text{+0.502}\,(0) = -1.302$$

$$\text{Logit2=-2.342-0.041}\,(26)\,\text{+0.796}\,(0)\,\text{+3.532}\,(0)\,\text{+4.238}\,(1)\,\text{+0.502}\,(0) = 0.83$$

Maka nilai peluang masing-masing peubah respon yaitu,

$$\pi_1 = P\,(Y \leq 1\,|X\,) = \frac{1}{1+\exp\,(\text{-Logit1})} = \frac{1}{1+\exp\,(1.302)} = 0.2138$$

$$\pi_2 = P\,(Y \leq 2\,|X\,) - P\,(Y \leq 1\,|X\,) = \frac{1}{1+\exp\,(-0.83)} - \frac{1}{1+\exp\,(1.302)} = 0.4825$$

$$\pi_3 = 1 - P\,(Y \leq 2\,|X\,) = 1 - \frac{1}{1+\exp\,(-0.83)} = 0.3036$$

Sehingga seorang yang berusia 26 tahun, belum menikah, lulusan SMA/SMK, dan belum memiliki penghasilan akan mendapatkan perkejaan dalam waktu satu tahun karena memiliki peluang yang paling besar.

Uji Kesesuaian Model

Setelah didapat estimasi parameter dan telah dibentuk model logit maka selanjutnya adalah pengujian kesesuaian model. Pengujian ini dilakukan untuk mengetahui apakah ada perbedaan antara hasil observasi dengan hasil prediksi. Hipotesis yang digunakan dalam pengujian adalah sebagai berikut.

H_0: model sesuai (tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

H_1: model tidak sesuai (ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

Berdasarkan tabel tersebut diketahui jika nilai $p - value$ uji (Hosmer & Lemenshow, 2000) 0.572 sudah lebih dari α sebesar 0.1 sehingga keputusannya gagal tolak H_0. Hal ini artinya model yang diperoleh dapat dikatakan jika tidak ada perbedaan yang signifikan antara hasil observasi dengan hasil prediksi model.

Interpretasi Model

Interpretasi dalam regresi logistik menggunakan nilai *odds ratio* yang menunjukkan perbandingan tingkat kecenderungan dari kategori yang ada dalam satu peubah penjelas.

Tabel 3.2: Hasil *Odds Ratio*

Variabel	*Odds Ratio*
X_3	0.96
$X_{4(2)}$	2.22
$X_{5(3)}$	34.19
$X_{5(4)}$	69.29
X_7	1.65

Peubah penjelas usia calon pekerja memiliki nilai *odds ratio* 0.96 dan koefisien negatif. Hal ini menunjukan bahwa seorang yang berusia lebih tua satu

tahun untuk pekerjaan kurang dari setahun memiliki resiko 0.96 lebih kecil dengan usia yang lebih muda satu tahun.

Pada peubah status pernikahan, kategori yang berpengaruh adalah seorang yang sudah menikah. Jika dilihat dari *odds ratio* pada kategori ini resiko mendapat kerja dalam kurun waktu kurang dari setahun yaitu 2.22 lebih besar dibanding yang belum menikah

Pada peubah jenjang pendidikan terakhir, kategori yang berpengaruh adalah lulusan SMA/SMK dan lulusan Perguruan tinggi. Untuk lulusan SMA/ SMK memiliki resiko 34.29 lebih besar dari lulusan SD dalam mendapat kerja dalam kurun waktu kurang dari setahun. Sedangkan ulusan Perguruan tinggi memiliki resiko 69.29 lebih besar dari lulusan SD untuk mendapat kerja yang lebih cepat.

Pada calon yang memiliki penghasilan sebelumnya memiliki resiko 1.65 kali lebih tinggi daripada yang belum meiliki penghasilan untuk mendapat pekerjan dengan cepat yaitu kurang dari satu tahun.

Ketepatan Klasifikasi

Salah satu kriteria kebaikan model adalah ketepatan klasifikasi. Suatu model yang baik digunakan adalah model yang memiliki ketepatan klasifikasi yang tinggi. Nilai ketepatan klasifikasi menggunakan data *testing* karena data *testing* merupakan suatu data yang dianggap baru yang digunakan untuk mengecek model yang dihasilkan oleh data *training*. Ketepatan klasifikasi ditampilkan pada data berikut.

Berdasarkan tabel, diketahui bahwa nilai APER (dalam %)

$$\left(\frac{4}{75}+\frac{2}{75}+\frac{11}{75}+\frac{12}{75}+\frac{1}{75}+\frac{2}{75}\right)*100\% = 42.67\%$$

Sehingga nilai ketepatan klasifikasinya adalah $1 - APER = 57.33\%$

Tabel 3.3: Ketepatan Klasifikasi Regresi Logistik Ordinal

Kelompok Pengamatan	Kelompok Prediksi		
	y_1	y_2	y_3
y_1	5	4	2
y_2	11	24	12
y_3	1	2	14

Nilai ketepatan klasfikasi yang didapat adalah 57.33 % dimana angka tersebut masih terbilang kecil. Oleh karena itu, untuk meningkatkan nilai ketepatan klasifikasi dapat dilalukan metode *Bootstrap Aggregating (Bagging)*.

Bootstrap Aggregating (Bagging)

Pemodelan regresi ordinal pada data *training* telah menghasilkan nilai ketepatan klasifikasi data *testing*. Model yang dihasilkan tersebut merupakan model tunggal dari data sebenarnya. Model tunggal menghasilkan ketepatan klasifikasi yang relatif masih kecil. *Bagging* ini merupakan salah satu metode yang pada umumnya mampu meningkatkan ketepatan klasifikasi. Oleh karena itu, pada data ini digunakan *bagging* dengan harapan mampu memperbaiki model sehingga mampu meningkatkan ketepatan klasifikasi.

Sama halnya dengan regresi logistik ordinal, pembentukan model *bagging* juga menggunakan data *training*. Kemudian data direplikasi sebanyak 50, 60, 70, 80, 90, dan 100. Program *bagging* ini dijalankan dengan program R. Berikut adalah hasil dari ketepatan klasifikasi *bagging*.

Pada tabel tersebut dapat dilihat bahwa nilai ketepatan klasifikasi replikasi ke-70 memiliki kenaikan yang tertinggi. Jika dibandingkan dari nilai ketepatan klasifikasi sebelumnya *bagging* dapat meningkatkan nilai ketepatan klasifikasi. Sehingga dapat dibuktikan bahwa metode *bagging* regresi logistik ordinal dapat meningkatkan nilai ketepatan klasifikasi pada regresi logistik ordinal.

Dengan demikian, ukuran *bootstrap* B tidak berpengaruh terhadap tingkat

Tabel 3.4: Hasil Klasifikasi *Bagging* Regresi Logistik Ordinal

Bagging Regresi Logistik Ordinal	
Replikasi	Ketepatan Klasifikasi
50	58.67%
60	58.67%
70	61.33%
80	60.00%
90	58.67%
100	60.00%

ketepatan klasifikasi. Selanjutnya model *bagging* yang diambil adalah model dengan ketepatan klasifikasi tertinggi yaitu pada replikasi ke-70. Sehingga model yang dibentuk adalah sebagai berikut.

$$\text{Logit1}=-4.565-0.038x_3+1.296x_{4(2)}+3.593x_{5(3)}+4.310x_{5(4)}+0.548x_7$$

$$\text{Logit2}=-2.621-0.038x_3+1.296x_{4(2)}+3.593x_{5(3)}+4.310x_{5(4)}+0.548x_7$$

Setelah didapatkan fungsi logit, selanjutnya dapat model diinterpretasikan dengan menggunakan nilai *odds ratio*. Berikut adalah hasil nilai *odds ratio*.

Tabel 3.5: Hasil *Odds Ratio*

Variabel	*Odds Ratio*
X_3	0.96
$X_{4(2)}$	3.65
$X_{5(3)}$	36.34
$X_{5(4)}$	74.44
X_7	1.73

Peubah penjelas usia calon pekerja memiliki nilai *odds ratio* 0.96 dan koefisien negatif. Hal ini menunjukan bahwa seorang yang berusia lebih tua satu tahun untuk pekerjaan kurang dari setahun memiliki resiko 0.96 lebih kecil dengan usia yang lebih muda satu tahun.

Pada peubah status pernikahan, kategori yang berpengaruh adalah seorang yang sudah menikah. Jika dilihat dari *odds ratio* pada kategori ini resiko mendapat kerja dalam kurun waktu kurang dari setahun yaitu 3.65 lebih besar dibanding yang belum menikah

Pada peubah jenjang pendidikan terakhir, kategori yang berpengaruh adalah lulusan SMA/SMK dan lulusan Perguruan tinggi. Untuk lulusan SMA/SMK memiliki resiko 36.34 lebih besar dari lulusan SD dalam mendapat kerja dalam kurun waktu kurang dari setahun. Sedangkan lulusan Perguruan tinggi memiliki resiko 74.44 lebih besar dari lulusan SD untuk mendapat kerja yang lebih cepat.

Pada calon yang memiliki penghasilan sebelumnya memiliki resiko 1.73 kali lebih tinggi daripada yang belum memiliki penghasilan untuk mendapat pekerjan dengan cepat yaitu kurang dari satu tahun.

BAB IV

PENUTUP

4.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, didapatkan kesimpulan sebagai berikut.

1. Regresi logistik ordinal merupakan metode untuk mengetahui hubungan antara peubah penjelas dengan peubah respon yang bersifat kategori bertingkat. Metode ini juga dapat digunakan sebagai proses pengklasifikasian. Berdasarkan hasil pembahasan, nilai klasifikasi ditampilkan pada tabel berikut.

Tabel 4.1: Ketepatan Klasifikasi Regresi Logistik Ordinal

Kelompok Pengamatan	Kelompok Prediksi		
	y_1	y_2	y_3
y_1	5	4	2
y_2	11	24	12
y_3	1	2	14

Berdasarkan tabel, diketahui bahwa nilai APER (dalam %)

$$\left(\frac{4}{75}+\frac{2}{75}+\frac{11}{75}+\frac{12}{75}+\frac{1}{75}+\frac{2}{75}\right)*100\% = 42.67\%$$

Sehingga nilai ketepatan klasifikasinya adalah $1 - APER = 57.33\%$

2. Hasil analisis *bagging* regresi logistik ordinal menunjukkan bahwa pada untuk replikasi *bootstrap* 70 kali diperoleh ketetapan klasifikasi untuk data *testing* sebesar 61.33 % meningkat dari ketepakan klasifikasi regresi

logistik ordinal yaitu 57.33 %. Hasil ini menunjukkan bahwa *bagging* regresi logistik ordinal telah meningkatkan ketetapan klasifikasi dan menurunkan tingkat kesalahan klasifikasi pada regresi logistik ordinal.

4.2 Saran

1. Berdasarkan hasil analisis dan pembahasan, dapat dikatakan bahwa hal yang berpengaruh pada jangka waktu dalam mencari kerja, adalah pendidikan terakhir yang telah dijalani calon tenaga kerja. Sehingga hendaknya pendidikan para calon tenaga kerja dapat terus ditingkatkan guna mempercepat mendapat pekerjaan.
2. Pada penelitian selanjutnya bisa ditambahkan peubah penjelas lain yaitu peubah yang memungkinkan lebih berpengaruh pada model atau dilakukan replikasi *bootstrap* yang lebih banyak lagi untuk hasil klasifikasi yang lebih baik.

DAFTAR PUSTAKA

Agresti, A. 2007. *Categorical Data Analysis*. New York : John Wiley and Sons.

Badan Pusat Statistik. 2014. *Survey Angkatan Kerja Nasional*. Jakarta: BPS

Breiman, L. 1994. *Bagging Predictor*. Technical report No. 421. Departement of statistics University of California.

Efron, B. dan Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Chapman Hall, New York

Hastie, T., Tibshirani, R., Friedman, J. 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer-Verlag, New York.

Hosmer, D.W., Lemenshow. 2000. *Applied Logistic Regression*. USA : John Wiley and Sons.

J. G. Dias dan J. K. Vermunnt. 2005. "A Bootstrap Based Aggregate Classifier for Model Based Clustering". *Journals of Analysis Statistics, Vol. 23:643-659*

Johnson, R. A., Wichern, D. W. 1992. *Applied Multivariate Statistical Analysist Third Edition*. New Jersey: Prentice Hall.

Reza, Ahmad. 2015. *Ketepatan Klasifikasi Pemilihan Metode Kontrasepsi di Kota Semarang Menggunakan Bootstrap Aggregating Regresi Logistik Multinomial. Jurnal Gaussian, Volume 4, Nomor 1: 11-20*. Universitas Diponegoro.

Supratikno, Andi. 2011. *Faktor-faktor Yang Mempengaruhi Lama Mencari Kerja Bagi Tenaga Kerja Terdidik di Kabupaten Semarang.* Skripsi. Universitas Diponegoro.

Suryaningrum. 2011. *Klasifikasi Kesejahteraan Rumah Tangga di Kota Malang dengan Pendekatan Bagging Regresi Logistik.* Skripsi. Institut Teknologi Sepuluh Nopember.

Sutton CD. 2005. "Classification and regression trees, Bagging, and Boosting". *Handbook of statistics, Vol. 24:303-329.*

LAMPIRAN-LAMPIRAN

LAMPIRAN 1

Tabel 4.2: Data SAKERNAS

No	y	x1	x2	x3	x4	x5	x6	x7	x8
1	3	3	2	22	1	3	2	0	1
2	3	3	1	30	1	1	2	0.6	1
3	3	1	1	46	2	3	1	3.75	1
4	1	3	1	17	1	3	2	0	1
5	3	3	1	25	1	3	2	0.6	1
6	2	1	2	37	3	2	2	0.5	2
7	3	3	1	20	1	3	2	0	1
8	3	1	2	47	3	2	2	1.7	1
9	2	3	1	27	1	3	2	0.8	1
10	2	3	1	20	1	3	2	0.8	1
11	1	3	1	23	1	3	2	0	2
12	1	4	2	25	2	3	2	0	2
13	1	3	2	25	1	4	2	0	2
14	2	3	2	18	1	3	2	0	1
15	2	3	1	38	1	2	2	1	1
16	2	3	2	28	3	3	2	0.5	1
17	1	2	2	37	2	1	2	1.7	2
18	2	3	2	24	2	3	2	0	1
19	2	4	1	46	1	3	2	0	1
20	1	1	1	34	2	4	1	0.55	1
21	1	3	2	32	3	3	2	0	2
22	3	1	1	55	3	1	2	1	1
23	1	3	1	19	1	3	2	1.2	2
24	2	4	1	21	1	3	2	0.7	2
25	3	3	2	22	1	3	2	0	1
26	1	3	2	22	2	4	2	0.95	2
27	1	3	1	18	1	3	2	0	1
28	2	1	1	34	2	3	2	0	1
29	1	3	1	21	1	3	2	0	1
30	2	3	1	25	2	3	2	0	1

No	y	x1	x2	x3	x4	x5	x6	x7	x8
31	2	3	2	20	2	3	2	0	1
32	2	3	2	19	1	3	2	0	1
33	2	3	2	20	1	3	2	0	1
34	2	4	2	20	1	3	2	0	1
35	1	3	2	24	1	4	2	0	1
36	1	3	1	25	1	4	2	0	1
37	2	1	1	22	3	3	2	0	1
38	3	4	1	24	1	2	1	0	1
39	2	3	1	20	1	3	2	0	1
40	2	3	1	37	1	3	2	0	1
41	1	4	1	36	2	4	2	0	1
42	1	1	1	26	2	3	2	0	1
43	3	3	2	30	1	4	2	0	1
44	1	3	1	25	1	4	1	5.75	1
45	2	3	1	19	1	3	2	0	1
46	2	3	1	19	1	3	2	0	1
47	3	3	1	24	1	3	2	0	1
48	2	3	2	23	1	3	2	0	1
49	2	3	1	30	1	3	2	0	1
50	1	3	1	24	1	4	2	0	1
51	3	3	2	36	1	4	2	0	1
52	1	3	1	32	1	4	2	0	1
53	2	3	1	31	1	3	2	0	1
54	2	3	2	17	1	3	2	0	1
55	3	3	1	15	1	1	2	0	1
56	3	3	2	24	1	2	2	0	1
57	1	3	1	44	1	3	2	1.56	1
58	2	3	2	18	1	3	2	0	1
59	2	3	1	19	1	3	2	0	1
60	2	3	1	18	1	3	2	0	1

No	y	x1	x2	x3	x4	x5	x6	x7	x8
61	2	3	1	25	1	3	1	0	1
62	2	3	2	17	1	3	2	0	1
63	2	3	1	24	1	3	1	0	1
64	1	3	2	23	1	4	2	2.7	1
65	1	3	2	19	1	3	2	0	1
66	2	3	2	35	1	4	2	0	1
67	2	3	2	34	1	4	2	0	1
68	1	3	1	22	1	3	2	0	1
69	3	3	1	24	1	3	2	0	1
70	1	3	1	28	1	4	2	0	1
71	2	3	2	19	1	3	2	0	1
72	2	3	1	22	1	3	2	0	1
73	2	3	2	19	1	3	2	0	1
74	2	3	2	28	2	4	2	0	1
75	1	3	1	23	1	4	2	0	1
76	1	3	1	22	1	3	2	0	1
77	1	3	2	24	1	4	1	3.1	1
78	2	3	1	21	1	3	2	0	1
79	3	4	1	35	1	2	2	2.9	1
80	1	3	2	19	1	3	2	2	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
361	3	3	1	18	1	2	2	0	1
362	2	4	1	35	2	2	2	0.36	1
363	2	3	2	31	1	4	2	0.3	1
364	2	3	1	20	1	3	2	1.05	1
365	1	3	1	24	1	3	2	0	1
366	3	3	1	23	1	3	2	0	1
367	1	3	2	19	1	3	2	0	1
368	3	3	2	30	1	3	2	0	1
369	3	3	1	21	1	1	2	0	1
370	2	3	1	26	1	3	2	0	1
371	1	1	1	50	2	3	2	0	2
372	1	3	1	18	1	3	2	0	1
373	3	3	1	23	1	3	2	0	1
374	2	3	2	18	1	3	2	0	1

LAMPIRAN 2

Tabel 4.3: Data Training

No	y	x1	x2	x3	x4	x5	x6	x7	x8
1	3	3	2	22	1	3	2	0	1
2	3	3	1	30	1	1	2	0.6	1
3	3	1	1	46	2	3	1	3.75	1
4	1	3	1	17	1	3	2	0	1
5	3	3	1	25	1	3	2	0.6	1
6	2	1	2	37	3	2	2	0.5	2
7	3	3	1	20	1	3	2	0	1
8	3	1	2	47	3	2	2	1.7	1
9	2	3	1	27	1	3	2	0.8	1
10	2	3	1	20	1	3	2	0.8	1
11	1	3	1	23	1	3	2	0	2
12	1	4	2	25	2	3	2	0	2
13	1	3	2	25	1	4	2	0	2
14	2	3	2	18	1	3	2	0	1
15	2	3	1	38	1	2	2	1	1
16	2	3	2	28	3	3	2	0.5	1
17	1	2	2	37	2	1	2	1.7	2
18	2	3	2	24	2	3	2	0	1
19	2	4	1	46	1	3	2	0	1
20	1	1	1	34	2	4	1	0.55	1
21	1	3	2	32	3	3	2	0	2
22	3	1	1	55	3	1	2	1	1
23	1	3	1	19	1	3	2	1.2	2
24	2	4	1	21	1	3	2	0.7	2
25	3	3	2	22	1	3	2	0	1
26	1	3	2	22	2	4	2	0.95	2
27	1	3	1	18	1	3	2	0	1
28	2	1	1	34	2	3	2	0	1
29	1	3	1	21	1	3	2	0	1
30	2	3	1	25	2	3	2	0	1

No	y	x1	x2	x3	x4	x5	x6	x7	x8
31	2	3	2	20	2	3	2	0	1
32	2	3	2	19	1	3	2	0	1
33	2	3	2	20	1	3	2	0	1
34	2	4	2	20	1	3	2	0	1
35	1	3	2	24	1	4	2	0	1
36	1	3	1	25	1	4	2	0	1
37	2	1	1	22	3	3	2	0	1
38	3	4	1	24	1	2	1	0	1
39	2	3	1	20	1	3	2	0	1
40	2	3	1	37	1	3	2	0	1
41	1	4	1	36	2	4	2	0	1
42	1	1	1	26	2	3	2	0	1
43	3	3	2	30	1	4	2	0	1
44	1	3	1	25	1	4	1	5.75	1
45	2	3	1	19	1	3	2	0	1
46	2	3	1	19	1	3	2	0	1
47	3	3	1	24	1	3	2	0	1
48	2	3	2	23	1	3	2	0	1
49	2	3	1	30	1	3	2	0	1
50	1	3	1	24	1	4	2	0	1
51	3	3	2	36	1	4	2	0	1
52	1	3	1	32	1	4	2	0	1
53	2	3	1	31	1	3	2	0	1
54	2	3	2	17	1	3	2	0	1
55	3	3	1	15	1	1	2	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
290	2	3	1	23	1	3	2	0	1
291	3	4	2	39	1	1	2	0	2
292	3	1	1	48	2	3	2	0	1
293	2	3	1	24	1	3	2	0	1
294	3	3	1	32	2	3	2	0	1
295	2	3	2	18	1	3	2	0	1
296	1	3	2	21	1	4	2	0	1
297	1	4	1	28	1	3	2	3.25	1
298	2	3	2	18	1	3	2	0	1
299	2	3	1	20	1	3	2	0	1

LAMPIRAN 3

Tabel 4.4: Data Testing

No	y	x1	x2	x3	x4	x5	x6	x7	x8
300	2	3	1	26	1	4	2	0	1
301	2	3	1	23	1	4	2	0	1
302	2	3	1	26	1	4	2	0	1
303	3	3	1	17	1	2	2	0	1
304	1	3	2	24	1	4	2	0	1
305	2	1	2	37	3	3	2	0	1
306	2	3	2	26	1	3	1	0	1
307	3	3	1	43	3	2	2	0	1
308	3	3	1	48	1	1	2	0	1
309	3	3	1	18	1	1	2	0	1
310	2	3	1	18	1	3	1	0	1
311	3	3	1	31	1	1	2	0	1
312	1	3	1	19	1	3	2	0	1
313	3	3	1	19	1	2	2	0	1
314	3	3	2	29	1	3	2	0	1
315	2	3	1	23	1	3	2	0	1
316	3	3	1	17	1	2	2	0	1
317	3	1	1	23	2	3	2	2	1
318	2	1	2	35	1	3	2	1.2	1
319	1	4	1	42	2	4	2	2.15	1
320	2	3	2	22	1	3	2	0	1
321	2	1	1	40	2	3	2	0	1
322	2	3	2	21	1	3	2	1.2	1
323	1	3	1	20	1	3	2	0	1
324	2	3	2	18	1	3	2	0	1
325	3	3	1	20	1	2	2	0	2
326	2	3	2	19	1	3	2	0	1
327	2	3	1	20	1	3	2	0	1
328	3	3	1	30	1	3	2	1.75	1
329	3	3	1	27	1	4	2	0	1
330	1	3	1	19	1	3	2	0	1

No	y	x1	x2	x3	x4	x5	x6	x7	x8
331	3	3	1	38	3	3	2	0	1
332	3	3	1	24	1	3	2	0	1
333	1	4	1	22	1	4	2	0	1
334	3	1	1	56	2	2	2	0	2
335	1	3	2	22	1	3	2	0	1
336	2	3	1	28	1	4	2	1.65	1
337	3	1	1	44	2	1	2	0	1
338	1	3	1	18	1	3	2	0	1
339	2	3	1	19	1	3	2	0	1
340	2	3	2	18	1	3	2	0	1
341	2	4	2	19	1	3	2	0	1
342	2	1	1	37	2	2	2	1.5	1
343	3	3	1	23	1	3	2	0	1
344	1	3	1	22	1	3	2	0.75	1
345	2	3	1	21	1	3	2	1.55	1
346	1	3	2	20	1	3	1	0.7	1
347	2	3	1	18	1	3	2	0	1
348	2	3	2	17	1	3	2	0	1
349	3	3	1	22	1	3	2	0	1
350	3	1	1	32	2	1	2	0	1
351	3	3	2	38	1	3	2	0	1
352	1	3	1	32	2	3	2	2.25	1
353	1	3	1	23	1	3	2	2.15	1
354	3	3	1	19	1	3	2	0	1
355	3	1	1	46	2	2	2	1.75	1
356	2	4	1	19	1	3	2	0	1
357	2	3	1	26	1	3	2	0	1
358	1	3	1	22	2	2	2	0.6	1
359	3	4	2	21	2	3	2	0	1
360	2	4	1	41	2	3	2	0.9	1
361	3	3	1	18	1	2	2	0	1
362	2	4	1	35	2	2	2	0.36	1
363	2	3	2	31	1	4	2	0.3	1
364	2	3	1	20	1	3	2	1.05	1
365	1	3	1	24	1	3	2	0	1
366	3	3	1	23	1	3	2	0	1
367	1	3	2	19	1	3	2	0	1
368	3	3	2	30	1	3	2	0	1
369	3	3	1	21	1	1	2	0	1
370	2	3	1	26	1	3	2	0	1
371	1	1	1	50	2	3	2	0	2
372	1	3	1	18	1	3	2	0	1
373	3	3	1	23	1	3	2	0	1
374	2	3	2	18	1	3	2	0	1

LAMPIRAN 4

Program R

```
#siapkan data
d1 < - sakernas
d1$y=as.ordered(d1$y)
d1$x1=as.factor(d1$x1)
d1$x2=as.factor(d1$x2)
d1$x4=as.factor(d1$x4)
d1$x5=as.factor(d1$x5)
d1$x6=as.factor(d1$x6)
d1$x8=as.factor(d1$x8)

summary(d1)
str(d1)
xtabs( y+x1, d1)
xtabs( y+x2, d1)
xtabs( y+x4, d1)
xtabs( y+x5, d1)
xtabs( y+x6, d1)
xtabs( y+x8, d1)

#partisi data (pembagian data training dan data testing)
ind=sample(2, nrow(d1), replace = TRUE, prob = c(0.2,0.8))
test1=d1[ind==1,]
train1=d1[ind==2,]
train=d1[1:299,]
test=d1[300:374,]
```

```
nrow(test)
nrow(train)
summary(test)
summary(train)

#model regresi logistik ordinal (proportional odds logistic regression)
library(MASS)
model1 <-polr(y x1+x2+x3+x4+x5+x6+x7+x8, train, Hess = TRUE)
summary(model1)

#menghitung p-value
(ctable=coef(summary(model1)))
p=pnorm(abs(ctable[, "t value"]), lower.tail = FALSE)*2
(ctable=cbind(ctable, "p value"=p))

#prediksi
pred <- predict(model1, train)
print(pred, digits = 3)

#confussion matrix data train
(tab <- table(pred,train$y))
1-sum(diag(tab))/sum(tab)
sum(diag(tab))/sum(tab)

#confussion matrix data test
pred1 <- predict(model1, test)
print(pred1, digits = 3)
```

```
(tab1 <- table(pred1,test$y))
1-sum(diag(tab1))/sum(tab1)
sum(diag(tab1))/sum(tab1)

#likelihood
library(lmtest)
lrtest(model1)

#odds ratio
model2 <-polr(y x2+x3+x4+x5, train, Hess = TRUE)
exp(coefficients(model2))
exp(coefficients(model1))

#bootstrap aggregatting
library(MASS)
library(foreach)

databagrlo <- sakernas
databagrlo$y=as.factor(databagrlo$y)
databagrlo$x4=as.factor(databagrlo$x4)
databagrlo$x5=as.factor(databagrlo$x5)

positions <- sample(nrow(databagrlo),size=floor((nrow(databagrlo)/5)*4))
training <- databagrlo[positions,]
testing <- databagrlo[-positions,]
```

```
lengthdivisor <- 2
iterations<- 50 #ubah replikasi

predictions <-foreach(m=1:iterations,.combine=cbind) %do%
{
trainingpositions <- sample(nrow(training), size=floor((nrow(training)/lengthdivisor)))
trainpos<-1:nrow(training) %in % trainingpositions
modelfit<-polr(y x3+x4+x5+x7,data=training[trainpos,])
pred2<-predict(modelfit,newdata=testing)
}
predictions<-rowMeans(predictions)

pred2<- predict(modelfit,newdata=testing)
(tab2 <- table(pred2,testing$y))
1-sum(diag(tab2))/sum(tab2)
sum(diag(tab2))/sum(tab2)

modelfit<-polr(y x3+x4+x5+x7,data=training, Hess = TRUE)
summary(modelfit)

(ctable=coef(summary(modelfit)))
p=pnorm(abs(ctable[, "t value"]), lower.tail = FALSE)*2
(ctable=cbind(ctable, "p value"=p))

#end
```

LAMPIRAN 4 Output

Summary data

```
y       x1       x2            x3          x4       x5       x6            x7          x8
 1:107   1: 49    1:252   Min.   :15.00   1:274    1: 40    1: 21   Min.   :0.0000   1:337
 2:142   2: 10    2:122   1st Qu.:20.00   2: 86    2: 41    2:353   1st Qu.:0.0000   2: 37
 3:125   3:276            Median :24.00   3: 14    3:224            Median :0.0000
         4: 39            Mean   :27.23            4: 69            Mean   :0.3555
                          3rd Qu.:31.00                             3rd Qu.:0.0000
                          Max.   :63.00                             Max.   :7.1000
```

Deskriptif Data

```
        x1
y     1   2   3   4
  1  10   2  83  12
  2  15   4 111  12
  3  24   4  82  15

        x2
y     1  2
  1 66 41
  2 91 51
  3 95 30

        x4
y     1   2   3
  1  74  31   2
  2 114  22   6
  3  86  33   6

        x5
y     1   2   3   4
  1   1   2  68  36
  2   3   8 110  21
  3  36  31  46  12

        x6
y     1   2
  1   8  99
  2   8 134
  3   5 120

        x8
y     1   2
  1  92  15
  2 133   9
  3 112  13
```

Summary data training (n=299)

```
 y       x1       x2            x3          x4       x5       x6            x7          x8
1: 90   1: 39    1:197   Min.   :15.00   1:216    1: 34    1: 18   Min.   :0.000   1:265
2:112   2: 10    2:102   1st Qu.:21.00   2: 72    2: 30    2:281   1st Qu.:0.000   2: 34
3: 97   3:218            Median :24.00   3: 11    3:175            Median :0.000
        4: 32            Mean   :27.49            4: 60            Mean   :0.365
                         3rd Qu.:31.00                             3rd Qu.:0.000
                         Max.   :63.00                             Max.   :7.100
```

Summary data testing (n=75)

```
 y       x1      x2            x3          x4      x5      x6            x7          x8
1:17   1:10    1:55    Min.   :17.00   1:58    1: 6    1: 3    Min.   :0.0000   1:72
2:30   2: 0    2:20    1st Qu.:19.00   2:14    2:11    2:72    1st Qu.:0.0000   2: 3
3:28   3:58            Median :23.00   3: 3    3:49            Median :0.0000
       4: 7            Mean   :26.17           4: 9            Mean   :0.3175
                       3rd Qu.:30.50                           3rd Qu.:0.0000
                       Max.   :56.00                           Max.   :2.2500
```

```
Tabel Resgresi Logistik

          Value Std. Error     t value      p value
x12 0.05839157 0.82401794 -0.07086202 9.435076e-01
x13 0.43870935 0.53257087 -0.82375768 4.100773e-01
x14 0.36029263 0.58655997 -0.61424689 5.390522e-01
x22 0.09505363 0.26912557 -0.35319434 7.239427e-01
x3  -0.04107636 0.01776411  2.31232317 2.075988e-02
x42 0.79645237 0.39799180 -2.00117785 4.537323e-02
x43 1.05672506 0.74021002 -1.42760167 1.534065e-01
x52 0.89393154 0.72082813 -1.24014519 2.149217e-01
x53 3.53186287 0.61488037 -5.74398373 9.247468e-09
x54 4.23834394 0.66207392 -6.40161741 1.537395e-10
x62 0.51905223 0.53353476 -0.97285552 3.306251e-01
x7  0.50223461 0.16164883 -3.10694853 1.890293e-03
x82 0.61007651 0.40048384 -1.52334864 1.276715e-01
1|2 -4.47402667 1.13200254 -3.95231152 7.739989e-05
2|3 -2.34161461 1.11966400 -2.09135473 3.649628e-02

Residual Deviance: 528.3557
AIC: 558.3557


Likelihood ratio test

Model 1: y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
Model 2: y ~ 1
 #Df  LogLik  Df  Chisq Pr(>Chisq)
1  15 -264.18
2   2 -327.23 -13 126.11  < 2.2e-16



Kepetapan klasifikasi untuk data training

pred  1  2  3
   1 44 21  6
   2 44 80 36
   3  2 11 55

APER 0.4013378

Kepetapan klasifikasi untuk data testing

pred1  1  2  3
  1  5  4  2
  2 11 24 12
  3  1  2 14

APER 0.4266667
```

SURAT PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya yang bertanda tangan di bawah ini, mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta:

Nama	:	Sharah Annisa
No. Registrasi	:	3125121986
Program Studi	:	Matematika

Menyatakan bahwa skripsi ini yang saya buat dengan judul "**Klasifikasi Jangka Waktu Mendapatkan Kerja di DKI Jakarta dengan Metode *Bootstrap Aggregating* Regresi Logistik Ordinal**" adalah :

1. Dibuat dan diselesaikan oleh saya sendiri.
2. Bukan merupakan duplikat skripsi yang pernah dibuat oleh orang lain atau jiplakan karya tulis orang lain.

Pernyataan ini dibuat dengan sesungguhnya dan saya bersedia menanggung segala akibat yang timbul jika pernyataan saya tidak benar.

Jakarta, Agustus 2017

Yang membuat pernyataan

Sharah Annisa

DAFTAR RIWAYAT HIDUP

SHARAH ANNISA. Lahir di Jakarta, 01 Maret 1994. Putri ketiga dari tiga bersaudara dari pasangan Bapak Wimbo Sasongko dan Ibu Millia Rachmawati. Saat ini penulis bertempat tinggal di Perumahan Bintara 3 B/57 RT 05 RW 07 Bintara Jaya Bekasi Barat 17136.

No. Ponsel : 085691288246

Email : sharah.annisa@gmail.com

Riwayat Pendidikan : Penulis mengawali pendidikan di TK Al-Amin, dan kemudian pada tahun 2000 melanjutkan pendidikan di SDN Pondok Kelapa 10 Pagi, lulus pada tahun 2006. Setelah itu, penulis melanjutkan ke SMPIT Tashfia Bekasi hingga tahun 2009. Kemudian kembali melanjutkan ke SMA Negeri 59 Jakarta dan lulus tahun 2012. Di Tahun yang sama penulis mengikuti SNMPTN tertulis dan diterima di Universitas Negeri Jakarta (UNJ), Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Jurusan Matematika, Program Studi Matematika. Di pertengahan tahun 2016 penulis telah memperoleh gelar Sarjana Sains untuk Jurusan Matematika, Program Studi Matematika, FMIPA, UNJ.

Riwayat Organisasi : Selama di bangku perkuliahan, penulis aktif di berbagai organisasi kemahasiswaaan. Pada tahun pertama, penulis mendapat kepercayaan sebagai staff Departemen Rohis BEMJ Matematika.

Riwayat Pekerjaan : Selain sebagai mahasiswa, penulis juga menjadi pengajar privat bagi siswa SD, SMP, dan SMA. Penulis juga pernah mengikuti program PKL di Badan Pusat Statistik (BPS) Jakarta selama 1 bulan.