



TESIS - SS14 2501

**PERBANDINGAN MODEL *NEURAL NETWORKS*
DENGAN *POISSON REGRESSION* DAN *NEGATIVE
BINOMIAL REGRESSION* PADA KASUS DEMAM
BERDARAH DENGUE (DBD) DI SURABAYA**

FASTHA AULIA PRADHANI
NRP. 1314 201 036

DOSEN PEMBIMBING
Dr. Kartika Fithriasari, M.Si
Dr. Brodjol Sutijo Suprih Ulama, M.Si

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016



TESIS - SS14 2501

**COMPARISON OF NEURAL NETWORKS MODEL,
POISSON REGRESSION, AND NEGATIVE
BINOMIAL REGRESSION IN THE CASE OF
DENGUE HAEMORRHAGIC FEVER (DHF) IN
SURABAYA**

FASTHA AULIA PRADHANI
NRP. 1314 201 036

SUPERVISOR
Dr. Kartika Fithriasari, M.Si
Dr. Brodjol Sutijo Suprih Ulama, M.Si

MAGISTER PROGRAM
STATISTICS DEPARTEMENT
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

**PERBANDINGAN MODEL *NEURAL NETWORKS*, DENGAN
POISSON REGRESSION DAN *NEGATIVE BINOMIAL
REGRESSION* PADA KASUS DEMAM BERDARAH DENGUE
(DBD) DI SURABAYA**

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

Di

Institut Teknologi Sepuluh Nopember

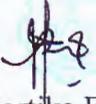
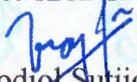
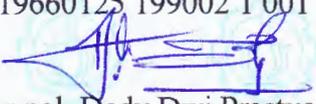
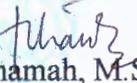
Oleh :

FASTHA AULIA PRADHANI
NRP. 1314 201 036

Tanggal Ujian : 28 Januari 2016

Periode Wisuda : Maret 2016

Disetujui Oleh :

- 
1. Dr. Kartika Fithriasari, M.Si (Pembimbing I)
NIP.19691212 199303 2 002
- 
2. Dr. Brodjol Sutjipto Suprih Ulama, M.Si (Pembimbing II)
NIP. 19660125 199002 1 001
- 
3. Dr. rer.pol. Dedy Dwi Prastyo, M.Si (Penguji)
NIP. 19831204 200812 1 002
- 
4. Irhamah, M.Si., Ph.D. (Penguji)
NIP. 19780406 200112 2 002



PERBANDINGAN MODEL *NEURAL NETWORKS* DENGAN *POISSON REGRESSION* PADA KASUS DEMAM BERDARAH DENGUE (DBD) DI SURABAYA

Nama mahasiswa : Fastha Aulia Pradhani
NRP : 1314201036
Pembimbing : Dr.Kartika Fithriasari, M.Si.
Co-Pembimbing : Dr. Brodjol Sutijo Suprih Ulama, M.Si

ABSTRAK

Demam berdarah atau demam dengue (DBD) adalah penyakit yang disebabkan oleh nyamuk *Aedes Aegypti* yang menyebarkan virus *dengue* pada tubuh manusia. Pada penelitian ini akan dilakukan pemodelan kasus DBD di Surabaya dengan menggunakan tiga model yaitu *Poisson Regression*, *Negative Binomial Regression*, dan *Neural Networks* (NN). Pemodelan kasus DBD telah banyak dilakukan, salah satunya menggunakan regresi poisson, namun pada banyak kasus asumsi *equidispersion* pada regresi poisson sulit dipenuhi. Salah satu alternatif yang dapat dipakai untuk mengatasi hal tersebut adalah memodelkan data dengan menggunakan model *negative binomial regression*. *Neural Network* merupakan salah satu teknik yang lebih fleksibel dalam penyelesaian kasus pemodelan hubungan variabel. Unit observasi yang digunakan ialah 31 kecamatan di Surabaya. Jumlah kasus DBD di tiap kecamatan sebagai variabel respon, dengan variabel bebas yaitu persentase rumah bebas jentik, rumah tangga miskin, PHBS, kepadatan penduduk, rasio tenaga medis, RT yang memiliki tempat sampah sehat, serta sarana Puskesmas. Data dibagi menjadi data *training* dan *testing* menggunakan *k-fold crossvalidation*. Arsitektur yang dipakai pada model *neural network* adalah *Multi Layer Perceptron*, dengan algoritma pembelajaran *backpropagation* dan metode perubahan bobot *Levenberg Marquardt*. Bobot awal diinisialisasi dengan nilai yang paling optimum setelah dilakukan pemilihan secara random sebanyak 100 kali. Kasus *overdispersi* terjadi pada model *poisson regression* yang dihasilkan, sehingga dilanjutkan dengan pemodelan menggunakan *negative binomial regression*, dan diperoleh hasil bahwa kasus *overdispersi* dapat teratasi, namun kasus *under estimate* belum dapat diatasi. Pada model (NN), arsitektur terbaik dihasilkan pada MLP (8-12-1). Selanjutnya dilakukan pemilihan model terbaik, dari ketiga model yang telah dihasilkan Kriteria pemilihan model terbaik ditentukan berdasarkan nilai CV testing, dan diperoleh hasil bahwa NN merupakan model terbaik yang dipakai untuk memprediksi banyaknya kasus DBD di Surabaya.

Kata kunci : DBD, *Poisson*, *Neural Network*

COMPARISON OF NEURAL NETWORKS MODEL, POISSON REGRESSION, AND NEGATIVE BINOMIAL REGRESSION IN THE CASE OF DENGUE HAEMORRHAGIC FEVER (DHF) IN SURABAYA

Name : Fastha Aulia Pradhani
NRP : 1314201036
Supervisor : Dr.Kartika Fithriasari, M.Si.
Co-Supervisor : Dr. Brodjol Sutijo Suprih Ulama, M.Si

ABSTRACT

Demam Berdarah or Dengue fever (DBD) is a disease caused by the mosquito *Aedes Aegypti* spreading dengue virus in the human body. This research will be carried out modeling of DBD cases in Surabaya using three models namely Poisson Regression, Negative Binomial Regression and Neural Networks (NN). Modeling of dengue cases has been done, one of them using Poisson regression, but in many cases the assumption equidispersion on Poisson regression difficult to reach. One alternative that can be used to overcome this problem is to modeling the data by using negative binomial regression models. Neural Network is one of the techniques that more flexible in solving the case of variable relation modeling. Observation units used are 31 districts in Surabaya. Number of dengue cases in each district as response variables, independent variable is the percentage of loose home larva, poor households, PHBS, population density, the ratio of medical personnel, RT which have a healthy litter, as well as means Puskesmas. Data was divided into training data and testing using the k-fold crossvalidation. Architecture used in the model was a neural network Multi Layer Perceptron with backpropagation learning algorithm and Levenberg Marquardt method of weight change. The initial weight was initialized with the most optimum value after randomly selecting 100 times. Overdispersion cases occurred in Poisson regression model was produced, which was followed by using a negative binomial regression modeling, and obtained results that overdispersion cases can be resolved, but the case can't be resolved under estimate. In the model (NN), best architecture resulting in MLP (8-12-1). Furthermore, the election of the best model, of the three models were produced criteria for selecting the best model was determined based on the value of CV testing, and obtained results that NN is the best model used to predict the number of dengue cases in Surabaya.

Keyword : DBD, *Poisson*, *Neural Network*

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT atas segala rahmat, karunia, rizki, dan hidayahnya yang diberikan kepada seluruh hambanya. Nikmat keimanan dan kesehatan merupakan salah satu nikmat berharga yang penulis rasakan selama proses pengerjaan Tugas Akhir ini, dengan kedua nikmat tersebut penulis akhirnya dapat menyelesaikan penulisan laporan Tesis yang berjudul **“Perbandingan Model *Neural Networks* dengan *Poisson Regression* dan *Negative Binomial Regression* pada Kasus Demam Berdarah Dengue (DBD) di Surabaya”**. Selama proses pengerjaan sampai tersusunnya laporan Tesis ini, penulis banyak dibantu oleh beberapa pihak. Untuk itu penulis ingin menyampaikan apresiasi dan ucapan terima kasih kepada:

1. Kedua orang tua tercinta, bapak dan ibu yang sudah banyak memberikan dukungan baik secara moril dan materil serta doa untuk kelancaran dan kesuksesan penulis.
2. Bapak Dr. Suhartono, MSc selaku Ketua Jurusan dan Ketua Program Studi Pasca Sarjana Statistika-FMIPA ITS
3. Ibu Dr. Kartika Fithriasari, M.Si. dan Bapak Dr. Brodjol Sutijo Suprih Ulama, M.Si selaku dosen pembimbing penulis yang selama ini sudah banyak sekali membantu penulis serta memberikan banyak pengalaman untuk penulis selama proses pengerjaan Tesis.
4. Bapak Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si serta Ibu Irhamah, M.Si., Ph.D. selaku dosen penguji penulis
5. Pemerintah, pimpinan Dikti, dan Direktur Pascasarjana ITS, Bapak Prof. Ir. Djauhar Manfaat, M.Sc, Ph.D melalui beasiswa *fresh graduate* yang telah mendukung penulis untuk menyelesaikan studi Magister di Statistika ITS
6. Adikku Fanina yang selalu memberikan motivasi dan dukungan.
7. Teman seperjuangan S2 Statistika 2014. “Semoga kelak kita dapat bertemu kembali dalam keadaan yang lebih baik daripada sekarang”
8. Pihak-pihak yang sudah banyak membantu penulis dalam proses pengerjaan Tesis ini, yang tidak dapat penulis sebutkan satu per satu.

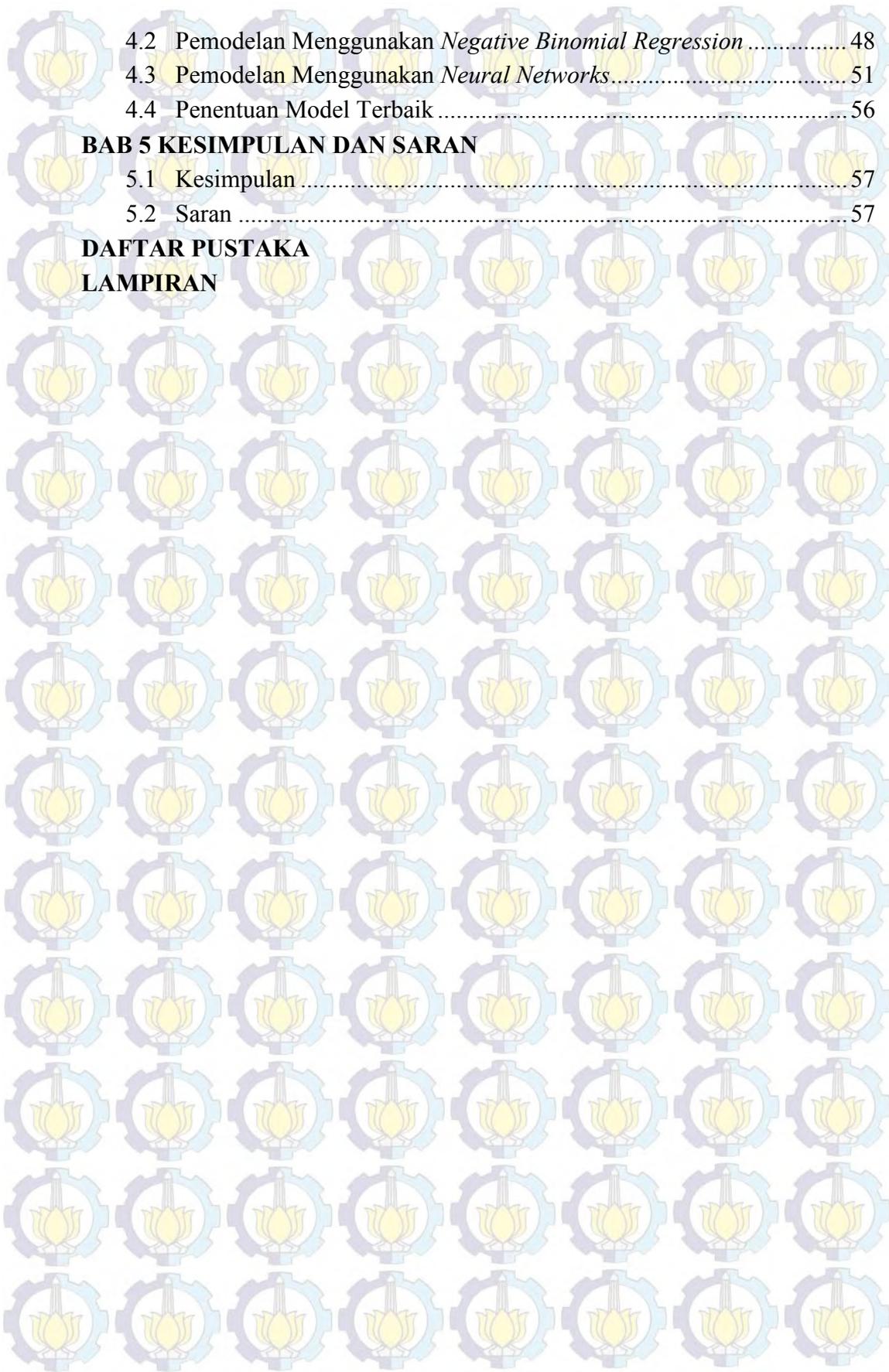
Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan, untuk itu penulis menerima segala macam bentuk saran dan kritik yang diberikan untuk perbaikan laporan Tesis ini. Terakhir, penulis berharap semoga laporan ini dapat memberikan banyak manfaat untuk pembaca.

Surabaya, Februari 2016

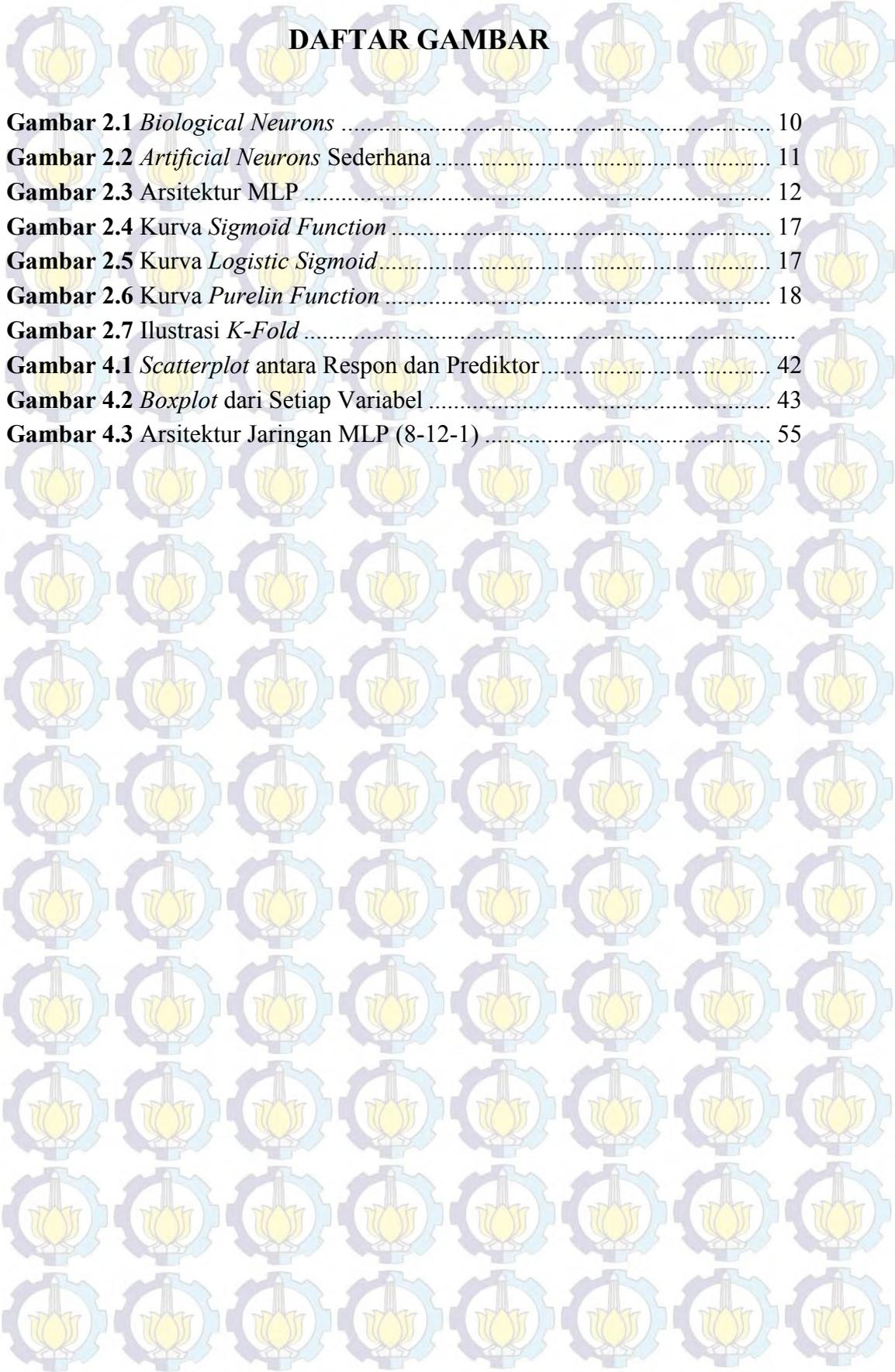
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	ix
BAB 1 PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Penelitian	7
BAB 2 KAJIAN PUSTAKA	
2.1 <i>Neural Networks</i>	9
2.2 <i>Multi Layer Perceptron</i>	11
2.3 <i>Backpropagation</i>	13
2.4 Fungsi Aktivasi.....	16
2.5 <i>Linear Model</i>	18
2.6 <i>General Linear Model</i>	19
2.7 <i>Poisson Regression</i>	20
2.8 Penaksiran Parameter Model <i>Poisson Regression</i>	22
2.9 Pengujian Parameter Model <i>Poisson Regression</i>	24
2.10 Multikolinieritas	25
2.11 <i>Overdispersi</i>	25
2.12 <i>Negative Binomial Regression</i>	26
2.13 Penaksiran Parameter Model <i>Negative Binomial Regression</i>	27
2.14 Pengujian Parameter Model <i>Negative Binomial Regression</i>	
2.15 <i>K-Fold Crossvalidation</i>	30
2.16 Demam Berdarah Dengue	32
BAB 3 METODOLOGI PENELITIAN	
3.1 Bahan dan Alat	35
3.2 Metode Penelitian.....	37
BAB 4 ANALISIS DAN PEMBAHASAN	
4.1 Pemodelan Menggunakan <i>Poisson Regression</i>	44



4.2	Pemodelan Menggunakan <i>Negative Binomial Regression</i>	48
4.3	Pemodelan Menggunakan <i>Neural Networks</i>	51
4.4	Penentuan Model Terbaik.....	56
BAB 5 KESIMPULAN DAN SARAN		
5.1	Kesimpulan.....	57
5.2	Saran.....	57
DAFTAR PUSTAKA		
LAMPIRAN		

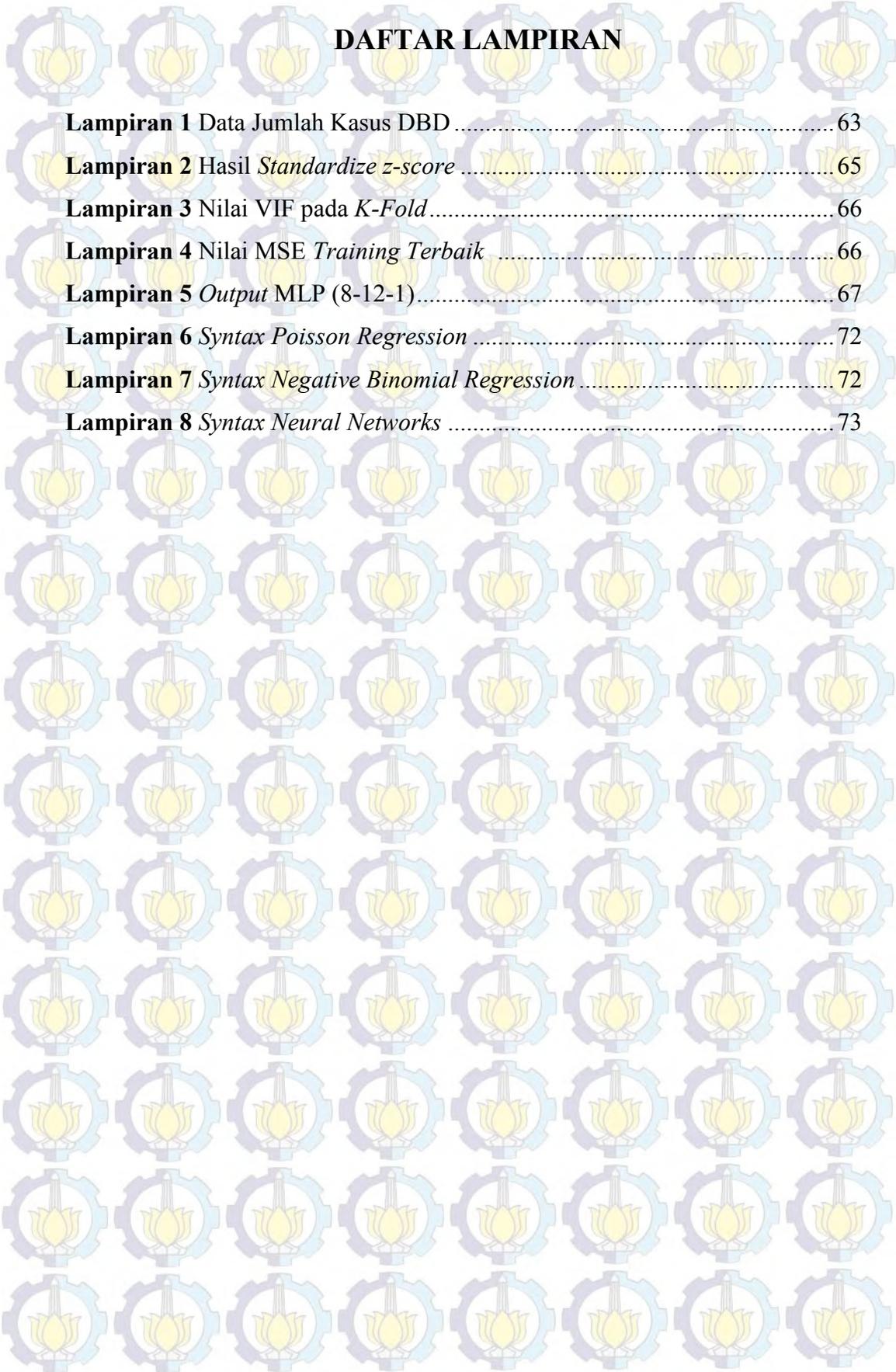


DAFTAR GAMBAR

Gambar 2.1 <i>Biological Neurons</i>	10
Gambar 2.2 <i>Artificial Neurons Sederhana</i>	11
Gambar 2.3 <i>Arsitektur MLP</i>	12
Gambar 2.4 <i>Kurva Sigmoid Function</i>	17
Gambar 2.5 <i>Kurva Logistic Sigmoid</i>	17
Gambar 2.6 <i>Kurva Purelin Function</i>	18
Gambar 2.7 <i>Ilustrasi K-Fold</i>	
Gambar 4.1 <i>Scatterplot</i> antara Respon dan Prediktor	42
Gambar 4.2 <i>Boxplot</i> dari Setiap Variabel	43
Gambar 4.3 <i>Arsitektur Jaringan MLP (8-12-1)</i>	55

DAFTAR TABEL

Tabel 3.1 Variabel Penelitian	35
Tabel 4.1 Hasil Korelasi Kasus DBD dengan Faktor	41
Tabel 4.2 Nilai VIF pada Setiap Variabel	43
Tabel 4.3 Hasil Estimasi <i>Poisson</i> (Model Umum)	44
Tabel 4.4 Hasil Estimasi <i>Poisson</i> (Semua Parameter)	45
Tabel 4.5 Hasil Estimasi <i>Poisson</i> (7 Parameter)	46
Tabel 4.6 Hasil Estimasi <i>Negative Binomial</i> (Model Umum)	48
Tabel 4.7 Hasil Estimasi <i>Negative Binomial</i> (Semua Parameter)	49
Tabel 4.8 Hasil Estimasi <i>Negative Binomial</i> (Lima Parameter)	50
Tabel 4.9 Bobot Awal MLP (8-12-1)	52
Tabel 4.10 Bobot Akhir MLP (8-12-1)	54



DAFTAR LAMPIRAN

Lampiran 1 Data Jumlah Kasus DBD	63
Lampiran 2 Hasil <i>Standardize z-score</i>	65
Lampiran 3 Nilai VIF pada <i>K-Fold</i>	66
Lampiran 4 Nilai MSE <i>Training Terbaik</i>	66
Lampiran 5 <i>Output MLP (8-12-1)</i>	67
Lampiran 6 <i>Syntax Poisson Regression</i>	72
Lampiran 7 <i>Syntax Negative Binomial Regression</i>	72
Lampiran 8 <i>Syntax Neural Networks</i>	73

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Demam berdarah atau demam dengue (disingkat DBD) adalah infeksi yang disebabkan oleh virus *dengue* dari genus *Flavivirus*, famili *Flaviviridae*. (Dinkes Surabaya, 2015). Virus *dengue* ditularkan atau disebarkan pada tubuh manusia melalui gigitan nyamuk *Aedes Aegypti*. Penularan sering terjadi pada musim hujan. DBD banyak ditemukan di daerah tropis dan sub-tropis. Sejak tahun 1968 hingga tahun 2009, World Health Organization (WHO) mencatat negara Indonesia sebagai negara dengan kasus DBD tertinggi di Asia Tenggara. Banyaknya penderita dan luas daerah penyebarannya semakin bertambah seiring dengan meningkatnya mobilitas dan kepadatan penduduk. Penyakit ini tidak hanya sering menimbulkan Kejadian Luar Biasa (KLB) tetapi juga menimbulkan dampak buruk sosial dan ekonomi. Kerugian sosial yang terjadi diantaranya kepanikan keluarga, kematian anggota keluarga, dan berkurangnya usia harapan penduduk (DINKES Jatim, 2012).

Di Indonesia Demam Berdarah pertama kali ditemukan di kota Surabaya pada tahun 1968, dimana sebanyak 58 orang terinfeksi dan 24 orang diantaranya meninggal dunia (Angka Kematian (AK) : 41,3 %) (Kementrian Kesehatan RI, 2010). Berdasarkan Data Profil Kesehatan Jawa Timur Tahun 2012 menunjukkan bahwa Surabaya merupakan kota dengan persentase banyaknya kasus DBD tertinggi, yang diikuti dengan Kab Bojonegoro, Jombang, Kediri, serta Bangkalan. Nilai kepadatan penduduk yang tinggi serta mobilitas yang padat juga turut berkontribusi terhadap tingginya persentase banyaknya kasus DBD di Surabaya. Selama beberapa tahun terakhir, kasus DBD di Surabaya selalu ditemukan dan jumlahnya fluktuatif. Dalam lima tahun terakhir, kasus DBD tertinggi terjadi pada tahun 2010 dengan 3.379 kasus, lalu di 2011 turun jauh

menjadi 1.008 kasus. Pada tahun 2012, kasus DBD naik menjadi 1091 dan kembali naik menjadi 2.207 kasus di tahun 2013 (Hakim, 2015).

Tingginya kasus DBD di Surabaya tentu saja tidak terlepas dari beberapa faktor yang mempengaruhinya. Faktor sosial ekonomi dan beberapa faktor lainnya seperti persentase rumah/bangunan bebas jentik nyamuk aedes, persentase rumah sehat, persentase rumah tangga yang berperilaku hidup bersih dan sehat, persentase rumah sehat, persentase rumah tangga yang memiliki tempat sampah sehat, rasio dokter umum, serta rasio banyaknya puskesmas diduga juga turut berpengaruh. Mengingat kota Surabaya mempunyai kasus DBD yang tinggi dibandingkan dengan Kabupaten/Kota di sekitarnya, maka untuk mengetahui faktor-faktor yang signifikan terhadap DBD perlu dibuat suatu pemodelan, sehingga diharapkan upaya untuk penanganan kasus DBD di Surabaya dapat lebih efisien dan terfokus.

Penelitian mengenai pemodelan DBD telah banyak dilakukan sebelumnya. Yussanti (2012) melakukan pemodelan DBD di Jawa Timur berdasarkan faktor iklim dan sosio ekonomi dengan pendekatan regresi panel semiparametrik dan diperoleh hasil bahwa terdapat dua variabel sosio-ekonomi yang signifikan yaitu pendapatan perkapita yang memberikan efek positif terhadap banyaknya kejadian DBD dan jumlah penduduk berusia 15 tahun kebawah yang memberikan efek negatif. Variabel curah hujan memberikan efek positif terhadap kejadian DBD bila intensitas curah hujan berada antar 1500 mm hingga 3670 mm. Suhu udara memberikan efek yang besar terhadap perubahan kejadian DBD pada suhu antara 22⁰C hingga 27⁰C, sedangkan kelembaban memberikan efek positif terhadap perubahan kejadian DBD ketika kelembaban berada pada interval 82% hingga 87%. Earnest *et al.* (2011) melakukan penelitian dengan membandingkan kedua model statistik untuk memprediksi peringatan DBD. Kedua model yang dipakai sebagai perbandingan adalah model *Autoregressive Integrated Moving Average* (ARIMA) dan Model *Knorr-Held* dua komponen. Kriteria pemilihan model terbaik didasarkan pada nilai MAPE terkecil yang dihasilkan dari kedua model. Hasil yang diperoleh menunjukkan bahwa nilai MAPE yang lebih rendah diperoleh dari model *Knorr-Held*, namun *Knorr-Held* relatif lebih sulit untuk mem-*fitting* model serta membutuhkan waktu yang relatif lebih lama untuk menjalankan modelnya.

Pemodelan DBD di Surabaya dengan menggunakan tiga buah metode juga pernah dilakukan oleh Hidayanti (2014). Ketiga metode tersebut adalah *negative binomial regression*, *poisson regression*, dan *Geographically Weighted Negative Binomial Regression* (GWNBR). Dari ketiga metode tersebut dilakukan perbandingan dan diperoleh hasil bahwa metode GWNBR menghasilkan model terbaik untuk pemodelan kasus DBD di Surabaya. Untuk mendapatkan gambaran tentang daerah yang memerlukan perhatian khusus atau prioritas utama dalam penanggulangan DBD, Hidayanti menggunakan menggunakan metode *Flexibly Shaped Spatial Scan Statistic* dimana hasil yang diperoleh menunjukkan kecamatan Benowo merupakan daerah yang memiliki resiko tertinggi.

Penelitian mengenai penerapan analisis spasial untuk pemeriksaan DBD di Guayaquil, Ecuador pernah dilakukan oleh Castillo *et al* (2011). Hasilnya menunjukkan bahwa autokorelasi spasial antar lokasi sangat tinggi. Pada tahun 2005-2007 dapat diketahui bahwa hotspot dari kejadian demam berdarah terletak di lingkungan berpenghasilan rendah dengan akses terbatas untuk utilitas publik. Khormi dan Kumar (2011) pernah melakukan penelitian yang berjudul Pemodelan DBD berdasarkan Faktor Sosial Ekonomi, Kewarganegaraan dan Kelompok Usia (Studi Kasus: GIS dan Pengindraan Jauh). Data yang dipakai adalah data kualitas lingkungan hidup dari gambar satelit beresolusi tinggi. Hal tersebut membuat Khormi dan Kumar memutuskan untuk menganalisis beberapa faktor sosial ekonomi seperti jumlah populasi, kepadatan populasi, dan kualitas lingkungan menggunakan GWR untuk membuat model prediksi, identifikasi tingkat resiko demam berdarah, dan untuk menentukan hubungan antara kasus DBD dengan faktor sosial ekonomi. Hasil dari penelitiannya menunjukkan bahwa terdapat hubungan positif yang kuat antara kasus demam berdarah dan-faktor sosial ekonomi.

Analisis regresi merupakan salah satu metode statistik yang digunakan untuk menyelidiki pola hubungan antara variabel prediktor dengan variabel respon. Terdapat dua jenis variabel dalam analisis regresi yakni variabel prediktor yang biasa dilambangkan dengan X dan variabel respon yang dilambangkan dengan Y dimana kedua variabel tersebut saling berkorelasi. Tipe data pada variabel Y dapat beragam, salah satunya data *count*. Poisson regression merupakan salah satu

metode analisis yang dipakai untuk memodelkan suatu data saat variabel Y bertipe data *count*. Beberapa penelitian dengan menggunakan poisson regression pernah dilakukan oleh Yulianingsih, Sukarsa, Sucipwati (2012) yang mengaplikasikannya pada kasus banyaknya siswa SMA / SMK yang tidak lulus UN di Bali. Berikutnya dilakukan oleh Sundari (2014) mengenai aplikasi *Poisson regression* untuk Memodelkan Hubungan Usia dan Perilaku Merokok terhadap banyaknya Kematian Penderita Penyakit Kanker Paru-Paru,

Asumsi penting pada analisis poisson regression adalah nilai ragam harus sama dengan nilai rata-ratanya yang biasa disebut dengan kondisi *equidispersion* (Famoye et al, 2004). Pada banyak kasus asumsi tersebut sangat sulit dipenuhi, kasus *overdispersi* seringkali terjadi. Overdispersi terjadi saat nilai varians data lebih besar daripada nilai meannya. *Binomial Negative Regression* merupakan salah satu metode alternatif yang sering dipakai untuk mengatasi kasus tersebut. Hal tersebut pernah dilakukan oleh Fatmasari (2014), yang melakukan pendekatan *Binomial Negative Regression* untuk mengatasi overdispersi pada *Poisson Regression* dengan studi kasus DBD di Kota Malang. Hasil analisis menunjukkan bahwa variabel yang signifikan mempengaruhi kasus DBD di kota Malang adalah kepadatan penduduk, sarana kesehatan, Gerakan Kemiskinan, PHBS, dan TPS. Nandasari (2014) juga menggunakan metode *negative binomial regression* untuk mengatasi kasus overdispersi pada pemodelan banyaknya kejadian luar biasa difteri di Jawa Timur. Berdasarkan nilai AIC dan devians yang dihasilkan dapat diketahui bahwa *negative binomial regression* lebih baik daripada model poisson regression.

Salah satu teknik pemodelan yang sedang berkembang saat ini adalah *Neural networks*, yaitu sebuah sistem pengolahan informasi yang secara umum memiliki kemampuan dan karakteristik menyerupai jaringan saraf biologis manusia (Fausett, 1994). *Neural network* telah berkembang secara pesat, sejak ditemukan pada tahun 1940-an, seiring dengan kecanggihan teknologi komputer saat ini. Elemen-elemen pemrosesan yang saling terhubung dengan sistem yang paralel menyebabkan pemrosesan pada *neural network* tidak membutuhkan waktu yang lama. Nilai *output* dari *neural network* diperoleh berdasarkan proses pembelajaran yang telah dilakukan sebelumnya.

Tidak seperti *poisson regression* yang mensyaratkan asumsi *equi-dispersion*, *neural networks* bersifat lebih fleksibel dalam penyelesaian kasus pemodelan hubungan variabel. Hal tersebut membuat penggunaan *neural networks* telah banyak dilakukan sebelumnya dalam penyelesaian berbagai kasus. Duddu & Pulugurtha (2012) yang melakukan estimasi tabrakan pada zona level untuk perencanaan transportasi. Estimasi dilakukan untuk total banyaknya dari kecelakaan, kerugian kecelakaan, banyaknya kerusakan properti dengan menggunakan metode *negative binomial regression* dan *neural networks*. Metode *negative binomial regression* dipilih karena kasus overdispersi terjadi pada ketiga variabel dependen (parameter dispersi > 0). Pada penelitian Duddu & Pulugharta, arsitektur *neural networks* yang dipakai menggunakan *multi layer perceptron* dengan algoritma pembelajaran *backpropagation*. Selanjutnya kedua metode tersebut dibandingkan dengan menggunakan kriteria *Mean Absolute Deviation* (MAD), persentil eror ke-50, persentil eror ke 85, *Root Mean Square Error* (RMSE). Berdasarkan hasil kriteria kebaikan model dapat diketahui bahwa model *neural networks* memiliki kemampuan yang lebih baik dalam memprediksi total banyaknya dari kecelakaan, kerugian kecelakaan, dan banyaknya kerusakan properti jika dibandingkan dengan model *negative binomial regression*.

Aydin & Tiryaki (2014) melakukan pemodelan *neural networks* untuk memprediksi kekuatan kompresi dari perlakuan panas kayu dan membandingkannya dengan regresi linear berganda, dan diperoleh hasil bahwa *neural networks* memberikan hasil prediksi yang lebih baik dengan nilai R^2 yang lebih besar dan MAPE yang lebih kecil daripada regresi linier berganda. Merujuk dari hasil penelitian Hidayanti (2014) bahwa faktor lokasi tidak memberikan pengaruh signifikan pada banyaknya kasus DBD di Surabaya serta penelitian Duddu & Pulugartha (2012) yang melakukan perbandingan *neural networks* dengan metode statistika konvensional, hal tersebut melatarbelakangi peneliti untuk melakukan penelitian mengenai Perbandingan Model *Neural Networks*, *Poisson Regression*, dan *Negative Binomial Regression* pada Kasus Demam Berdarah Dengue (DBD) di Surabaya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana model *poisson regression* yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya?
2. Bagaimana model *negative binomial regression* yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya?
3. Bagaimana model *neural networks* yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya?
4. Bagaimana model terbaik yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka penelitian ini memiliki tujuan sebagai berikut.

1. Menentukan model *poisson regression* yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya.
2. Menentukan model *negative binomial regression* yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya.
3. Menentukan model *neural networks* yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya.
4. Menentukan model terbaik yang dipakai untuk memprediksi banyaknya kasus DBD di Kota Surabaya.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

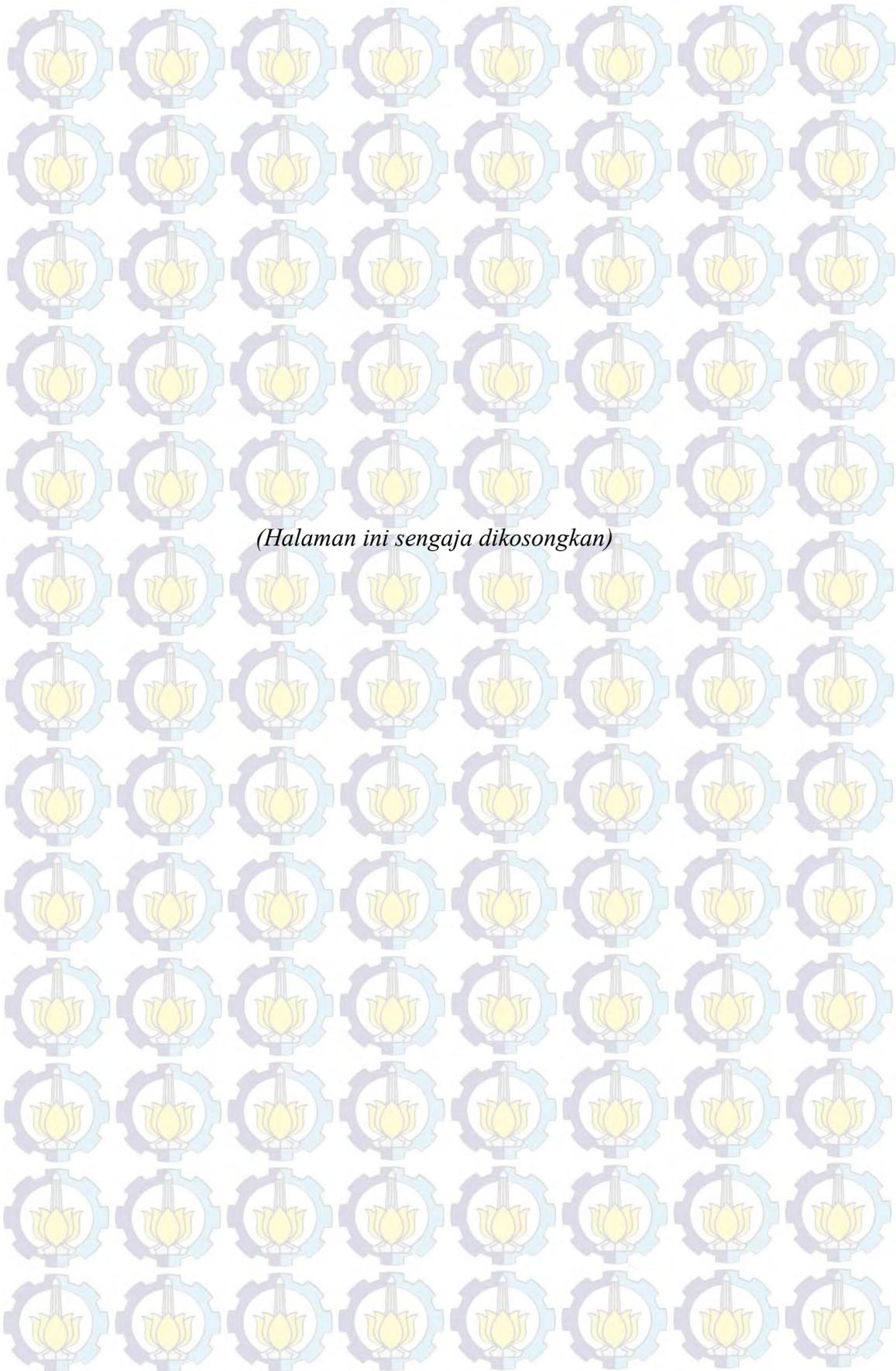
Manfaat untuk pemerintah: Memberikan alternatif model yang dapat dipakai untuk memprediksi banyaknya kasus penyakit DBD di Surabaya, sehingga diharapkan dapat menjadi tambahan informasi serta bahan rujukan untuk pengambilan kebijakan pada tiap kecamatan di Surabaya dalam program pencegahan dan penanggulangan DBD di masa depan.

Manfaat untuk peneliti : Mengkaji dan menerapkan aplikasi dari ilmu statistik khususnya pemodelan dengan *poisson regression*, *negative binomial regression* serta *neural networks*, sehingga dapat diaplikasikan dan diterapkan untuk penyelesaian banyaknya kasus DBD di Kota Surabaya.

1.5 Batasan Penelitian

Pada penelitian ini, ruang lingkup permasalahan dibatasi dengan beberapa hal sebagai berikut.

1. Data yang dipakai pada penelitian ini adalah data banyaknya kasus DBD di kota Surabaya pada tahun 2013.
2. Arsitektur pemodelan yang dipakai menggunakan *Multi Layer Perceptron* dengan algoritma pembelajaran *backpropagation*. Fungsi aktivasi yang dipakai adalah *logistic sigmoid* pada *hidden layer* dan *purelin function* pada *output layer*



BAB 2

KAJIAN PUSTAKA

Pada Bab ini akan diberikan penjelasan teori mengenai *neural networks*, *poisson regression*, dan *negative binomial regression* serta beberapa teori yang berkaitan dengan penelitian ini.

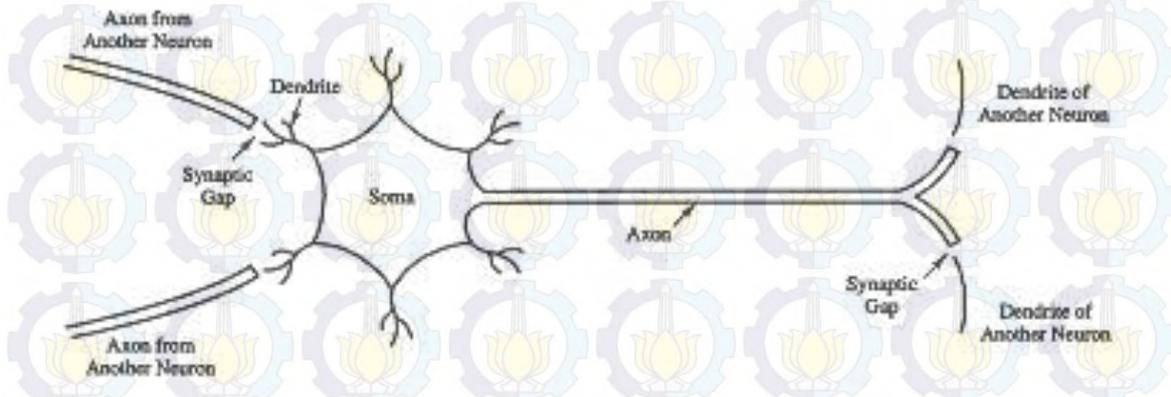
2.1 *Neural networks*

Artificial Neural networks atau *neural networks* adalah sistem pemrosesan informasi yang memiliki karakteristik kemampuan menyerupai jaringan biologis manusia. Terdapat 3 jenis komponen dari jaringan biologis yang mirip dengan komponen *neural networks* yaitu dendrit, soma, dan akson. Beberapa dendrit menerima sinyal dari neuron lainnya. Sinyal adalah impuls listrik yang dipancarkan di celah sipnasis saat proses kimia. Aktivitas dari pemancar kimia yang mengubah sinyal masuk memiliki cara yang mirip dengan aktivitas bobot pada *neural networks*. Soma atau sel tubuh menghitung sinyal yang masuk. Ketika input diterima, sel tubuh akan melepas atau memberikan rangsangan, rangsangan tersebut mengirimkan sinyal melalui akson pada sel tubuh yang lain (Fausett, 1994). Bentuk dari jaringan biologis manusia ditampilkan pada Gambar 2.1.

Neural networks dikembangkan sebagai generalisasi model matematika pada jaringan syaraf biologis otak manusia, dengan beberapa asumsi sebagai berikut (Fausett, 1994).

- a. Pemrosesan informasi terjadi pada beberapa elemen sederhana yang disebut neuron
- b. Sinyal diterima atau dilewatkan antar neuron melalui penghubung (*connection-link*)
- c. Setiap penghubung berhubungan dengan bobot / *weight* masing-masing dan merupakan bobot dari sinyal yang melalui penghubung tersebut

- d. Setiap neuron menggunakan fungsi aktivasi pada *net input*-nya (*sum of weight input signal*) untuk menentukan sinyal *output*



Gambar 2.1 *Biological Neurons* (Fausett,1994)

Neural networks memiliki beberapa karakteristik sebagai berikut.

1. Pola koneksinya antar *neuron* yang disebut dengan arsitektur
2. Metode untuk menentukan bobot pada koneksi (disebut dengan *training, learning*, atau algoritma)
3. Fungsi aktivasi

Neural networks terdiri dari sejumlah besar dari elemen pemrosesan sederhana yang dinamakan *neuron, units, cells*, atau *nodes*. Setiap *neuron* berhubungan dengan *neuron* lainnya melalui suatu penghubung langsung, yang masing-masing memiliki suatu *weights*. *Weights* menggambarkan informasi yang digunakan oleh jaringan untuk menyelesaikan suatu permasalahan.

Setiap *neuron* memiliki sebuah *internal state*, yang dinamakan dengan *activation* atau *activity level*, yang merupakan sebuah fungsi dari input yang diterima. Secara khusus, sebuah *neuron* mengirim aktivasinya sebagai sinyal untuk beberapa *neuron* lainnya.

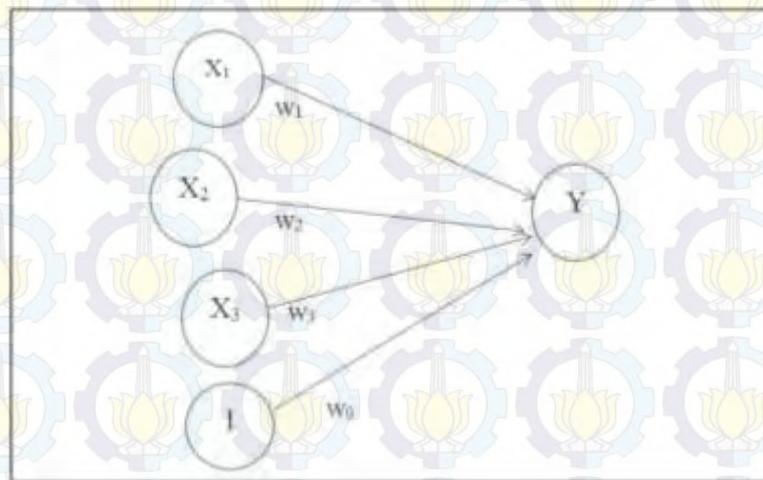
Sebagai contoh, anggap sebuah *neuron* yang diilustrasikan pada Gambar 2.2, yang menerima *inputs* dari *neuron* X_1, X_2, X_3 . Aktivasi sinyal dari *neuron*

tersebut secara berturut-turut adalah x_1 , x_2 , dan x_3 . Bobot yang dihubungkan dari X_1 , X_2 , X_3 pada *neuron Y* adalah w_1 , w_2 , dan w_3 . Input ke *neuron Y* dinyatakan dengan *net input*, y_{in} adalah jumlahan bobot sinyal dari *neuron X*, X_1 , X_2 , X_3 yang ditampilkan pada Persamaan (2.1)

$$y_{in} = w_1x_1 + w_2x_2 + w_3x_3 \quad (2.1)$$

Aktivasi sinyal y dari *neuron Y* diperoleh dari fungsi pada *net input*-nya

$$y = f(y_{in}).$$



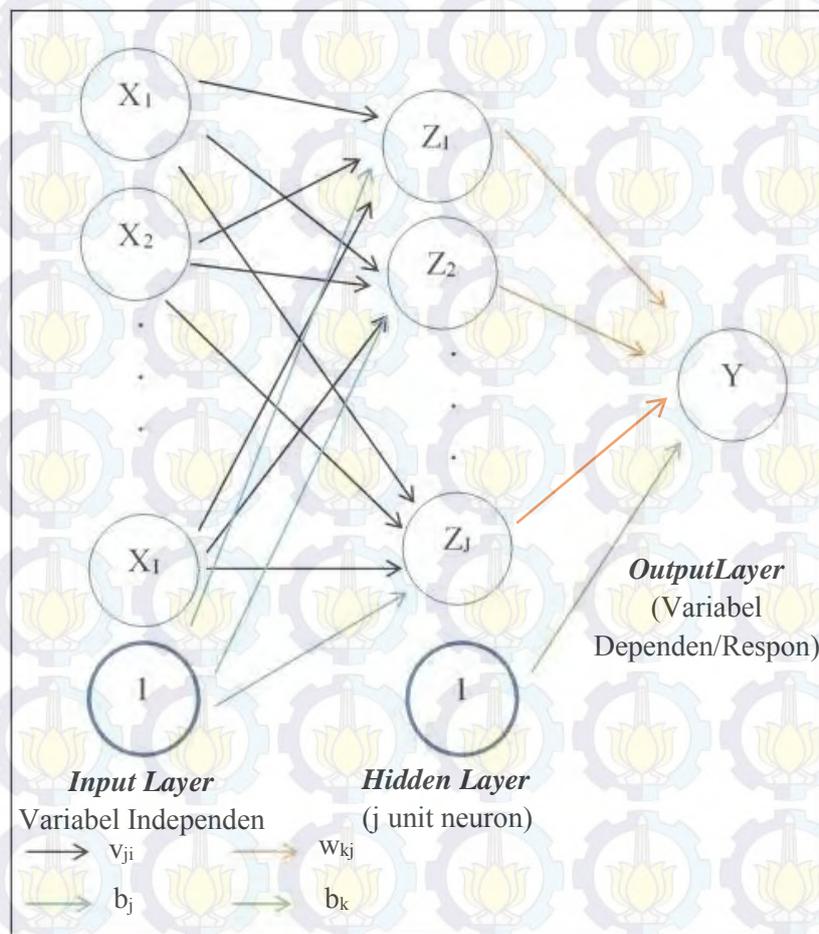
Gambar 2.2 *Artificial Neuron Sederhana* (Fausett, 1994)

2.2 *Multi Layer Perceptron*

Arsitektur dasar dari *neural networks* yang paling sederhana terdiri dari sebuah layer *input* dan sebuah layer *output* yang memiliki satu atau beberapa node (Fausett, 1994). Model *Feed Forward Neural networks* (FFNN) merupakan salah satu bentuk model NN yang sangat fleksibel dari fungsi-fungsi *nonlinear* dan dapat dipakai untuk berbagai aplikasi. *Multi Layer Feed Forward Neural networks* atau *Multi Layer Perceptron* (MLP) merupakan suatu *perceptron* dengan adanya tambahan satu atau lebih *hidden layer*, yang terletak antara jaringan input dengan output. Salah satu fungsi dari *hidden layer* adalah menjadi perantara antara input dengan jaringan *output*.

Menurut Haykin (1999), sumber node pada *input layer* jaringan menyediakan masing-masing elemen dari pola aktivasi (vektor *input*), yang

merupakan sinyal input yang diaplikasikan pada *neuron* pada layer kedua (*hidden layer*). Sinyal *output* pada layer kedua dipakai sebagai input pada layer ketiga, dan seterusnya. Secara khusus *neuron* pada setiap layer dari jaringan memiliki sinyal input dan output dari layer sebelumnya. Kumpulan sinyal *output* dari *neuron* pada *output* terakhir jaringan merupakan keseluruhan respon dari jaringan yang berasal dari sumber node pada layer input pertama. Gambar arsitektural MLP yang telah dijelaskan ditampilkan pada Gambar 2.3.



Gambar 2.3 Arsitektural MLP (Fausett, 1994)

Berdasarkan Gambar 2.3 dapat diketahui terdapat i sumber node, j neuron pada layer *hidden*, dan k neuron pada *output* layer, jaringan seperti Gambar 2.3 disebut dengan arsitektur jaringan i - j - k . Dalam arsitektur MLP dengan satu *hidden layer*, nilai dari *output* atau y_k dapat dihitung melalui Persamaan 2.2

$$y_k = \varphi_k \left(b_k + \sum_{j \rightarrow k} w_{kj} \varphi_j \left(b_j + \sum_{i \rightarrow j} v_{ji} x_i \right) \right) \quad (2.2)$$

Dengan :

y_k = nilai dugaan dari variabel *output*

x_i = variabel input sebanyak I, ($i=1, 2, 3, \dots, I$)

i = indeks untuk *input layer*

j = indeks untuk *hidden layer*

k = indeks untuk *output layer*

φ_j = fungsi aktivasi di neuron ke-j pada *hidden layer*

φ_k = fungsi aktivasi di neuron ke-k pada *output layer*

b_j = bias pada neuron ke j pada *hidden layer* ($j= 1, 2, \dots, J$)

b_k = bias pada neuron ke k pada *output layer* ($k= 1, 2, \dots, K$)

v_{ji} = bobot dari input ke- i yang menuju ke neuron j pada *hidden layer*

w_{kj} = bobot dari neuron ke- j yang menuju ke *output layer*

MLP banyak diterapkan dalam berbagai kasus untuk menyelesaikan beberapa kesulitan dan permasalahan dengan melakukan pembelajaran *supervised* dan menggunakan algoritma *error backpropagation*. Algoritma ini berdasarkan aturan *error corection learning*.

2.3 Backpropagation

Pada tahun 1974, Werbos pertama kali mencetuskan metode pelatihan *backpropagation* atau biasa disebut dengan metode propagasi balik, dan kemudian pada tahun 1986 metode tersebut dikembangkan kembali oleh Rumelhart. Metode ini merupakan metode pelatihan terbimbing (*supervised learning*) yang pada awalnya dirancang untuk model FFNN, yang tersusun dari beberapa layer dan sinyal dialirkan secara searah dari input menuju output, tetapi dalam perkembangannya metode ini diadaptasi untuk pelatihan pada model NN lainnya.

Pada umumnya, pembelajaran dari *backpropagation* terdiri dari 2 tahapan yaitu *forward* dan *backward* (Haykin,1999). Secara singkat tahapan kerja dari metode pelatihan ini adalah input set pelatihan yang diasosiasikan dengan target

pelatihannya, selanjutnya input sel pelatihan tersebut dimasukkan ke dalam lapisan input, dan dilakukan pemrosesan dan diperoleh *output* yang hasilnya masuk pada lapisan selanjutnya (*hidden layer*), berikutnya dilakukan kembali tahapan pemrosesan informasi dan hasilnya masuk pada lapisan berikutnya (*output layer*). Di lapisan *output* juga dilakukan pemrosesan informasi dan keluaran pada jaringan merupakan sebuah *output*. Tahapan selanjutnya dilakukan perbandingan antara *output* dengan target. Perbedaan yang timbul antara *output* dengan target pada jaringan merupakan sebuah error. Selanjutnya error tersebut akan dipakai untuk meng-*update* bobot yang relevan dengan cara mempropagasikan kembali error.

Algoritma pelatihan *backpropagation* menurut Fausett (1994) terdiri dari 3 tahapan yaitu:

1. Input nilai data pelatihan / *training* sehingga diperoleh nilai output (*feed-forward*)
2. Propagasi balik dari nilai *error* yang diperoleh
3. Penyesuaian bobot koneksi untuk meminimalkan nilai *error*

Ketiga tahapan diulangi terus menerus sampai didapatkan nilai *error* yang sesuai dan diperoleh bobot terbaik dari NN, selanjutnya dilakukan pengujian atau *testing* dengan menggunakan data *testing*. Pada tahapan terakhir, hanya tahap pertama dari algoritma diatas yang diperlukan.

Langkah-langkah dalam algoritma *backpropagation* disajikan sebagai berikut.

0. Inisialisasi bobot awal, penentuan nilai η , dan θ , serta set maksimum *epoch*
1. Saat kondisi berhenti tidak terpenuhi, lakukan langkah 2 sampai dengan langkah 9
2. Untuk setiap pasangan pola pelatihan, lakukan langkah 3 sampai dengan 8

Tahap Umpan Maju (*Feed-forward*)

3. Setiap unit input X_i mengirimkan sinyal ke unit lapisan tersembunyi
4. Pada setiap unit di lapisan tersembunyi Z_j , sinyal output lapisan tersembunyi dihitung dengan menerapkan fungsi aktivasi terhadap penjumlahan sinyal-sinyal input berbobot x_i :

$$z_j = \varphi_j \left(b_j + \sum_{i=1}^I x_i v_{ji} \right) \quad (2.3)$$

Kemudian dikirim ke semua unit di lapisan atas

5. Setiap unit di lapisan output Y_k dihitung sinyal outputnya dengan menerapkan fungsi aktivasi terhadap penjumlahan sinyal input berbobot z_j bagi lapisan ini:

$$y_k = \varphi_k(b_k + \sum_{j=1}^J z_j w_{kj}) \quad (2.4)$$

Tahap Propagasi Balik

6. Setiap unit output (y_k) mempunyai pola target (t_k), lalu informasi kesalahan lapisan output (δ_k) dihitung, δ_k dikirim ke lapisan dibawahnya dan digunakan untuk menghitung besar koreksi bobot dan bias (Δw_{kj} & Δb_k) antara lapisan tersembunyi dengan lapisan output.

$$\delta_k = (t_k - y_k) \varphi_k(b_k + \sum_{j=1}^J z_j w_{kj}) \quad (2.5)$$

$$\Delta w_{kj} = \eta \delta_k z_j \quad (2.6)$$

$$\Delta b_k = \alpha \delta_k \quad (2.7)$$

1. Pada tiap unit di lapisan tersembunyi dilakukan perhitungan informasi kesalahan lapisan tersembunyi (δ_j). δ_j kemudian dipakai untuk menghitung besar koreksi bobot dan bias (Δw_{ji} & Δb_j) antara lapisan input dan tersembunyi.

$$\delta_j = (\sum_{k=1}^K \delta_k w_{jk}) \varphi_j(b_j + \sum_{i=1}^I x_i v_{ji}) \quad (2.8)$$

$$\Delta v_{ji} = \eta \delta_j x_i \quad (2.9)$$

$$\Delta b_j = \alpha \delta_j \quad (2.10)$$

Tahap Update Bobot dan Bias

8. Pada setiap unit output y_k dilakukan pengupdaten bias dan bobot ($j=1, \dots, J$; $k=1, \dots, K$) sehingga bias dan bobot baru menjadi

$$w_{kj}(\text{baru}) = w_{kj}(\text{lama}) + \Delta w_{kj} \quad (2.11)$$

Dari unit ke -1 sampai dengan p di lapisan tersembunyi juga dilakukan pengupdaten bobot dan bias

$$b_j(\text{baru}) = b_j(\text{lama}) + \Delta b_j \quad (2.12)$$

9. Tes kondisi berhenti.

Jika kondisi berhenti telah terpenuhi, yaitu nilai ambang (θ) telah tercapai atau pelatihan sudah mencapai iterasi maksimum, maka pelatihan jaringan dapat dihentikan.

Pada penelitian ini, metode yang dipakai untuk melakukan perubahan bobot adalah algoritma *Levenberg-Marquardt*, yaitu salah satu metode optimasi yang digunakan untuk meminimumkan jumlah kuadrat error (*sum square error*) (Bishop, 1995). Bobot baru yang dihasilkan dapat dihitung berdasarkan Persamaan (2.13).

$$w_{baru} = w_{lama} - (Z'Z + \lambda I)^{-1} Z' \epsilon \quad (2.13)$$

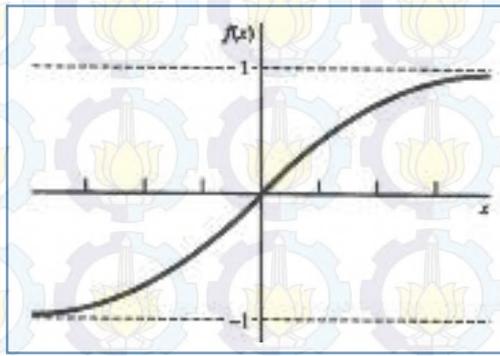
I adalah matriks identitas, λ adalah koefisien kombinasi yang nilainya selalu positif, dan Z merupakan matriks jacobian, yaitu turunan pertama dari ϵ , sedangkan matriks Hessian didefinisikan sebagai $H = Z'Z$.

Berikutnya dilanjutkan dengan menghitung keluaran jaringan dengan algoritma berikut.

1. Setelah diperoleh bobot akhir optimum dari algoritma *backpropagation*, bobot tersebut dipakai untuk menghitung keluaran jaringan.
2. Keluaran jaringan dihitung berdasarkan arsitektur terbaik yang telah diperoleh pada proses *learning*, pola masukan, dan fungsi aktivasi yang dipakai.

2.4 Fungsi Aktivasi

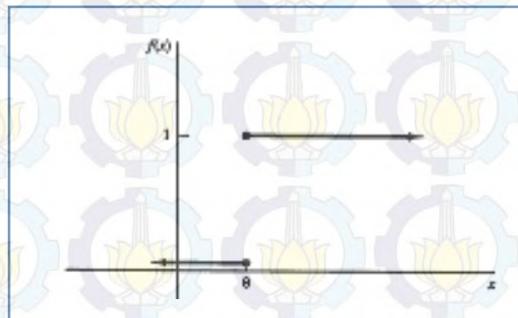
Pada *neural networks*, setiap neuron memiliki fungsi aktivasi yang dipakai untuk menentukan suatu *output* dari sebuah neuron yang telah diberikan suatu *input*. Penggunaan fungsi aktivasi tergantung pada kebutuhan dari *output*. Terdapat beberapa fungsi aktivasi baik linear maupun *non linear*, namun untuk beberapa kasus, fungsi aktivasi *non linear* lebih banyak digunakan. Dalam algoritma pembelajaran *backpropagation*, suatu fungsi aktivasi yang dipakai seharusnya memiliki beberapa karakteristik tertentu seperti kontinyu, *differentiable*, dan diharapkan memiliki *derivative* yang mudah untuk dihitung sehingga akan mengurangi beban perhitungan selama pelatihan (Fausett, 1995).



Gambar 2.4 Kurva *sigmoid function* (Fausett, 1994)

Fungsi aktivasi yang banyak dipakai dalam pembelajaran *backpropagation* karena memiliki beberapa karakteristik yang diharapkan adalah fungsi aktivasi non linear, yaitu *sigmoid*. *Logistic function* dan *hyperbolic tangen function* merupakan yang sering dipakai. Bentuk kurva dari fungsi sigmoid berbentuk seperti huruf s, disajikan pada Gambar 2.4.

Pada penelitian ini, fungsi aktivasi yang dipakai pada *hidden layer* adalah fungsi aktivasi *logistic sigmoid*. Bentuk kurvanya ditampilkan pada Gambar 2.5.



Gambar 2.5 Kurva *Logistic sigmoid* (Fausett, 1994)

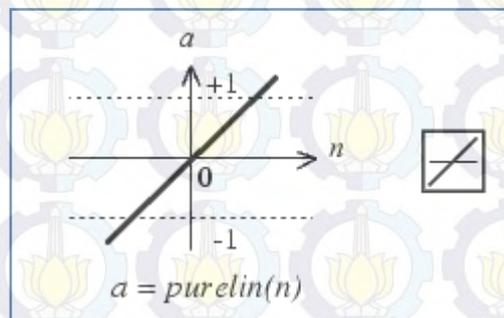
Fungsi aktivasi *logistic sigmoid* merupakan salah satu fungsi aktivasi yang sering dipakai saat output yang diharapkan antara 0 dan 1. Bentuk persamaan dari *logistic sigmoid* ditampilkan pada Persamaan (2.14).

$$\varphi_j(v) = \frac{1}{1 + \exp(-av)} \quad (2.14)$$

Dengan : $v_j = b_j + \sum_{i=1}^I x_i v_{ji}$.

a = slope dari fungsi sigmoid

Pada *output layer*, fungsi aktivasi yang dipakai adalah *purelin*. Pada Gambar 2.6 menunjukkan fungsi aktivasi yang berbentuk linier.



Gambar 2.6 Kurva *Purelin Function* (Fausett, 1994)

. Fungsi linier ini akan membawa *input* ke *output* yang sebanding, dalam artian *output* sebanding dengan jumlah bobot *output*. Persamaan (2.1) menampilkan bentuk persamaan dari fungsi linier.

$$\varphi(v_k) = v_k \quad (2.15)$$

2.5 *Linear Model*

Salah satu bentuk dari model linear adalah model regresi linear sederhana. Model ini menampilkan hubungan antara dua variabel. Contohnya tinggi dan berat badan seseorang, lebar dan tinggi sebuah amplop, temperature dan hasil dari proses industry, dll. Untuk sebuah hubungan yang linier, bentuk modelnya sebagai berikut.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.16)$$

dimana y adalah variabel *dependent* atau respon dan x adalah variabel *independent* atau prediktor. Variabel random ε adalah kondisi error pada model. Error tidak diartikan sebuah kesalahan, tetapi pada statistika menggambarkan adanya efek dari faktor diluar kontrol.

2.6 *Generalized Linear Model (GLM)*

GLM merupakan model yang terdiri dari model regresi klasik dan model *ANOVA* dengan respon yang kontinu maupun diskrit. *GLM* berhubungan dengan variabel prediktor melalui persamaan dengan bentuk linear. *GLM* terdiri dari 3 komponen yaitu *random component*, *systematic component*, dan *link function* (Agresti, 2007).

a. *Random Component*

Random Component dari *GLM* merupakan sebuah variabel respon (Y) yang memiliki distribusi probabilitas. Distribusi dari Y_i adalah anggota dari keluarga eksponensial seperti distribusi Gaussian (normal), binomial, poisson, gamma, atau *inverse gaussian*. Pada beberapa contoh kasus, pengamatan Y berbentuk biner, seperti “sukses” atau “gagal”. Secara umum, Y_i adalah banyaknya dari pengamatan sukses dari suatu percobaan tertentu, yang diasumsikan berdistribusi binomial. Untuk beberapa kasus lainnya, nilai Y dapat berupa sebuah *count*, yang selanjutnya diasumsikan bahwa distribusi untuk Y dapat diaplikasikan untuk semua *non-negative integers*, seperti *poisson* atau *negative binomial*. Jika setiap pengamatan memiliki variabel Y yang kontinu, dapat diasumsikan bahwa distribusinya adalah normal.

b. *Systematic Component*

Systematic Component dari *GLM* merupakan sebuah variabel prediktor, yang merupakan fungsi linear dalam persamaan model pada sisi sebelah kanan. *Systematic Component* merupakan variabel $\{x_j\}$ dengan formula yang tersaji pada Persamaan (2.17). Kombinasi linear dari variabel prediktor dinamakan *linear predictor*.

$$\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.17)$$

Salah satu keuntungan dari *GLMs* adalah struktur dari *linear predictor* memiliki struktur yang sama dengan *linear model*.

c. **Link Function**

Link function menunjukkan ekspektasi dari Y, rata-rata dari distribusi probabilitasnya dengan $\mu = E(Y)$. Komponen ketiga dari GLM adalah *link function* yaitu sebuah fungsi dari $g(\cdot)$ yang mengubah μ ke *linear predictor* seperti yang ditunjukkan pada persamaan berikut:

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.18)$$

Karena *link function* bersifat *invertible*, maka dapat juga ditulis sebagai berikut:

$$\mu = g^{-1}(\eta) = g^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (2.19)$$

GLM dapat dianggap sebagai sebuah transformasi dari *linear model* pada ekspektasi Y atau sebagai model regresi non linear untuk respon. *Inverse link* $g^{-1}(\cdot)$ juga disebut sebagai *mean function*. *Identity link* merupakan salah satu *link function* yang paling sederhana yang ditunjukkan melalui persamaan

$$\eta = g(\mu) = \mu \text{ dan } \mu = g^{-1}(\eta) = \eta \quad (2.20)$$

Link function yang lain yaitu $g(\mu) = \ln(\mu)$ merupakan model log dari *mean*. Fungsi log diaplikasikan untuk hasil yang positif, sehingga *link function* sesuai ketika μ merupakan suatu hasil yang tidak boleh bernilai negatif, seperti data *count* (Agresti, 2007). Suatu GLM yang menggunakan *log link* disebut dengan *loglinear model*. Persamaan (2.21) menunjukkan suatu *loglinear model*.

$$\ln(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.21)$$

2.7 **Poisson Regression**

Beberapa variabel respon diskrit memiliki *counts* sebagai peluang hasil yang mungkin keluar dari suatu percobaan. Misal Y adalah banyaknya peserta pada bulan Maret. *Counts* juga terdapat dalam ringkasan variabel kategorik pada tabel kontingensi. GLM untuk *count data* diasumsikan berdistribusi *poisson* pada *random component*.

Suatu variabel random Y dikatakan memiliki distribusi *poisson* dengan parameter μ jika nilai dari variabel Y berupa integer ($y = 0, 1, 2, \dots$) dengan nilai peluangnya dihitung menggunakan persamaan

$$\Pr\{Y = y\} = \frac{e^{-\mu} \mu^y}{y!} \quad (2.22)$$

Dengan μ adalah rata-rata banyaknya kejadian yang berdistribusi poisson, dengan $\mu > 0$. Parameter μ bergantung pada unit tertentu atau periode waktu tertentu, jarak, volume, luas, dsb (Rodriguez, 2007). *Poisson distribution* digunakan untuk memodelkan kejadian yang relatif jarang terjadi selama periode waktu yang dipilih. *Mean* dan varians dari distribusi ini dapat ditunjukkan melalui

$$E(Y) = \text{var}(Y) = \mu \quad (2.23)$$

Suatu model log linier poisson adalah GLM yang diasumsikan berdistribusi poisson untuk Y dan menggunakan fungsi *loglink*. Fungsi *loglink* diperoleh dari turunan keluarga eksponensial, yang ditampilkan pada Persamaan berikut.

$$\begin{aligned} f(y) &= \frac{e^{-\mu} \mu^y}{y!} \\ &= e^{-\mu} \frac{1}{y!} \exp(\ln(\mu^y)) \\ &= e^{-\mu} \frac{1}{y!} \exp(y \ln(\mu)) \\ &= a(\mu) b(y) \exp(d(y) c(\mu)) \end{aligned} \quad (2.24)$$

Maka : $a(\mu) = e^{-\mu}$, $b(y) = \frac{1}{y!}$, $d(y) = y$, dan $c(\mu) = \ln(\mu)$, komponen $c(\mu)$ merupakan sebuah *link function*, sehingga untuk k variabel prediktor, model *loglinear poisson* memiliki bentuk sebagai berikut

$$\ln \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (2.25)$$

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) \quad (2.26)$$

dengan

$$\mathbf{x}_i = [1 \ x_{1i} \ x_{2i} \ \dots \ x_{ki}]^T \text{ dan}$$

$$\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]^T$$

dimana i merupakan unit observasi yaitu $i = 1, 2, \dots, n$

Pada *poisson regression* nilai \hat{y} dapat ditaksir melalui μ , hal tersebut dapat ditunjukkan melalui persamaan

$$y = E(Y|X)$$

$$\text{Karena } E(Y|X) = \mu, \text{ maka } y = \mu = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (2.27)$$

Nilai yang dihasilkan pada Persamaan (2.27) berupa positif kontinu, sehingga untuk menghasilkan nilai y yang integer, dilakukan pembulatan ke bawah pada nilai μ .

2.8 Penaksiran Parameter Model *Poisson Regression*

Penaksiran parameter model *poisson regression* dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) yaitu dengan cara memaksimumkan fungsi *likelihood*. Fungsi *likelihood* dari *poisson regression* disajikan pada persamaan berikut

$$\begin{aligned} \ln L(\boldsymbol{\beta}) &= \ln \left(\prod_{i=1}^n \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \right) \\ &= \sum_{i=1}^n \ln \left(\frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \right) \\ &= \sum_{i=1}^n (\ln(e^{-\mu_i}) + \ln(\mu_i^{y_i}) - \ln(y_i!)) \\ &= \sum_{i=1}^n (-\mu_i + y_i \ln \mu_i - \ln(y_i!)) \\ &= \sum_{i=1}^n (-e^{\mathbf{x}_i^T \boldsymbol{\beta}} + y_i \ln e^{\mathbf{x}_i^T \boldsymbol{\beta}} - \ln(y_i!)) \\ &= -\sum_{i=1}^n e^{\mathbf{x}_i^T \boldsymbol{\beta}} + \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} - \sum_{i=1}^n \ln(y_i!) \end{aligned} \quad (2.28)$$

Taksiran MLE untuk parameter β_j dinyatakan dengan $\hat{\beta}_j$ ($j=1,2,\dots,k$) yang merupakan penyelesaian dari turunan pertama fungsi logaritma natural dari *Likelihood*. Tahapan berikutnya, penurunan kedua terhadap $\boldsymbol{\beta}^T$ pada

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = -\sum_{i=1}^n \mathbf{x}_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) + \sum_{i=1}^n y_i \mathbf{x}_i \quad (2.29)$$

Kemudian Persamaan (2.29) disamakan dengankan nol untuk mendapatkan penyelesaiannya, namun penyelesaian dengan cara tersebut sering kali tidak

mendapatkan hasil yang eksplisit, sehingga alternatif yang bisa digunakan untuk mendapatkan penyelesaian dari MLE adalah dengan metode iterasi numerik *Newton-Rapshon*. Algoritma untuk optimalisasi metode *Newton-Rapshon* adalah sebagai berikut (Cameron & Travedi, 1998).

1. Menentukan nilai taksiran awal parameter ($(\beta_{(0)})$) yang diperoleh dari metode *Ordinary Least Square* (OLS) yaitu .

$$\beta_{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

dengan

$$\mathbf{X}_{n \times (k+1)} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \text{ dan}$$

$$\mathbf{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]^T$$

2. Membentuk vektor gradien \mathbf{g}

$$\mathbf{g}^T(\beta_{(m)})_{(k+1) \times 1} = \left[\frac{\partial \ln L(\beta)}{\partial \beta_0} \quad \frac{\partial \ln L(\beta)}{\partial \beta_1} \quad \frac{\partial \ln L(\beta)}{\partial \beta_2} \quad \cdots \quad \frac{\partial \ln L(\beta)}{\partial \beta_k} \right]$$

dimana k merupakan banyaknya parameter yang ditaksir

3. Membentuk matriks Hessian \mathbf{H} :

$$\mathbf{H}(\beta_{(m)})_{(k+1) \times (k+1)} = \begin{bmatrix} \frac{\partial^2 \ln L(\beta)}{\partial \beta_0^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \beta_1} & \cdots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_0 \beta_k} \\ & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1^2} & \cdots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \beta_k} \\ & & \ddots & \vdots \\ & & & \frac{\partial^2 \ln L(\beta)}{\partial \beta_k^2} \end{bmatrix}$$

dengan $\text{Var}(\beta) = -E[H^{-1}(\beta)]$

4. Memasukkan nilai $(\beta_{(0)})$ ke dalam elemen-elemen vektor \mathbf{g} dan matriks \mathbf{H} , sehingga diperoleh vektor $\mathbf{g}(\beta_{(0)})$ dan matriks $\mathbf{H}(\beta_{(0)})$
5. Mulai dari $m=0$ dilakukan iterasi pada persamaan :

$$\boldsymbol{\beta}_{(m+1)} = \boldsymbol{\beta}_{(m)} - \mathbf{H}_{(m)}^{-1} \mathbf{g}_{(m)}$$

Nilai $\boldsymbol{\beta}_{(m)}$ merupakan sekumpulan penaksir parameter yang konvergen pada iterasi ke= m .

6. Jika belum didapatkan penaksir parameter yang konvergen, maka dilakukan kembali langkah 5. Iterasi berhenti pada saat $\|(\boldsymbol{\beta}_{(m+1)}) - \boldsymbol{\beta}_{(m)}\| \leq \varepsilon$, dimana ε merupakan bilangan yang sangat kecil sekali.

2.9 Pengujian Parameter Model *Poisson Regression*

Salah satu metode yang dipakai untuk menentukan statistik uji dalam pengujian parameter model *poisson regression* adalah *Maximum Likelihood Ratio Test* (MLRT) dengan hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_j \neq 0 \text{ dengan } j=1, 2, \dots, k$$

Statistik uji untuk kelayakan model *poisson regression* adalah sebagai berikut.

$$D(\boldsymbol{\beta}) = -2 \ln \left[\frac{L(\omega)}{L(\Omega)} \right] = 2[\ln L(\Omega) - L(\omega)] \quad (2.30)$$

Dengan $L(\omega)$ merupakan fungsi likelihood maksimum di bawah H_0 , dan $L(\Omega)$ adalah fungsi likelihood maksimum di bawah populasi. Keputusan yang akan diambil adalah tolak H_0 jika $D(\boldsymbol{\beta}) > \chi^2_{v,\alpha}$ dengan v adalah banyaknya parameter model dibawah populasi dikurangi dengan banyaknya parameter dibawah H_0 . Parameter model *poisson regression* yang telah dihasilkan dari estimasi parameter belum tentu mempunyai pengaruh yang signifikan terhadap model. Untuk itu perlu dilakukan pengujian terhadap parameter model *poisson regression* secara individu dengan menggunakan hipotesis berikut.

$$H_0 : \beta_j = 0 \text{ (pengaruh variabel ke-} j \text{ tidak signifikan)}$$

$$H_1 : \beta_j \neq 0 \text{ (pengaruh variabel ke-} j \text{ signifikan) dengan } j=0, 1, 2, \dots, k$$

Statistik uji yang digunakan adalah :

$$z = \frac{\beta_j}{SE(\beta_j)} \quad (2.31)$$

$SE(\beta_j)$ adalah nilai *standar error* atau tingkat kesalahan dari parameter β_j .

Keputusan yang akan diambil adalah tolak H_0 jika $|Z_{hit}| > Z_{\alpha/2}$, dengan α adalah tingkat signifikansi yang digunakan.

2.10 Multikolinieritas

Salah satu syarat yang harus dipenuhi dalam pembentukan model regresi poisson dengan beberapa variabel adalah tidak adanya kasus multikolinieritas (Myers, 1990). Multikolinieritas adalah kondisi terdapatnya hubungan linier atau korelasi yang tinggi antara satu variabel prediktor dengan variabel prediktor yang lain. Dalam model regresi, adanya korelasi antar variabel prediktor menyebabkan estimasi parameter regresi yang dihasilkan akan memiliki error yang sangat besar. Pendeteksian kasus multikolinieritas menurut Hocking (1996) dapat dilihat melalui beberapa cara yaitu sebagai berikut:

1. Jika koefisien korelasi Pearson (r_{jj}) antar variabel prediktor lebih dari 0,95 maka terdapat korelasi antar variabel tersebut.
2. Nilai VIF (*Varian Inflation Factor*) lebih besar dari 10 menunjukkan adanya multikolinieritas antar variabel prediktor. Nilai VIF dinyatakan sebagai berikut:

$$VIF_j = \frac{1}{1-R_j^2}, \quad (2.32)$$

Dengan R_j^2 adalah koefisien determinasi antara x_j dengan variabel prediktor lainnya.

Solusi untuk mengatasi adanya kasus multikonlinieritas yaitu dengan mengeluarkan variabel prediktor yang tidak signifikan dalam model atau dengan cara mengelompokkan variabel yang saling berkorelasi cukup tinggi dalam sebuah komponen yang membentuk variabel baru yaitu menggunakan *Principial Component Analysis* (PCA), sehingga mereduksi banyaknya dimensi regresi dan antar variabel baru tersebut tidak saling berkorelasi cukup tinggi.

2.11 *Overdispersi*

Overdispersi dalam *Poisson regression* terjadi apabila nilai varians data lebih besar daripada nilai meannya. Jika dalam *Poisson regression* terdapat kasus overdispersi, namun tetap menggunakan *Poisson regression* maka akan menyebabkan estimasi dari parameter koefisien regresinya tetap konsisten namun tidak efisien. Hal ini akan berdampak pada nilai standar error yang menjadi *underestimate*, sehingga tidak valid. Overdispersi (McCullagh & Nelder, 1989) dapat dituliskan sebagai berikut:

$$\text{Var}(Y) > E(Y) \quad (2.33)$$

Overdispersi dapat dideteksi dengan nilai dispersi Pearson Chi-square dan devians yang dibagi dengan derajat bebasnya. Jika hasil pembagian kedua nilai tersebut lebih besar dari 1, maka dapat dikatakan terjadi overdispersi pada data. Overdispersi dapat terjadi apabila ada pengamatan missing pada peubah penjelas, adanya pencilan data, perlunya interaksi dalam model, peubah penjelas perlu ditransformasi atau kesalahan spesifikasi link. Jika pada *Poisson regression* terjadi overdispersi, maka salah satu alternatif yang dapat digunakan adalah *Negative Binomial regression* (Hardin & Hilbe, 2007).

2.12 *Negative Binomial Regression*

Sebagai penerapan dari GLM, maka distribusi *Negative Binomial* memiliki tiga komponen yaitu komponen random, komponen sistematik dan fungsi link (Greene, 2008). Pada *Negative Binomial regression* yang dihasilkan dari distribusi *Mixture Poisson-Gamma*. Untuk membentuk suatu model regresi pada distribusi *Mixture Poisson-Gamma* dinyatakan dalam bentuk $\mu = \alpha\beta$ dan $\theta = \frac{1}{\alpha}$ sehingga diperoleh mean dan varians dalam bentuk:

$$E(Y) = \mu \text{ dan } V[Y] = \mu + \theta\mu^2 \quad (2.33)$$

dengan fungsi massa peluang *Negative Binomial* :

$$f(y, \mu, \theta) = \frac{\Gamma(y+1/\theta)}{\Gamma(1/\theta)y!} \left(\frac{1}{1+\theta\mu}\right)^{1/\theta} \left(\frac{\theta\mu}{1+\theta\mu}\right)^y \quad (2.34)$$

Saat $\theta = 0$ maka distribusi *Negative Binomial* memiliki varians $V[Y] = \mu$. Distribusi *Negative Binomial* akan mendekati suatu *Poisson distribution* yang

mengasumsikan mean dan varians sama yaitu $E[Y] = V[Y] = \mu$. Fungsi distribusi keluarga eksponensial dari distribusi *negative binomial* (Greene, 2008) adalah:

$$f(y, \mu, \theta) = \exp \left\{ y \ln \left(\frac{\theta \mu}{1 + \theta \mu} \right) + \frac{1}{\theta} \ln \left(\frac{1}{1 + \theta \mu} \right) + \ln \left(\frac{\Gamma(y+1/\theta)}{\Gamma(1/\theta) y!} \right) \right\} \quad (2.35)$$

Kontribusi variabel prediktor dalam model *negative binomial regression* dinyatakan dalam bentuk kombinasi linier antara parameter (η) dengan parameter regresi yang akan diestimasi yaitu:

$$\eta_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} \quad (2.36)$$

Atau dalam matriks dituliskan dalam bentuk

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \quad (2.37)$$

dengan $\boldsymbol{\eta}$ adalah vektor ($n \times 1$) dari observasi, \mathbf{X} adalah matriks ($n \times (k+1)$) dari variabel prediktor, $\boldsymbol{\beta}$ adalah matriks ($(k+1) \times 1$) dari koefisien regresi. Nilai ekspektasi dari variabel respon Y adalah diskrit dan bernilai positif. Maka untuk mentransformasikan nilai η_i (bilangan riil) ke rentang yang sesuai dengan rentang pada respon Y yang diperlukan suatu fungsi link $g(\cdot)$ (Greene, 2008) yaitu:

$$g(\mu_i) = \ln \mu_i = \mathbf{X}_i^T \boldsymbol{\beta} \quad (2.38)$$

2.13 Penaksiran Parameter Model *Negative Binomial Regression*

Estimasi parameter dari *negative binomial regression* digunakan metode maksimum *likelihood* dengan prosedur Newton Raphson. Metode ini membutuhkan turunan pertama dan kedua dari fungsi *likelihood*. Y_i mempunyai fungsi massa probabilitas distribusi *negative binomial* sebagai berikut ini:

$$f(y_i | \mu_i, \theta) = \frac{\Gamma(y_i+1/\theta)}{\Gamma(1/\theta)\Gamma(y_i+1)} \left(\frac{1}{1+\theta\mu_i} \right)^{1/\theta} \left(\frac{\theta\mu_i}{1+\theta\mu_i} \right)^{y_i}; y_i = 0, 1, 2, \dots \quad (2.39)$$

Karena fungsinya saling bebas, maka fungsi *log-likelihood* adalah:

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^n \frac{\Gamma(y_i+1/\theta)}{\Gamma(1/\theta)\Gamma(y_i+1)} \left(\frac{1}{1+\theta\mu_i} \right)^{1/\theta} \left(\frac{\theta\mu_i}{1+\theta\mu_i} \right)^{y_i}$$

$$\text{Dengan } \frac{\Gamma(y_i+1/\theta)}{\Gamma(1/\theta)} = \prod_{r=1}^{y_i-1} (r + \theta^{-1})$$

$$L(\boldsymbol{\beta}, \theta) = \prod_{i=1}^n \left(\prod_{r=1}^{y_i-1} (r + \theta^{-1}) \right) \frac{1}{(y_i!)} \left(\frac{1}{1+\theta\mu_i} \right)^{1/\theta} \left(\frac{\theta\mu_i}{1+\theta\mu_i} \right)^{y_i}$$

$$L(\boldsymbol{\beta}, \theta) = \ln\{L(\boldsymbol{\beta}, \theta)\}$$

$$= \sum_{i=1}^n [(\sum_{r=1}^{y_i-1} \ln(r + \theta^{-1})) - \ln(y_i!) + y_i \ln(\theta \mu_i) - (\theta^{-1} y_i) \ln(1 + \theta \mu_i)]$$

Turunan pertama dari fungsi log-likelihood terhadap koefisien $\boldsymbol{\beta}$ adalah:

$$\frac{\partial L(\boldsymbol{\beta}, \theta)}{\partial \beta_0} = \sum_{i=1}^n [y_i - (y_i + \theta^{-1}) \left(\frac{\theta \mu_i}{1 + \theta \mu_i} \right)] = \sum_{i=1}^n \left[\frac{y_i - \mu_i}{1 + \theta \mu_i} \right] = 0$$

⋮

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta}, \theta)}{\partial \beta_k} &= \sum_{i=1}^n \left[y_i x_{ik} - (y_i + \theta^{-1}) \left(\frac{\theta \mu_i x_{ik}}{1 + \theta \mu_i} \right) \right] \\ &= \sum_{i=1}^n \left[\frac{(y_i - \mu_i) x_{ik}}{1 + \theta \mu_i} \right] \end{aligned}$$

$$= \sum_{i=1}^n \left[\frac{\mu_i}{1 + \theta \mu_i} \frac{(y_i - \mu_i) x_{ik}}{\mu_i} \right] = 0$$

Bentuk persamaan matriks dari turunan pertama fungsi log-likelihood terhadap parameter $\boldsymbol{\beta}$ yaitu: $\mathbf{q} = \mathbf{X}^T \mathbf{W} \mathbf{z}$, dengan \mathbf{X} adalah matriks ($n \times (k+1)$) dari variabel prediktor, \mathbf{W} adalah matriks *weight* diagonal ke- i dan \mathbf{z} adalah vektor matriks dengan baris ke- i , dengan masing masing elemennya adalah:

$$w_i = \frac{\mu_i}{1 + \theta \mu_i} \quad \text{dan} \quad z_i = \frac{(y_i - \mu_i)}{\mu_i} \quad i = 1, 2, \dots, n$$

Turunan pertama dari fungsi log-likelihood terhadap parameter dispersi θ adalah:

$$f'(\theta) = \frac{\partial L(\boldsymbol{\beta}, \theta)}{\partial \theta}$$

$$= \sum_{i=1}^n \left[-\theta^{-2} \sum_{r=0}^{y_i-1} \frac{1}{r + \theta^{-1}} + \frac{y_i}{\theta} + \theta^{-2} \ln(1 + \theta \mu_i) \frac{(\theta^{-1} + y_i) \mu_i}{1 + \theta \mu_i} \right] = 0$$

$$f'(\theta) = \frac{\partial L(\boldsymbol{\beta}, \theta)}{\partial \theta} = \sum_{i=1}^n \left[-\theta^{-2} \sum_{r=0}^{y_i-1} \frac{1}{r + \theta^{-1}} + \theta^{-2} \ln(1 + \theta \mu_i) + \frac{y_i - \mu_i}{\theta(1 + \theta \mu_i)} \right] = 0$$

Turunan parsial kedua fungsi log-likelihood terhadap parameter koefisien regresi $\boldsymbol{\beta}$ adalah:

$$\frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_0^2} = - \sum_{i=1}^n \left[\frac{(1 + \theta y_i) \mu_i}{(1 + \theta \mu_i)^2} \right]$$

$$\frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_0 \partial \beta_j} = \sum_{i=1}^n \left[\frac{-\mu_i x_{ij} (1 + \theta \mu_i) - (y_i - \mu_i) (x_{ij} \theta \mu_i)}{(1 + \theta \mu_i)^2} \right]$$

$$= - \sum_{i=1}^n \left[\frac{(1 + \theta y_i) x_{ij} \mu_i}{(1 + \theta \mu_i)^2} \right]$$

Misalkan turunan parsial pertama dari $L(\boldsymbol{\beta}, \theta)$ terhadap $\beta_j, j \leq k$ adalah:

$$\frac{\partial L(\boldsymbol{\beta}, \theta)}{\partial \beta_j} = \sum_{i=1}^n \left[\frac{(y_i - \mu_i) x_{ij}}{1 + \theta \mu_i} \right] = 0, \text{ maka turunan parsial kedua terhadap } \beta_u, u \leq k$$

adalah:

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_u \partial \beta_j} &= \sum_{i=1}^n \left[\frac{-\mu_i x_{iu} x_{ij} (1 + \theta \mu_i) - (y_i - \mu_i) (x_{iu} x_{ij} \theta \mu_i)}{(1 + \theta \mu_i)^2} \right] \\ &= - \sum_{i=1}^n \left[\frac{(1 + \theta y_i) x_{iu} x_{ij} \mu_i}{(1 + \theta \mu_i)^2} \right] \end{aligned}$$

Ekspektasi dari turunan kedua log-likelihood adalah:

$$E \left(- \frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \beta_u \partial \beta_j} \right) = \sum_{i=1}^n \left[\frac{x_{iu} x_{ij} \mu_i}{(1 + \theta \mu_i)} \right] \quad (2.40)$$

Jika persamaan (2.40) dinyatakan dalam matriks **I** yaitu matriks yang mengandung ekspektasi negatif dari turunan kedua log-likelihood maka:

$$\mathbf{I} = \mathbf{X}^T \mathbf{W} \mathbf{X},$$

dengan **X** adalah matriks dari variabel prediktor, **W** adalah matriks weight diagonal ke-*i* dengan elemen : $w_i = \frac{\mu_i}{1 + \theta \mu_i}$

Turunan kedua fungsi log-likelihood terhadap parameter dispersi θ adalah

$$\begin{aligned} f''(\theta) &= \frac{\partial^2 L(\boldsymbol{\beta}, \theta)}{\partial \theta^2} \\ &= \sum_{i=1}^n \left[\theta^{-3} \sum_{r=0}^{y_i-1} \frac{(2r + \theta^{-1})}{(r + \theta^{-1})^2} - 2\theta^{-3} \ln(1 + \theta \mu_i) + \frac{-\theta^2 \mu_i}{(1 + \theta \mu_i)} - \frac{(y_i - \mu_i)(1 + 2\theta \mu_i)}{(\theta + \theta^2 \mu_i)^2} \right] \end{aligned}$$

Langkah- langkah estimasi parameter *negative binomial regression* dilakukan dengan langkah sebagai berikut:

1. Menentukan estimasi awal θ , misal $\hat{\theta}_1 = 0,1$
2. Menentukan estimasi maksimum *likelihood* dari parameter $\boldsymbol{\beta}$ menggunakan prosedur iterasi Fisher scoring dengan asumsi $\theta = \hat{\theta}_1$

$$\hat{\boldsymbol{\beta}}_{i+1} = \hat{\boldsymbol{\beta}}_i + (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{z}_i$$

Iterasi berakhir jika diperoleh $\|\hat{\boldsymbol{\beta}}_{i+1} - \hat{\boldsymbol{\beta}}_i\| \leq \varepsilon$.

3. Menggunakan $\hat{\boldsymbol{\beta}}$ untuk menghasilkan estimasi dari parameter θ dengan menggunakan prosedur iterasi Newton-Raphson satu variabel, iterasi berakhir jika diperoleh $|\hat{\theta}_{i+1} - \hat{\theta}_i| \leq \varepsilon$

$$\hat{\theta}_{i+1} = \hat{\theta}_i - \frac{f'(\theta_i)}{f''(\theta_i)}$$

4. Jika $|\hat{\theta}_{i+1} - \hat{\theta}_i| \leq \varepsilon$ selesai; bila tidak, gunakan parameter $\theta = \hat{\theta}_{i+1}$ dan kembali ke langkah 2, nilai ε merupakan bilangan positif yang sangat kecil.

2.14 Pengujian Parameter Model *Negative Binomial Regression*

Uji kesesuaian model *negative binomial regression* dengan uji devians sebagai berikut:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_j \neq 0 \quad j = 1, 2, \dots, k$$

Statistik Uji:

$$D(\hat{\beta}) = -2 \ln \Lambda = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) = 2[\ln L(\hat{\Omega}) - L(\hat{\omega})] \quad (2.41)$$

Kriteria Penolakan: Tolak H_0 jika nilai $D(\hat{\beta}) > \chi^2_{(\alpha; k)}$, artinya paling sedikit ada satu variabel yang memberikan pengaruh pada model.

Setelah dilakukan uji serentak, dilanjutkan dengan pengujian parsial setiap variabel prediktornya dengan menggunakan uji Wald. Hipotesisnya adalah sebagai berikut:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, \quad j = 0, 1, 2, \dots, k$$

Statistik uji:

$$W_j = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right)^2 \quad (2.42)$$

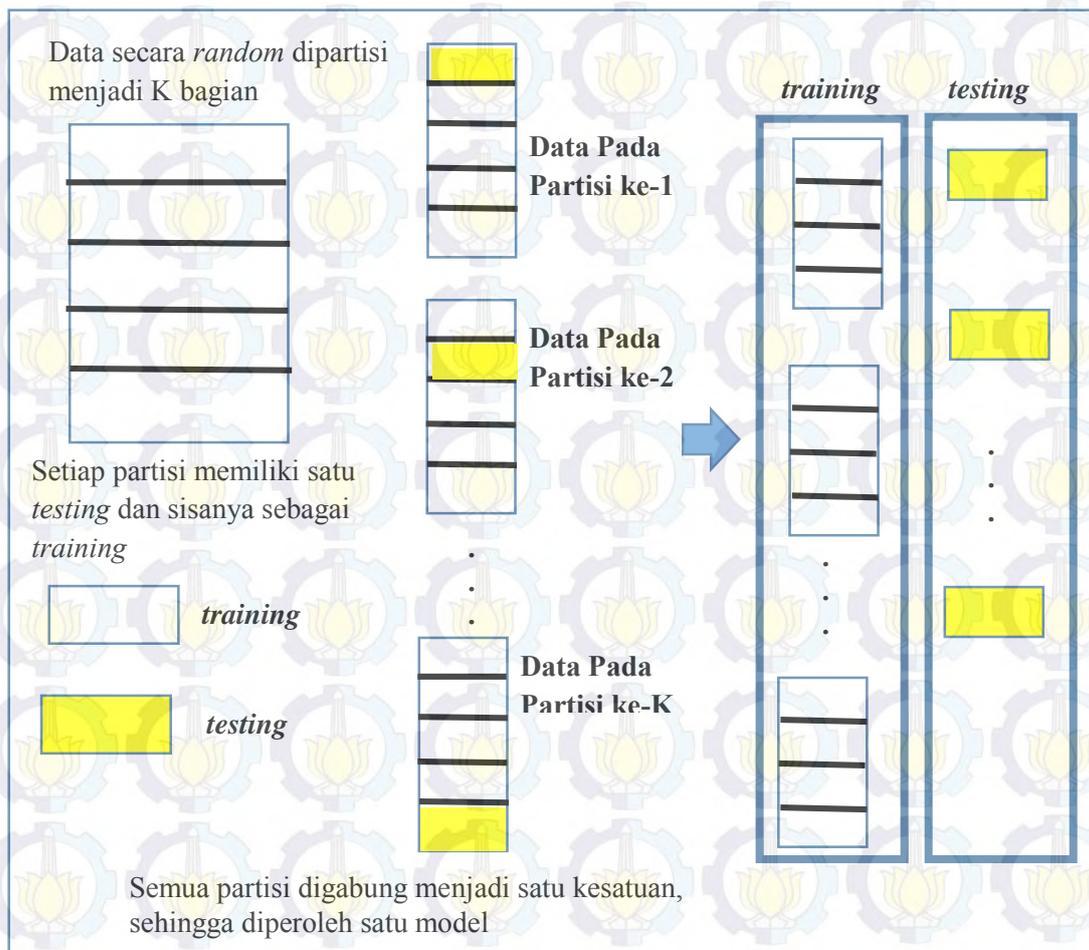
Kriteria Penolakan: Tolak H_0 jika nilai $W_j > \chi^2_{(\alpha; k)}$, artinya variabel j memberikan pengaruh pada model.

2.15 *K-Fold Cross Validation*

Permasalahan yang dihadapi dalam *neural networks* adalah cara penentuan nilai bobot dan bias yang optimum. Secara umum, solusi yang dilakukan untuk mengatasi permasalahan tersebut adalah dengan membagi data menjadi data *training* dan *testing*. Data *training* dipakai sebagai penentuan nilai bobot dan bias

terbaik dan *testing* digunakan untuk menguji dan memvalidasi bobot dan bias yang dihasilkan.

Metode yang sering dipakai untuk membagi data menjadi *training* dan *testing* adalah *Hold out*, yaitu salah satu metode *cross validation* yang paling sederhana. Kenyatannya pendekatan *hold out* seringkali masih beresiko karena *randomisasi* pembagian data training maupun testing yang tidak merata, dapat menyebabkan hasil yang kurang sesuai (Caffrey, 2013). *K-fold* merupakan salah satu metode yang dipakai untuk mengatasi permasalahan pada *hold-out*. Metode ini membagi data dengan ukuran yang hampir sama pada *training* maupun *testing*. Setiap iterasi ke-*k*, beberapa data akan dipilih untuk menjadi data *testing*, sementara sisanya akan dipakai sebagai data *training*. Ilustrasi dari *k-fold* dijelaskan pada Gambar 2.4



Gambar 2.4 Ilustrasi *K-Fold*

Selanjutnya nilai *cross validation error estimate* pada partisi $(1,2,\dots,K)$ dihitung sebagai berikut

$$CV_{(K)} = \sum_{k=1}^K \frac{n_k}{n} MSE_k \quad (2.43)$$

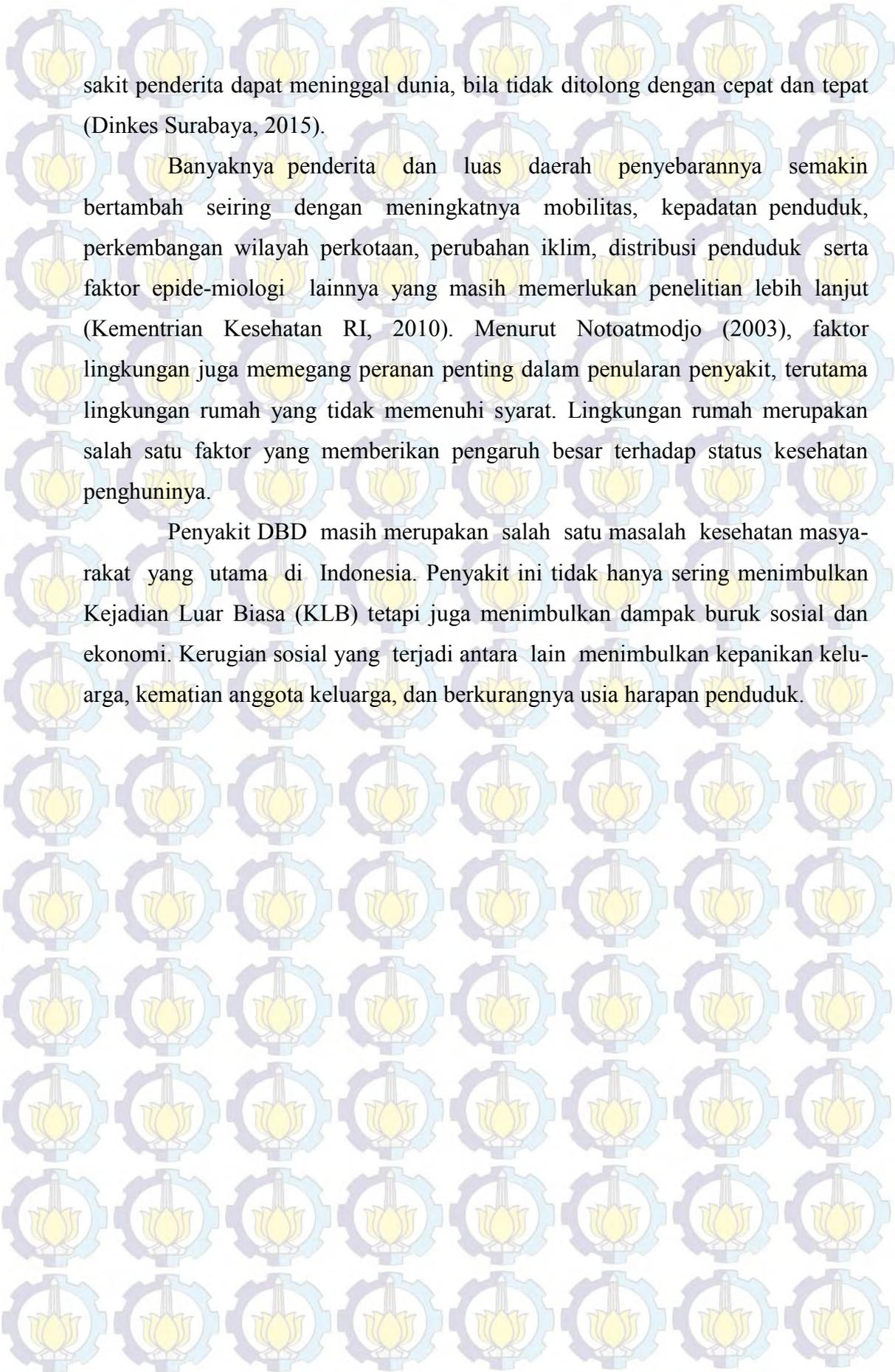
dimana MSE_k merupakan MSE pada data *testing* ke- k , dengan $k = (1,2,\dots,K)$ dan dihitung dengan Persamaan berikut.

$$MSE_k = \frac{\sum_{i \in k} (y_i - \hat{y}_i)^2}{n_k} \quad (2.44)$$

dengan y_i merupakan data pada partisi ke- k , \hat{y}_i ialah nilai taksiran data pada partisi ke- k , n_k yaitu banyaknya data pada partisi ke- k , dan n merupakan banyaknya data keseluruhan. CV merupakan salah satu kriteria terbaik untuk memilih model, sehingga pada tahapan berikutnya model terbaik dipilih berdasarkan CV terkecil. Saat banyaknya partisi dari *fold* sama dengan banyaknya observasi ($K=n$), metode tersebut adalah kasus khusus dari *k-fold* yang biasa disebut dengan *Leave-One-Out (LOO) Cross Validation*.

2.16 Demam Berdarah Dengue (DBD)

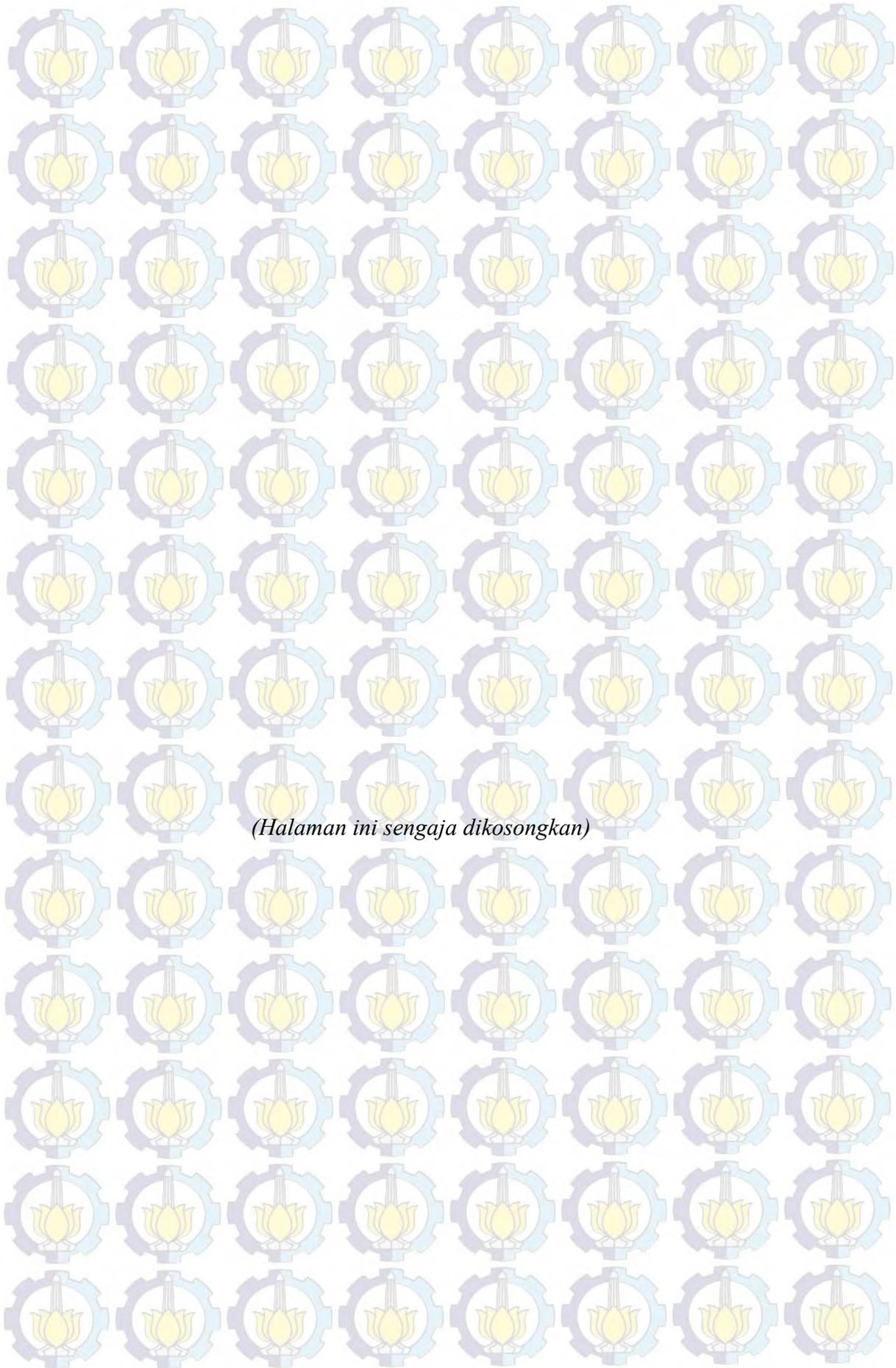
Penyakit Demam Berdarah Dengue (DBD) atau *Dengue Haemorrhagic Fever* (DHF) mulai dikenal di Indonesia sejak tahun 1968 di Surabaya dan Jakarta (Dinkes Jatim, 2013). DBD banyak ditemukan di daerah tropis dan subtropis. Penyakit ini disebabkan oleh virus Dengue dari genus *Flavivirus*, famili *Flaviviridae*. DBD ditularkan ke manusia melalui gigitan nyamuk *Aedes* yang terinfeksi virus Dengue. Virus Dengue penyebab Demam Dengue (DD), Demam Berdarah Dengue (DBD) dan *DengShock Syndrome* (DSS) termasuk dalam kelompok *B Arthropod Virus* (*Arbovirosis*) yang sekarang dikenal sebagai genus *Flavivir* famili *Flaviviride*, dan mempunyai 4 jenis serotipe, yaitu: Den-1, Den-2, Den-3, Den-4. Cara bekerja virus dengue seminggu setelah digigit nyamuk yang mengandung virus, seseorang bisa jatuh sakit atau tidak sakit tetapi namun tetap sebagai sumber berpenular bagi orang lain (*carrier*). Kurang dari tujuh hari setelah



sakit penderita dapat meninggal dunia, bila tidak ditolong dengan cepat dan tepat (Dinkes Surabaya, 2015).

Banyaknya penderita dan luas daerah penyebarannya semakin bertambah seiring dengan meningkatnya mobilitas, kepadatan penduduk, perkembangan wilayah perkotaan, perubahan iklim, distribusi penduduk serta faktor epide-miologi lainnya yang masih memerlukan penelitian lebih lanjut (Kementerian Kesehatan RI, 2010). Menurut Notoatmodjo (2003), faktor lingkungan juga memegang peranan penting dalam penularan penyakit, terutama lingkungan rumah yang tidak memenuhi syarat. Lingkungan rumah merupakan salah satu faktor yang memberikan pengaruh besar terhadap status kesehatan penghuninya.

Penyakit DBD masih merupakan salah satu masalah kesehatan masyarakat yang utama di Indonesia. Penyakit ini tidak hanya sering menimbulkan Kejadian Luar Biasa (KLB) tetapi juga menimbulkan dampak buruk sosial dan ekonomi. Kerugian sosial yang terjadi antara lain menimbulkan kepanikan keluarga, kematian anggota keluarga, dan berkurangnya usia harapan penduduk.



(Halaman ini sengaja dikosongkan)

BAB 3

METODOLOGI PENELITIAN

3.1 Bahan dan Alat

Dalam penelitian ini digunakan alat dan bahan untuk mendukung proses penelitian yaitu sebagai berikut.

1. Jurnal yang berkaitan dengan metode yang dipakai dalam penelitian seperti: *Journal Association for European Transport and Contributors 2012, Neural networks to Estimate Crashes at Zona Level for Transportation Planning, Neural networks and Statistical Model* dan referensi yang terkait.
2. Paket program komputer, yaitu software statistik Matlab R2009A.
3. Pada penelitian ini, perbandingan metode *neural networks* dengan *poisson regression* akan diterapkan pada data Profil Kesehatan Kota Surabaya Tahun 2013, yang diperoleh dari Dinas Kesehatan Kota Surabaya. Penelitian ini juga menggunakan data kemiskinan tahun 2013 yang diambil dari Badan Pusat Statistik (BPS) kota Surabaya.

Unit observasi yang digunakan dalam penelitian ini ialah 31 kecamatan di Kota Surabaya. Adapun variabel penelitian yang digunakan dalam penelitian ini terdiri dari variabel respon (y) yaitu banyaknya kasus DBD di tiap kecamatan dan variabel prediktor (x) yang ditunjukkan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian (Profil Kesehatan Kota Surabaya, 2013)

Variabel	Nama Variabel	Definisi Operasional
x_1	Persentase rumah/bangunan bebas jentik nyamuk Aedes	Hasil bagi antara banyaknya rumah atau bangunan yang bebas jentik nyamuk Aedes terhadap banyaknya bangunan yang dipantau dikalikan 100%

Tabel 3.1 Variabel Penelitian (Lanjutan)

Variabel	Nama Variabel	Definisi Operasional
x_2	Persentase rumah tangga miskin	Hasil bagi antara banyak rumah tangga miskin terhadap banyak rumah tangga seluruhnya dikalikan 100%.
x_3	Persentase rumah tangga yang berperilaku hidup bersih dan sehat (PHBS)	Hasil bagi dari banyaknya rumah tangga yang berperilaku hidup bersih dan sehat terhadap banyaknya rumah tangga yang dipantau di tiap kecamatan di kota Surabaya dikalikan dengan 100%.
x_4	Kepadatan Penduduk	Hasil bagi dari jumlah penduduk terhadap luas wilayah di tiap kecamatan di Kota Surabaya dalam satuan (orang/km ²).
x_5	Rasio tenaga medis (dokter umum)	Rasio antara banyak dokter umum dengan jumlah penduduk di suatu wilayah. Menurut standar WHO, satu dokter umum melayani 2.500 penduduk. Berikut merupakan perhitungan rasio dokter umum: $Rasio = \frac{\text{banyak dokter umum}}{\text{jumlah penduduk}} \times 2.500$
x_6	Persentase rumah tangga yang memiliki tempat sampah sehat	Hasil bagi antara banyaknya rumah yang memiliki tempat sampah sehat dengan banyak rumah yang dipantau dikalikan 100%.
x_7	Rasio sarana kesehatan Puskesmas	Rasio antara banyaknya Puskesmas dengan jumlah penduduk di suatu wilayah. Menurut standar Dinas Kesehatan, satu Puskesmas melayani 30.000 penduduk. Berikut merupakan perhitungan Puskesmas: $Rasio = \frac{\text{banyak Puskesmas}}{\text{jumlah penduduk}} \times 30.000$
x_8	Persentase Rumah Sehat	Hasil bagi antara banyaknya rumah kategori sehat dengan banyaknya rumah yang dipantau dikalikan 100%

3.2 Metode Penelitian

Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

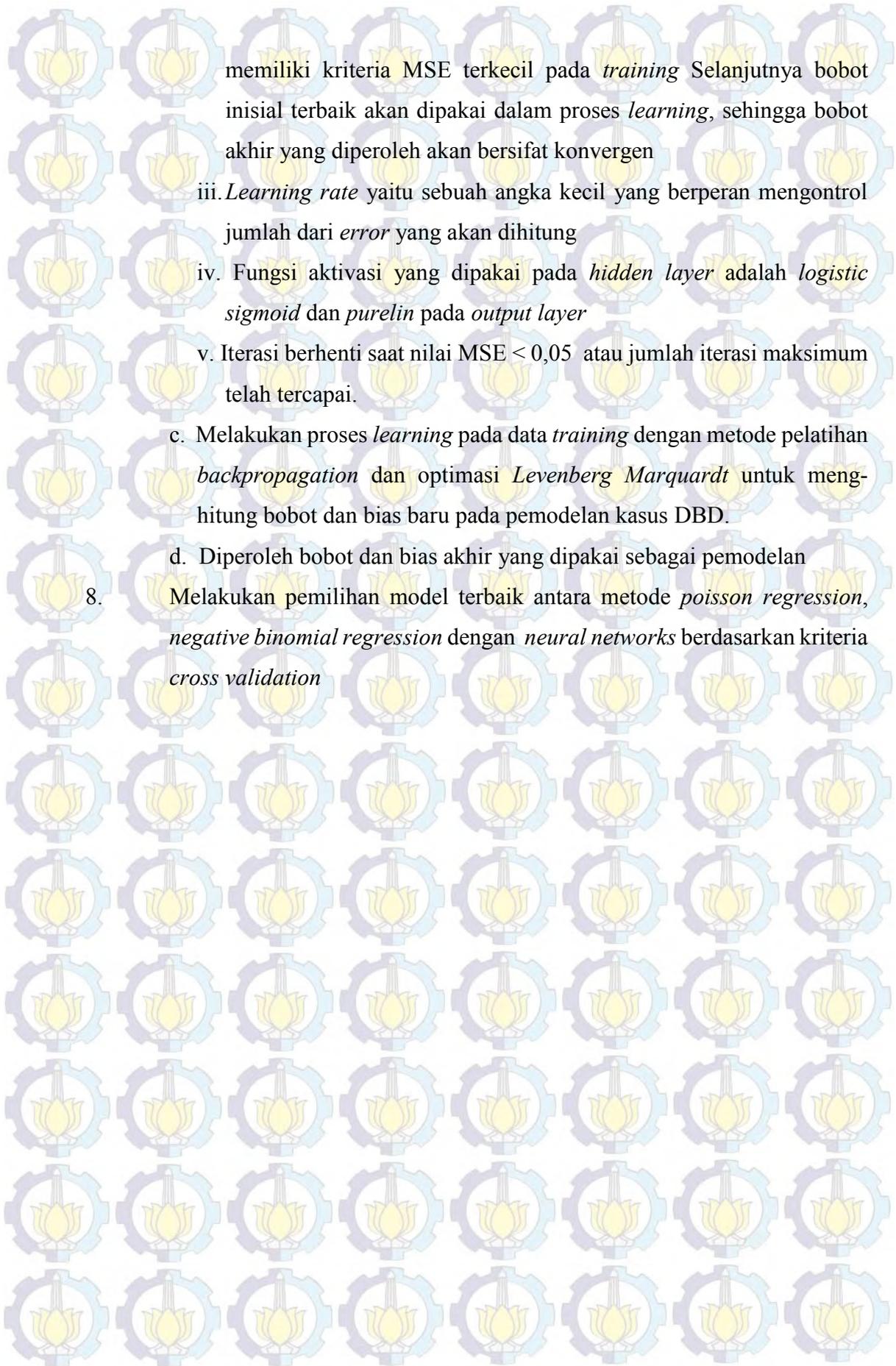
1. Melakukan identifikasi hubungan antar variabel banyaknya kasus DBD dengan faktor-faktor yang mempengaruhinya menggunakan diagram pencar serta korelasi.
2. Mengecek ada atau tidaknya *outlier* pada data dengan menggunakan *box-plot* serta nilai *standardize z score* pada setiap variabel untuk mengindikasikan ada atau tidaknya kasus overdispersi
3. Membagi data menjadi data *training* dan *testing*, dengan metode *k-fold cross validation* sebanyak 5. Pembagian dilakukan sesuai ilustrasi yang telah dijelaskan pada Sub bab 2.14.
4. Melakukan pengujian multikolinieritas antar variabel prediktor terhadap kasus DBD di kota Surabaya, pada data keseluruhan (sebelum dilakukan pembagian *training* maupun *testing*) serta pada data *training*. Jika tidak terdapat kasus multikolinieritas pada data, maka dapat dilanjutkan pada langkah berikutnya
5. Melakukan pemodelan *poisson regression* pada data banyaknya kasus DBD di kota Surabaya dengan menggunakan data keseluruhan serta pada data *training* Langkah yang dipakai pada kedua pemodelan tidak jauh berbeda, yaitu sebagai berikut.
 - a. Menaksir parameter model *poisson regression* dengan metode MLE
 - b. Menguji signifikansi parameter yang dihasilkan secara serentak maupun parsial. Jika terdapat beberapa parameter yang tidak signifikan pada pengujian parsial, maka dilakukan kembali langkah ke a dan b hanya dengan variabel yang signifikan pada pengujian parsial, setelah semua parameter signifikan dapat dilanjutkan pada langkah berikutnya.
 - c. Melakukan uji dispersi model *poisson regression*
 - d. Menghitung nilai MSE pada data *training*
 - e. Menghitung nilai MSE pada data *testing*
 - f. Melakukan interpretasi pada model terpilih

Pada pemodelan dengan menggunakan data keseluruhan, tahapan yang dipakai hanya pada point a sampai b.

6. Melakukan pemodelan *negative binomial regression* pada data kasus DBD di kota Surabaya dengan menggunakan data keseluruhan serta pada data *training*. Langkah yang dipakai pada kedua pemodelan tidak jauh berbeda, yaitu sebagai berikut.
 - a. Menaksir parameter model *negative binomial* dengan metode MLE
 - b. Menguji signifikansi parameter yang dihasilkan pada model *negative binomial regression* secara serentak maupun parsial. Jika terdapat beberapa parameter yang tidak signifikan pada pengujian parsial, maka dilakukan kembali langkah ke a dan b hanya dengan variabel yang signifikan pada pengujian parsial, setelah semua parameter signifikan dapat dilanjutkan pada langkah berikutnya.
 - c. Melakukan uji dispersi model *negative binomial regression*
 - d. Menghitung nilai MSE pada data *training*
 - e. Menghitung MSE pada data *testing*
 - f. Melakukan interpretasi pada model terpilih

Pada pemodelan dengan menggunakan data keseluruhan, tahapan yang dipakai hanya pada point a sampai b.

7. Melakukan pemodelan *neural networks* pada data kasus DBD di kota Surabaya dengan menggunakan data *training*.
 - a. Menentukan jumlah *input unit*, *hidden unit*, dan *output unit*. Jumlah node untuk *hidden unit* ditentukan berdasarkan *trial* dan *error*.
 - b. Menentukan *training setting* dan menetapkan fungsi aktivasi yang dipakai
 - i. Banyaknya iterasi maksimum, yaitu banyaknya siklus penyesuaian bobot yang dilakukan selama proses pelatihan berjalan. Pada penelitian ini, jumlah iterasi maksimum ditentukan sebanyak 1000.
 - ii. Inisialisasi bobot awal. Bobot awal akan di-*set* dengan suatu nilai tertentu, yang diperoleh secara random dengan melakukan iterasi sebanyak 100 kali. Dari hasil iterasi akan dipilih satu bobot yang



memiliki kriteria MSE terkecil pada *training* Selanjutnya bobot inisial terbaik akan dipakai dalam proses *learning*, sehingga bobot akhir yang diperoleh akan bersifat konvergen

iii. *Learning rate* yaitu sebuah angka kecil yang berperan mengontrol jumlah dari *error* yang akan dihitung

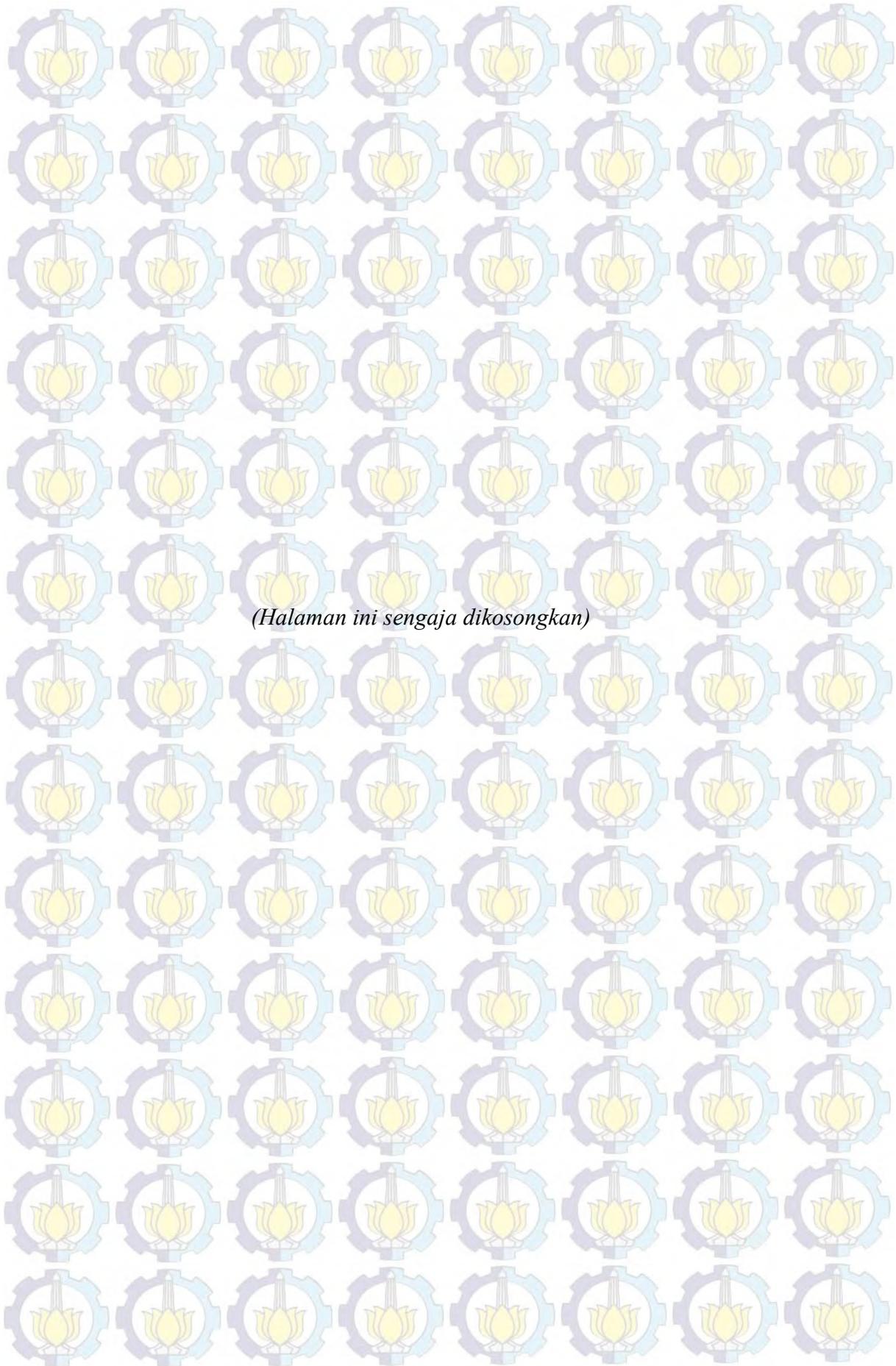
iv. Fungsi aktivasi yang dipakai pada *hidden layer* adalah *logistic sigmoid* dan *purelin* pada *output layer*

v. Iterasi berhenti saat nilai $MSE < 0,05$ atau jumlah iterasi maksimum telah tercapai.

c. Melakukan proses *learning* pada data *training* dengan metode pelatihan *backpropagation* dan optimasi *Levenberg Marquardt* untuk menghitung bobot dan bias baru pada pemodelan kasus DBD.

d. Diperoleh bobot dan bias akhir yang dipakai sebagai pemodelan

8. Melakukan pemilihan model terbaik antara metode *poisson regression*, *negative binomial regression* dengan *neural networks* berdasarkan kriteria *cross validation*



(Halaman ini sengaja dikosongkan)

BAB 4

ANALISIS DAN PEMBAHASAN

Pada bagian ini akan dilakukan analisis dan pembahasan dari data yang telah diolah dengan menggunakan beberapa metode analisis yang sudah dibahas pada bagian sebelumnya. Sebelum dilakukan pemodelan banyaknya kasus DBD di Surabaya, terlebih dahulu akan dilakukan identifikasi pola hubungan antara variabel banyaknya kasus DBD di Surabaya dengan faktor-faktor yang mempengaruhinya. Identifikasi dilakukan melalui pembuatan diagram pencar (*scatter plot*) dan korelasi antara variabel respon dengan beberapa prediktor yang mempengaruhinya. Nilai koefisien korelasi dan *p_value* yang dihasilkan disajikan pada Tabel 4.1.

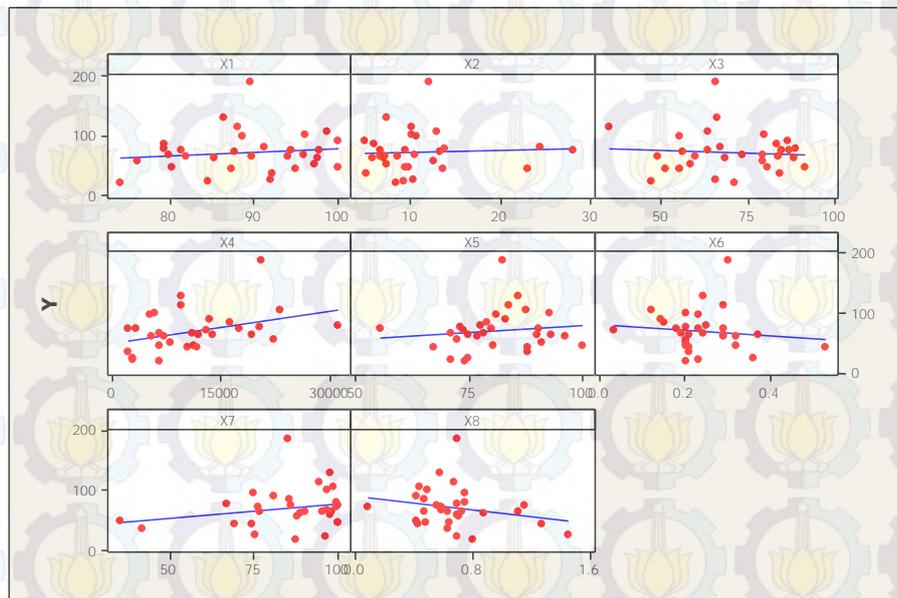
Tabel 4.1 Hasil Korelasi antara Banyaknya Kasus DBD dengan Faktor yang Berpengaruh

Nilai	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Koef.Korelasi	0.13	0.046	-0.067	0.379	0.138	-0.121	0.223	-0.238
P value	0.487	0.807	0.72	0.035	0.46	0.517	0.227	0.198

Berdasarkan Tabel 4.1 dapat diketahui bahwa variabel prediktor yang memiliki hubungan dengan variabel banyaknya kasus DBD pada alpha sebesar 0.2 adalah variabel kepadatan penduduk (X₄) dan persentase rumah sehat (X₈). Selain dengan menggunakan koefisien korelasi, pola hubungan juga dapat diketahui melalui hasil pembuatan diagram pencar yang disajikan pada Gambar 4.1

Pada Gambar 4.1 menunjukkan bahwa pada variabel kepadatan penduduk (X₄) memiliki hubungan yang positif dengan banyaknya kasus DBD di Surabaya artinya jika terdapat peningkatan kepadatan penduduk maka akan mengakibatkan peningkatan pula pada banyaknya kasus DBD yang terjadi di Surabaya. Sedangkan pada variabel persentase rumah sehat (X₈) memiliki pola hubungan yang negatif dengan banyaknya kasus DBD

artinya semakin tinggi persentase rumah sehat maka banyaknya kasus DBD akan semakin rendah.

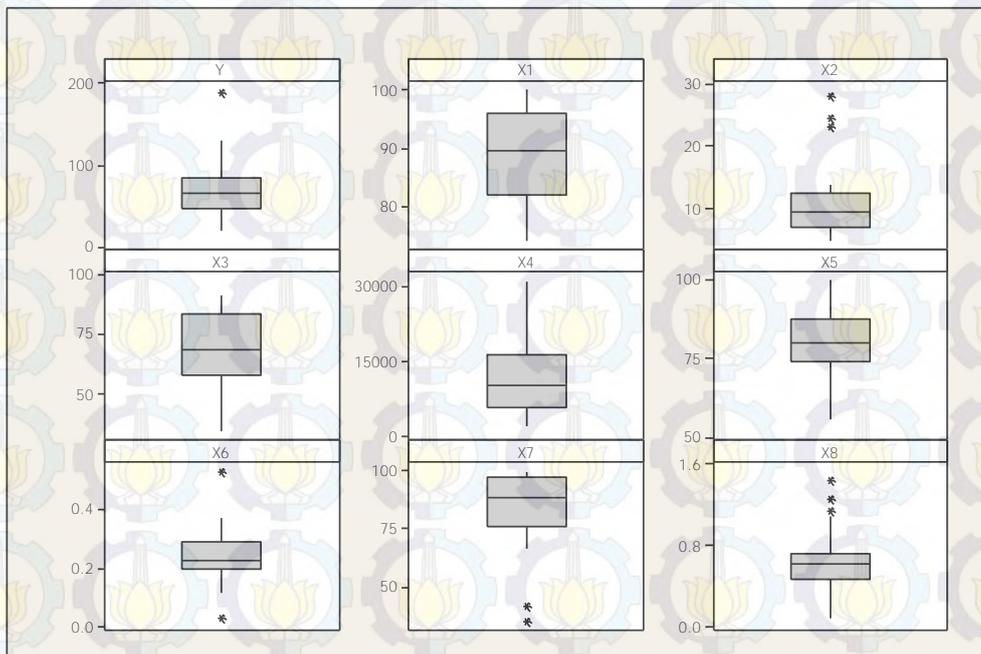


Gambar 4.1 *Scatterplot* antara Variabel Respon dengan Prediktor

Berikutnya akan dilakukan pendeteksian *outlier* pada data secara *univariat*.

Pendeteksian akan dilakukan melalui pembuatan *box-plot* yang disajikan pada Gambar 4.2

Berdasarkan Gambar 4.2 dapat diketahui bahwa hampir semua variabel memiliki pengamatan yang *outlier* kecuali variabel X_1 , X_3 , X_4 , dan X_5 . *Outlier* paling banyak terdapat pada variabel persentase rumah tangga miskin (X_2) serta persentase rumah sehat (X_8). Selain menggunakan *boxplot*, pengecekan *outlier* secara *univariat* juga akan dilakukan melalui nilai *standardize z-score* dari setiap variabel, yang hasilnya ditampilkan pada Lampiran 2 Variabel yang memiliki pengamatan *outlier* akan memiliki nilai *standardize z_score* lebih besar atau lebih kecil 2.5. Hasil *standardize z_score* dari setiap variabel menunjukkan bahwa variabel yang tidak memiliki *outlier* adalah variabel X_1 , X_3 , X_4 . Pada tahapan berikutnya, akan dilanjutkan dengan pemodelan kasus DBD dengan menggunakan tiga metode yang berbeda.



Gambar 4.2 *Boxplot* dari setiap Variabel

Tahapan pertama sebelum dilakukan pemodelan adalah melakukan pengecekan multikolinieritas antar variabel prediktor. Pada pemodelan *poisson regression* dan *negative binomial regression*, mensyaratkan tidak terjadinya kasus multikolinieritas antar variabel prediktor. Pengecekan multikolinieritas dilakukan berdasarkan nilai VIF, yang hasilnya ditampilkan pada Tabel 4.2.

Tabel 4.2 Nilai VIF pada Setiap Variabel

Variabel	VIF
X1	2.466
X2	1.891
X3	1.225
X4	1.631
X5	2.551
X6	3.573
X7	1.077
X8	3.348

Berdasarkan Tabel 4.2 semua nilai pada setiap variabel prediktor bernilai kurang dari 10, hal ini mengindikasikan tidak terjadinya kasus multikolinieritas pada data, sehingga dapat dilanjutkan pada tahapan pemodelan.

4.1 Pemodelan Kasus DBD dengan menggunakan *Poisson Regression*

Data banyaknya kasus DBD merupakan data bertipe integer positif, maka data banyaknya kasus DBD dapat diasumsikan mengikuti distribusi poisson, oleh sebab itu pada bagian ini data akan dilakukan pembentukan model dengan *poisson regression*. Namun, pada bagian pertama pembentukan model akan dilakukan pada model umum sebelum dikelompokkan menjadi *5 fold*. Prosedur pertama adalah pengujian parameter secara serentak. Pengujian secara serentak dapat diindikasikan melalui nilai devians yang di-hasilkan dari model. Nilai *devians* yang diperoleh dari model umum sebesar 313.36 dengan *p_value* sebesar 5.88×10^{-63} . Jika nilai devians yang dihasil-kan pada model umum dibandingkan dengan nilai tabel $\chi^2_{(8,0.05)}$ sebesar 15.51 ataupun alpha 5% untuk *p_value*, maka dapat diartikan bahwa minimal ada satu parameter yang berpengaruh terhadap model. Sehingga dapat dilanjutkan pada prosedur berikutnya yaitu pengujian parameter secara parsial. Hasil dari pengujian parsial ditampilkan pada Tabel 4.3.

Tabel 4.3 Hasil Estimasi Parameter *Poisson Regression* (Model Umum)

Parameter	Estimasi	Std. Error	Z hitung	<i>P value</i>
$\hat{\beta}_0$	2.89E+00	3.41E-01	8.471	< 2e-16
$\hat{\beta}_1$	1.38E-02	4.61E-03	3.003	0.00267
$\hat{\beta}_2$	-2.06E-02	5.38E-03	-3.826	0.00013
$\hat{\beta}_3$	-3.68E-03	1.59E-03	-2.318	0.02044
$\hat{\beta}_4$	3.08E-05	3.71E-06	8.309	< 2e-16
$\hat{\beta}_5$	6.08E-05	3.79E-03	0.016	0.9872
$\hat{\beta}_6$	1.45E+00	4.86E-01	2.982	0.00287
$\hat{\beta}_7$	4.75E-03	1.45E-03	3.269	0.00108
$\hat{\beta}_8$	-7.83E-01	1.67E-01	-4.69	2.73E-06

Pada Tabel 4.3 dapat diketahui bahwa hampir semua nilai $|Z_{hitung}|$ kecuali β_5 bernilai lebih besar dari $\frac{z_{0,05}}{2}$ sebesar 1.96, begitu pula dengan nilai p_value pada semua parameter kecuali β_5 bernilai lebih kecil dari alpha 5%. Artinya semua parameter kecuali β_5 signifikan terhadap model. Berikutnya analisis akan dilakukan pada model yang telah dikelompokkan menjadi 5 *fold*, namun variabel X_5 tetap akan disertakan dalam pemodelan.

Dengan prosedur yang sama, sebelumnya akan dilakukan pengecekan multikolinieritas pada data yang telah dikelompokkan menjadi 5 *fold*. Ringkasan hasilnya ditampilkan pada Lampiran 3.

Nilai VIF yang dihasilkan pada setiap variabel yang telah dikelompokkan menjadi 5 *fold* bernilai kurang dari 10, hal itu menunjukkan bahwa tidak terjadi kasus multikolinieritas dan dapat dilanjutkan pada tahapan pemodelan berikutnya yaitu pengujian parameter baik secara serentak maupun parsial. Hasil dari pengujian secara serentak yang didasarkan pada nilai devians sebesar 1253.45 dengan p_value sebesar 27035×10^{-265} , menunjukkan bahwa minimal ada satu parameter yang berpengaruh terhadap model, hal tersebut dapat diketahui melalui nilai tabel $\chi^2_{(8,0.05)}$ sebesar 15.51 yang lebih kecil daripada nilai *devians* serta p_value yang bernilai kurang dari alpha 5%. Setelah pengujian parameter secara serentak telah terpenuhi, tahapan berikutnya adalah pengujian secara parsial. Hasil estimasi serta Z-hitung dalam pengujian parsial ditampilkan pada Tabel

4.4

Tabel 4.4 Hasil Estimasi Parameter *Poisson Regression* dengan Semua Parameter

Parameter	Estimasi	Std. Error	Z-hitung	P-value
$\hat{\beta}_0$	2.89E+00	1.71E-01	16.943	< 2e-16
$\hat{\beta}_1$	1.38E-02	2.31E-03	6.006	1.90E-09
$\hat{\beta}_2$	-2.06E-02	2.69E-03	-7.652	1.97E-14
$\hat{\beta}_3$	-3.68E-03	7.95E-04	-4.636	3.55E-06
$\hat{\beta}_4$	3.08E-05	1.85E-06	16.618	< 2e-16
$\hat{\beta}_5$	6.08E-05	1.90E-03	0.032	0.974
$\hat{\beta}_6$	1.45E+00	2.43E-01	5.963	2.47E-09
$\hat{\beta}_7$	4.75E-03	7.26E-04	6.539	6.21E-11
$\hat{\beta}_8$	-7.83E-01	8.35E-02	-9.38	< 2e-16

Pada Tabel 4.4 dengan alpha 5% , dapat diketahui bahwa terdapat satu parameter yang tidak signifikan terhadap model yaitu parameter β_5 . Hal tersebut dapat diindikasikan berdasarkan nilai p_value yang lebih besar dari alpha 5% dan juga nilai $|Z_{hitung}|$ yang kurang dari $z_{\frac{0,05}{2}}$ sebesar 1.96. Maka untuk langkah selanjutnya variabel X_5 tidak disertakan dalam pemodelan. Hasil dari estimasi parameternya ditampilkan pada Tabel 4.5

Tabel 4.5 Hasil Estimasi Parameter *Poisson Regression* dengan 7 Parameter

Parameter	Estimasi	Std. Error	Z-hitung	P-value
$\hat{\beta}_0$	2.89E+00	1.70E-01	17.004	< 2e-16
$\hat{\beta}_1$	1.39E-02	1.50E-03	9.298	< 2e-16
$\hat{\beta}_2$	-2.06E-02	2.48E-03	-8.308	< 2e-16
$\hat{\beta}_3$	-3.68E-03	7.90E-04	-4.662	3.13E-06
$\hat{\beta}_4$	3.08E-05	1.81E-06	16.98	< 2e-16
$\hat{\beta}_6$	1.45E+00	2.39E-01	6.077	1.23E-09
$\hat{\beta}_7$	4.75E-03	7.25E-04	6.552	5.66E-11
$\hat{\beta}_8$	-7.83E-01	8.34E-02	-9.388	< 2e-16

Nilai devians yang dihasilkan sebesar 1253.451 dengan p_value sebesar 19471×10^{-266} . Hal tersebut mengindikasikan bahwa minimal ada satu parameter yang berpengaruh terhadap model.

Pada pengujian secara parsial dapat diketahui berdasarkan nilai p_value pada setiap parameter di Tabel 4.2, nilainya kurang dari alpha 0.05, selain itu nilai $|Z_{hitung}|$ dari setiap parameter yang dihasilkan juga bernilai lebih besar dari $z_{\frac{0,05}{2}}$ sebesar 1,96, artinya semua parameter telah signifikan terhadap model. Sehingga diperoleh model poisson regression sebagai berikut.

$$\ln(\hat{\mu}) = 2.89 + 0.0139 X_1 - 0.0206 X_2 - 0.00368 X_3 + 0.000031 X_4 + 1.45 X_6 + 0.00475 X_7 - 0.783 X_8$$

Berdasarkan model tersebut dapat dilakukan interpretasi bahwa setiap kenaikan satu persen rumah atau bangunan bebas jentik nyamuk aedes (X_1) maka rata-rata banyaknya kasus DBD akan bertambah sebesar $\exp(0.0139) = 1.01 \approx 1$ kasus dengan asumsi variabel prediktor yang lain dianggap

konstan. Jika rumah tangga miskin (X_2) naik sebesar satu persen maka rata-rata banyaknya kasus DBD akan menurun sebesar $\exp(0.0206) = 1.02 \approx 1$ kasus dengan asumsi variabel prediktor yang lain dianggap konstan. Setiap PHBS (X_3) bertambah sebesar 1 persen maka rata-rata banyaknya kasus DBD akan berkurang sebesar $\exp(0.00368) = 1.003 \approx 1$ kasus. Saat kepadatan penduduk (X_4) meningkat 1 jiwa per km^2 maka rata-rata banyaknya kasus DBD akan naik sebesar $\exp(0.000031) = 1.00003 \approx 1$ kasus dengan asumsi variabel prediktor yang lain dianggap konstan. Sedangkan saat terjadi penambahan satu persen rumah tangga yang memiliki sampah sehat (X_6), rata-rata banyaknya kasus DBD akan meningkat sebesar $\exp(1.45) = 4.263 \approx 4$ kasus dengan asumsi variabel prediktor yang lain dianggap konstan. Begitu pula dengan kenaikan rumah sehat (X_8) sebesar 1 persen akan menurunkan rata-rata banyaknya kasus DBD sebesar $\exp(0.783) = 2.18 \approx 2$ kasus, dengan asumsi variabel prediktor yang lain dianggap konstan. Setelah dilakukan interpretasi pada kelima variabel yang signifikan berpengaruh terhadap rata-rata banyaknya kasus DBD, terlihat bahwa terdapat beberapa variabel yang kurang sesuai dengan kondisi sebenarnya. Seperti pada variabel persentase rumah atau bangunan bebas jentik nyamuk aedes (X_1), persentase rumah tangga miskin (X_2) dan persentase rumah tangga yang memiliki sampah sehat (X_6), ketiga hal tersebut terlihat kontradiktif dengan kondisi yang sebenarnya, sehingga pada tahapan berikutnya dilanjutkan dengan pengecekan asumsi *equidispersi* terhadap model.

Pengecekan asumsi *equidispersi* dapat diketahui melalui nilai rasio antara devians dengan derajat bebas dari model. Pada model poisson regression dengan 7 parameter, nilai rasio antara devians dengan derajat bebas yang dihasilkan sebesar 10.81. Nilai tersebut lebih besar dari 1, sehingga dapat diindikasikan bahwa kasus overdispersi terjadi pada model poisson regression sebelumnya. Kondisi demikian dikhawatirkan berpengaruh terhadap kesalahan estimasi parameternya, yang mengakibatkan ketidaksesuaian interpretasi model dengan kondisi sebenarnya. Sehingga pada penelitian ini, penanggulangan kasus tersebut diatasi dengan memodelkan

data dengan *negative binomial regression*, sesuai dengan teori yang terdapat pada Subbab 2.10. Pembahasan mengenai *binomial negative regression* disajikan pada bagian selanjutnya.

4.2 Pemodelan Kasus DBD dengan menggunakan *Negative Binomial Regression*

Seperti pada bagian sebelumnya, pada bagian pertama pemodelan akan dilakukan pada model umum yaitu model keseluruhan tanpa pembagian data *training* maupun *testing*. Pada pengujian parameter secara serentak, hasilnya menunjukkan bahwa minimal ada satu parameter yang berpengaruh signifikan terhadap model. Hal tersebut dapat diketahui berdasarkan nilai devians yang dihasilkan yaitu sebesar 31.69 yang bernilai lebih besar jika dibandingkan dengan nilai tabel $\chi^2_{(8,0.05)}$. Begitu pula dengan *p_value* yang dihasilkan sebesar 0.000105426, yang bernilai kurang dari alpha 5%. Berikutnya dilakukan dengan pengujian parameter secara parsial, yang dilakukan dengan membandingkan nilai $|Z_{hitung}|$ ataupun *p_value* yang terdapat pada Tabel 4.6.

Tabel 4.6 Hasil Estimasi Parameter *Negatif Binomial Regression* (Model Umum)

Parameter	Estimasi	Std. Error	Z_hitung	P_value
$\hat{\beta}_0$	2.40E+00	1.08E+00	2.217	0.0267
$\hat{\beta}_1$	1.78E-02	1.47E-02	1.21	0.2263
$\hat{\beta}_2$	-1.90E-02	1.74E-02	-1.089	0.2761
$\hat{\beta}_3$	-2.50E-03	5.18E-03	-0.481	0.6303
$\hat{\beta}_4$	3.10E-05	1.23E-05	2.513	0.012
$\hat{\beta}_5$	-5.15E-04	1.17E-02	-0.044	0.965
$\hat{\beta}_6$	1.18E+00	1.49E+00	0.791	0.4292
$\hat{\beta}_7$	5.85E-03	4.49E-03	1.302	0.1928
$\hat{\beta}_8$	-7.15E-01	4.72E-01	-1.516	0.1296

Berdasarkan Tabel 4.6 dapat diketahui bahwa hanya terdapat satu parameter yang signifikan yaitu β_4 . Hal tersebut dapat diketahui melalui nilai

$|Z_{hitung}|$ dari parameter β_4 yang bernilai lebih besar dari $\frac{z_{0,05}}{2}$ yaitu 1.96 dan nilai p_value -nya yang lebih kecil dari α 5%.

Dengan prosedur yang sama, *crossvalidation 5 fold* juga akan diterapkan dalam pemodelan ini. Hasil estimasi parameternya ditampilkan pada Tabel 4.7

Tabel 4.7 Hasil Estimasi Parameter *Negatif Binomial Regression* dengan Semua Parameter

Parameter	Estimasi	Std. Error	Z-hitung	P-value
$\hat{\beta}_0$	2.40E+00	5.42E-01	4.433	9.29E-06
$\hat{\beta}_1$	1.78E-02	7.36E-03	2.42	0.01551
$\hat{\beta}_2$	-1.90E-02	8.71E-03	-2.178	0.0294
$\hat{\beta}_3$	-2.50E-03	2.59E-03	-0.963	0.33572
$\hat{\beta}_4$	3.10E-05	6.17E-06	5.026	5.01E-07
$\hat{\beta}_5$	-5.15E-04	5.86E-03	-0.088	0.93007
$\hat{\beta}_6$	1.18E+00	7.44E-01	1.581	0.11388
$\hat{\beta}_7$	5.85E-03	2.25E-03	2.605	0.00919
$\hat{\beta}_8$	-7.15E-01	2.36E-01	-3.031	0.00243

Pengujian parameter secara serentak dilakukan melalui nilai *devians* yang dihasilkan pada model yaitu 126.7973 dengan p_value sebesar 1.303×10^{-23} , maka dapat diartikan bahwa minimal ada satu parameter yang berpengaruh terhadap model. Hal tersebut ditunjukkan melalui nilai *devians* yang lebih besar dari nilai tabel $\chi^2_{(8,0.05)}$ sebesar 15,51 serta nilai p_value yang kurang dari α 5%.

Parameter yang signifikan pada pengujian parsial dapat diketahui melalui nilai $|Z_{hitung}|$ maupun p_value dari setiap parameter pada Tabel 4.7.

Pada Tabel 4.8 dapat diketahui bahwa parameter β_3 , β_5 , dan β_6 tidak signifikan terhadap model, hal tersebut dapat diketahui berdasarkan $|Z_{hitung}|$ yang bernilai lebih kecil dari $\frac{z_{0,05}}{2}$ yaitu 1.96 dan nilai p_value yang lebih besar dari α 5%. Oleh sebab itu pada tahapan berikutnya akan dilakukan pemodelan hanya dengan beberapa variabel yang signifikan, dan hasilnya ditampilkan pada Tabel 4.9. Berdasarkan hasil dari pengujian parameter secara serentak

menunjukkan bahwa minimal ada satu parameter yang berpengaruh terhadap model. Hal tersebut dapat diindikasikan melalui nilai devians yang dihasilkan yaitu sebesar 126.67 dengan p_value nya 1.210×10^{-25} . Jika dibandingkan dengan nilai tabel $\chi^2_{(5,0.05)}$ sebesar 11.07, nilai devians jauh lebih besar, begitu pula dengan p_value nya yang memiliki nilai kurang dari alpha 5%.

Tabel 4.8 Hasil Estimasi Parameter Regresi Binomial dengan 5 Parameter

Parameter	Estimasi	Std. Error	Z_hitung	P_value
$\hat{\beta}_0$	2.26E+00	5.14E-01	4.403	1.07E-05
$\hat{\beta}_1$	1.72E-02	4.90E-03	3.515	0.000439
$\hat{\beta}_2$	-1.74E-02	7.65E-03	-2.273	0.023024
$\hat{\beta}_4$	3.06E-05	6.01E-06	5.084	3.69E-07
$\hat{\beta}_7$	6.31E-03	2.29E-03	2.761	0.00577
$\hat{\beta}_8$	-3.97E-01	1.32E-01	-3	0.002702

Berdasarkan hasil pengujian parsial dengan $|Z_{hitung}|$ dan p_value yang ditampilkan pada Tabel 4.10, hasilnya juga menunjukkan bahwa nilai $|Z_{hitung}|$ yang dihasilkan bernilai lebih besar dari $\frac{z_{0,05}}{2}$ yaitu 1,96, begitu pula dengan p_value nya yang kurang dari alpha 5%. Sehingga model yang dapat dibentuk berdasarkan Tabel 4.9 adalah sebagai berikut.

$$\ln(\hat{\mu}) = 2.26 + 0.0172 X_1 - 0.0174 X_2 + 0.0000306 X_4 + 0,00631 X_7 - 0,397 X_8$$

Dari model diatas dapat diartikan kenaikan satu persen rumah atau bangunan bebas jentik nyamuk aedes (X_1) maka rata-rata banyaknya kasus DBD akan bertambah sebesar $\exp(0.0172) = 1.01 \approx 1$ kasus dengan asumsi variabel prediktor yang lain dianggap konstan. Jika rumah tangga miskin (X_2) naik sebesar satu persen maka rata-rata banyaknya kasus DBD akan menurun sebesar $\exp(0.0174) = 1.02 \approx 1$ kasus dengan asumsi variabel prediktor yang lain dianggap konstan jika kepadatan penduduk di Surabaya (X_4) meningkat 1 jiwa per km^2 maka rata-rata banyaknya kasus DBD akan naik sebesar \exp

$(0.0000306) = 1.000031 \approx 1$ kasus dengan asumsi variabel prediktor yang lain dianggap konstan. Hal ini sesuai dengan pernyataan dari Kementerian Kesehatan RI karena jika ditinjau dari cara penularannya, DBD ditularkan melalui nyamuk *Aedes Aegypti* yang sangat menyukai tempat yang padat, sehingga semakin padat penduduk, peluang lingkungan menjadi kumuh akan lebih besar yang berdampak pula terhadap peningkatan penderita DBD di Surabaya. Begitu pula dengan kenaikan rumah sehat (X_8) sebesar 1 persen akan menurunkan rata-rata banyaknya kasus DBD sebesar $\exp(0.397) = 1.48 \approx 1$ kasus, dengan asumsi variabel prediktor yang lain dianggap konstan.

Prosedur terakhir dalam pemodelan dengan *binomial negative regression* adalah pengecekan terjadi atau tidaknya kasus overdispersi yang diindikasikan melalui nilai rasio antara devians dengan derajat bebas dari model *negative binomial regression*. Nilai yang dihasilkan sebesar 1.07. Nilai ini mendekati 1. Jika dibandingkan dengan nilai rasio yang dihasilkan oleh *poisson regression*, nilai rasio pada *binomial negative regression* jauh lebih kecil dibandingkan dengan nilai rasio dispersi *poisson regression*, maka dapat diartikan bahwa model *binomial negatif regression* dapat mengatasi overdispersi pada *poisson regression*, namun kasus over ataupun under estimate belum dapat diatasi. Hal tersebut juga diindikasikan disebabkan tidak terdapatnya hubungan antara variabel prediktor dengan respon pada beberapa variabel prediktor yang dipakai pada penelitian ini.

4.3 Pemodelan Kasus DBD dengan menggunakan *Neural Networks*

Pada bagian ketiga data banyaknya kasus DBD di Surabaya akan dilakukan pemodelan dengan menggunakan metode *Neural Networks* (NN). Pemodelan dengan NN bersifat lebih fleksibel, tidak mensyaratkan asumsi multikolinieritas maupun *equidispersi* seperti pada kedua pemodelan sebelumnya. NN merupakan salah satu sistem pengolahan informasi yang memiliki 3 lapisan yaitu *input*, *hidden*, serta *output*. Hasil pengolahan ataupun *output* diperoleh melalui proses *learning*. Pada penelitian ini, metode *learning* yang dipakai adalah *backpropagation* dengan arsitektur jaringan *Multi Layer Perceptron* (MLP). *Training setting* dan fungsi aktivasi yang

dipakai telah dijelaskan pada Bab 3. Seperti pada bagian sebelumnya, metode *crossvalidation k-fold* juga akan dipakai untuk melakukan proses estimasi dan pemilihan model terbaik.

Salah satu tujuan dari pemodelan NN adalah mendapatkan taksiran parameter. Parameter dalam NN adalah bobot. Bobot pada NN meliputi bobot *input* ke *hidden*, dan bobot *hidden* ke *output*. Sedangkan bobot pada bias terdiri dari bobot bias ke *hidden* dan bobot bias ke *output*. Tahapan pertama adalah melakukan penentuan inisialisasi bobot awal untuk mendapatkan parameter dari model. Pada penelitian ini, bobot awal akan di-*set* dengan suatu nilai tertentu, yang diperoleh melalui iterasi secara *random* pada setiap kemungkinan jumlah *node* pada *hidden layer* (2 sampai 20) sebanyak 100 kali iterasi. Bobot awal yang dipilih adalah bobot yang menghasilkan kriteria MSE paling minimum.

Pada bagian selanjutnya bobot awal optimum yang dipakai adalah bobot dengan nilai penaksir yang telah dibulatkan, hal tersebut dikarenakan untuk penyesuaian dengan kedua metode sebelumnya. Bobot awal optimum diperoleh pada iterasi ke-35, dengan jumlah *node* pada *hidden layer* sebanyak 12. Nilai MSE *training* terbaik pada setiap *node hidden layer* ditampilkan pada Lampiran 4. Sehingga arsitektur jaringan yang terbentuk adalah MLP (8-12-1). Nilai-nilai bobot awal optimum pada setiap lapisan untuk jaringan MLP(8-12-1) ditampilkan pada Tabel 4.9.

Tabel 4.9 Bobot Awal MLP (8-12-1)

Bobot Awal	X₁	X₂	X₃	X₄	X₅	X₆	X₇	X₈
v₁	-1.68	0.38	-1.91	0.28	-0.92	-1.64	1.70	-1.22
v₂	2.05	0.10	0.37	0.75	-1.04	-2.16	1.10	-1.65
v₃	-0.04	-0.24	1.93	-1.95	0.45	1.60	-2.06	0.16
v₄	0.94	1.77	-1.11	0.66	-0.39	-1.71	1.78	1.63
Hidden- Input v₅	-0.99	-1.67	-2.13	0.95	-1.65	0.87	1.25	0.61
v₆	-2.08	0.25	-1.32	1.82	-0.45	-0.19	-2.06	0.80
v₇	0.64	0.70	0.93	1.81	-1.91	1.67	0.99	-1.47
v₈	-0.69	-2.29	1.13	-1.57	1.62	-0.69	-0.07	-1.43
v₉	0.56	0.53	0.49	1.68	-1.31	-1.78	-1.74	-1.73
v₁₀	1.69	1.65	-0.16	-1.47	0.88	-1.37	-1.88	0.80

Tabel 4.9 Bobot Awal MLP (8-12-1) (Lanjutan)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
v_{11}	-2.02	-0.53	-0.94	-1.43	1.26	1.07	-2.14	0.04
v_{12}	0.47	-0.33	0.75	-2.60	-2.56	0.26	0.57	-0.08
Bias Hidden- Input	b_1	b_2	b_3	b_4	b_5	b_6		
	3.82	-3.13	2.43	-1.74	1.04	0.35		
	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}		
	0.35	-1.04	1.74	2.43	-3.13	3.82		
Hidden- Output	w_1	w_2	w_3	w_4	w_5	w_6		
	0.72	0.93	0.52	-0.51	0.35	0.52		
	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}		
	0.17	0.64	0.85	0.28	-0.96	-0.63		
Bias Hidden- Output	b_k							
	-0.95							

Setelah dilakukan tahap *learning* dengan inisialisasi bobot di atas, maka dapat dibentuk suatu model NN dengan bobot akhir yang diperoleh dari hasil *learning*. Model MLP(8-12-1) adalah sebagai berikut.

$$Y = b_k + w_{11}a_1 + w_{12}a_2 + \dots + w_{112}a_{12}$$

Dengan

$$a_1 = f_1(v_1) = \frac{1}{1 + \exp(-v_1)}$$

$$a_2 = f_2(v_2) = \frac{1}{1 + \exp(-v_2)}$$

⋮

$$a_{12} = f_{12}(v_{12}) = \frac{1}{1 + \exp(-v_{12})}$$

Dimana:

$$v_1 = b_1 + v_{11}X_1 + v_{12}X_2 + v_{13}X_3 + \dots + v_{18}X_8$$

$$v_2 = b_2 + v_{21}X_1 + v_{22}X_2 + v_{23}X_3 + \dots + v_{28}X_8$$

⋮

⋮

$$v_{12} = b_{12} + v_{121}X_1 + v_{122}X_2 + v_{123}X_3 + \dots + v_{128}X_8$$

Bobot akhir yang dihasilkan terdiri dari 4 bobot, yaitu bobot dari input ke hidden (bobot akhir hidden input), bobot bias input ke hidden layer (bobot akhir bias hidden input), bobot dari hidden ke output layer (bobot akhir hidden output), serta bobot bias hidden layer ke output (bobot akhir bias hidden output). *Output* selengkapnya ditampilkan pada Lampiran 5. Keempat bobot pada MLP (8-12-1) ditampilkan pada Tabel 4.10.

Tabel 4.10 Bobot Akhir MLP (8-12-1)

Bobot Akhir		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Hidden-Input	v_1	-0.42	-0.70	0.84	-1.45	-2.93	-4.18	3.75	-0.06
	v_2	2.04	1.07	0.38	0.49	-1.87	-0.37	0.88	-1.59
	v_3	0.66	2.55	-1.10	0.11	6.04	0.70	-1.27	-3.32
	v_4	1.03	0.57	-0.89	-0.59	-4.05	-2.57	1.91	1.48
	v_5	-1.69	-1.42	-2.89	0.93	-2.75	0.98	-1.09	1.77
	v_6	-1.78	-0.23	0.32	1.07	3.12	-2.32	-2.12	1.62
	v_7	0.79	1.71	-0.38	2.22	-1.89	0.96	0.71	-2.58
	v_8	-1.12	1.15	2.90	1.59	4.89	-0.63	-1.64	-1.50
	v_9	2.74	1.03	-1.85	-0.05	-4.10	-1.58	-1.18	-3.27
	v_{10}	2.45	1.70	0.57	-1.24	1.05	-3.11	-2.02	-1.99
	v_{11}	-0.44	0.25	0.72	-2.45	1.17	1.43	-5.31	-3.00
	v_{12}	1.26	-0.54	1.13	-1.77	-2.24	0.17	1.57	0.29
Bias Hidden-Input	b_1	b_2	b_3	b_4	b_5	b_6			
	4.32	-2.66	5.54	-1.16	-3.41	-1.07			
	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}			
	-0.24	-1.51	2.89	3.99	-0.54	4.11			
	w_1	w_2	w_3	w_4	w_5	w_6			
	0.96	-0.61	-1.04	-2.61	-0.89	0.14			
Hidden-Output	w_7	w_8	w_9	w_7	w_8	w_9			
	1.04	-1.50	1.09	2.09	-1.09	-0.21			
Bias Hidden-Output	b_k								
	-1.28								

Sehingga model NN setelah dilakukan substitusi dengan bobot akhir adalah sebagai berikut.

$$Y = -1.28 + 0.96a_1 - 0.61a_2 - 1.04a_3 - 2.61a_4 - 0.89a_5 + 0.14a_6 + 1.04a_7 - 1.50a_8 + 1.09a_9 + 2.09a_{10} - 1.09a_{11} - 0.21a_{12}$$

Dimana:

$$a_1 = f_1(v_1) = \frac{1}{1 + \exp(-v_1)} \text{ dengan}$$

$$v_1 = 4.32 - 0.42X_1 - 0.70X_2 + 0.84X_3 - 1.45X_4 - 2.93X_5 - 4.18X_6 + 3.75X_7 - 0.06X_8$$

$$a_2 = f_2(v_2) = \frac{1}{1 + \exp(-v_2)} \text{ dengan}$$

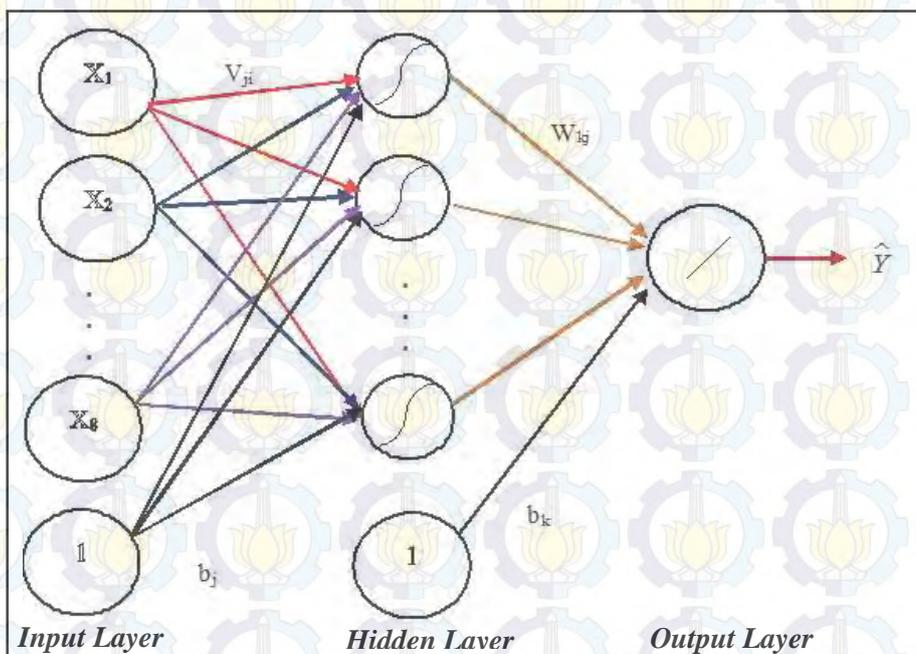
$$v_2 = -2.66 + 2.04X_1 + 1.07X_2 + 0.38X_3 + 0.49X_4 - 1.87X_5 - 0.37X_6 + 0.88X_7 - 1.59X_8$$

⋮

$$a_{12} = f_{12}(v_{12}) = \frac{1}{1 + \exp(-v_{12})} \text{ dengan}$$

$$v_{12} = 4.11 + 1.26X_1 - 0.54X_2 + 1.13X_3 - 1.77X_4 - 2.24X_5 + 0.17X_6 + 1.57X_7 + 0.29X_8$$

Arsitektur jaringan dari model MLP(8-12-1) ditampilkan pada Gambar 4.3.



Gambar 4.3 Arsitektur Jaringan MLP (8-12-1)

4.4 Penentuan Model Terbaik untuk Memprediksi Kasus DBD

Pada bagian terakhir, ketiga model yang telah dibentuk sebelumnya akan dilakukan perbandingan untuk menentukan model yang terbaik dalam memprediksi kasus DBD di Surabaya. Perbandingan dilakukan berdasarkan nilai CV, yang diperoleh dari penjumlahan MSE testing dikalikan dengan proporsi data pada masing-masing *fold*. Ringkasannya disajikan pada Tabel 4.11

Tabel 4.1 Ringkasan nilai CV pada Ketiga Metode

Partisi	nk/n	<i>Poisson</i>		<i>Negatife Binomial</i>		<i>NN</i>	
		MSE	(nk/n)*MSE	MSE	(nk/n)*MSE	MSE	(nk/n)*MSE
1	0.19	472.5	91.45	459.83	89	48.77	9.44
2	0.19	628.5	121.65	697.17	134.94	48.34	9.36
3	0.19	375.66	72.71	345.17	66.81	84.36	16.33
4	0.19	2187.33	423.35	2717	525.87	24.83	4.81
5	0.23	343.42	77.55	273	61.65	26.18	5.91
CV		786.71		878.26		45.84	

Berdasarkan Tabel 4.12 dapat diketahui bahwa dari ketiga model, *Neural Network* merupakan model terbaik yang dapat dipakai untuk memprediksi banyaknya kasus DBD di Surabaya. Hal tersebut dapat diindikasikan berdasarkan nilai CV yang terkecil. Nilai MSE pada model NN merupakan rata-rata dari MSE pada 100 data yang diperoleh dari hasil *running* sebelumnya.

BAB 5

KESIMPULAN DAN SARAN

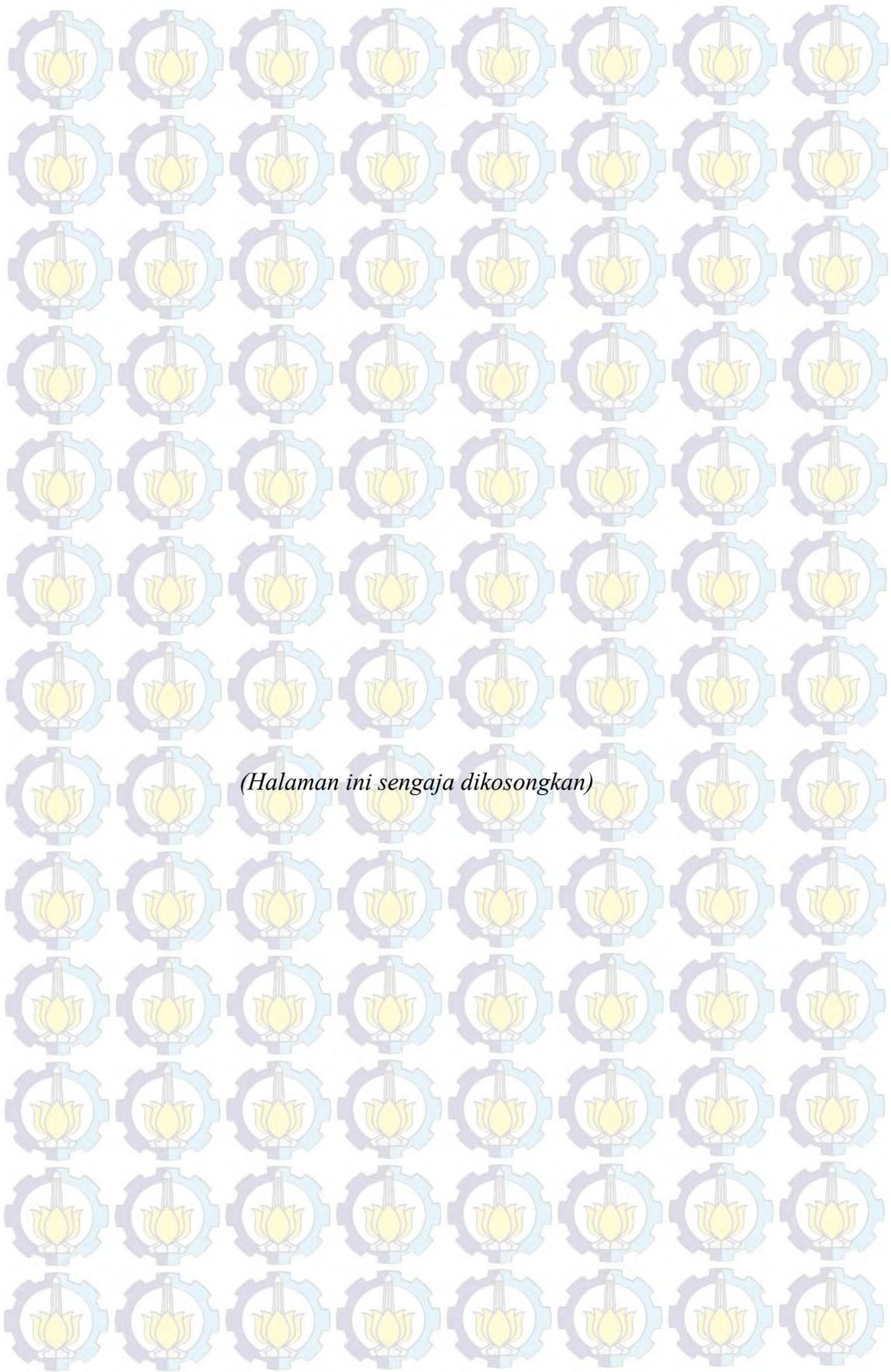
5.1 Kesimpulan

Setelah dilakukan analisis dan penjelasan dari pengolahan data pada bagian sebelumnya. Kesimpulan dari setiap *subbab* yang telah dibahas adalah sebagai berikut.

1. Semua variabel signifikan pada model *poisson regression* kecuali variabel rasio tenaga medis, namun pada model yang telah dihasilkan terjadi kasus *over-dispersi*.
2. Variabel yang signifikan pada model *negative binomial regression* adalah persentase bangunan bebas jentik nyamuk, persentase RT miskin, kepadatan penduduk, rasio saran kesehatan puskesmas, dan persentase rumah sehat Pada model ini kasus overdispersi dapat diatasi
3. Model *Neural Network* terbaik yang dipakai untuk memodelkan banyaknya kasus DBD yaitu model dengan arsitektur jaringan MLP (8-12-1).
4. Model terbaik yang dipakai untuk memprediksi kasus DBD di Surabaya adalah model *neural network*

5.2 Saran

Pada penelitian berikutnya dapat ditambahkan variabel-variabel lain yang juga berpengaruh terhadap banyaknya kasus DBD. Metode analisis yang dipakai untuk penyelesaian suatu kasus juga dapat dilakukan pengembangan, seperti pengembangan dari gabungan metode *Neural Network* dengan *Poisson Regression*.



(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1. Data Jumlah Kasus DBD di Kota Surabaya beserta Faktor-Faktor yang Mempengaruhinya sebelum Dilakukan *K-Fold Crossvalidation*

No	Y	X1	X2	X3	X4	X5	X6	X7	X8
1	68	79.87	6.80	79.03	11102.49	71.03	0.24	97.03	0.59
2	130	86.37	7.39	65.87	9484.19	85.72	0.24	97.66	0.57
3	22	84.63	9.19	46.65	2816.77	70.97	0.23	96.51	0.69
4	75	81.40	6.68	84.70	2322.49	55.43	0.23	100.00	0.54
5	35	92.25	5.07	84.40	2189.61	87.86	0.21	41.54	0.62
6	25	92.09	10.28	65.43	2934.50	74.68	0.36	75.17	1.45
7	75	97.76	9.40	87.06	3326.19	90.40	0.29	86.11	1.15
8	43	95.05	13.56	55.13	11735.890	87.90	0.53	74.55	1.27
9	78	79.24	13.74	88.61	20324.710	72.93	0.20	66.98	0.69
10	57	76.19	12.52	79.33	22287.820	72.14	0.20	87.93	0.70
11	81	91.20	24.41	66.76	31193.820	77.29	0.25	99.64	0.74
12	44	87.43	23.03	51.07	10397.940	67.00	0.21	69.51	0.42
13	74	94.50	28.00	63.41	17605.940	80.01	0.18	76.29	0.58
14	72	87.67	13.17	55.77	13025.900	73.53	0.03	99.83	0.08
15	19	74.15	8.44	71.05	6534.830	74.15	0.20	87.56	0.79
16	64	81.93	8.57	59.65	19310.670	74.62	0.29	90.35	0.72
17	107	98.78	12.87	63.10	23203.670	87.30	0.12	98.62	0.43
18	85	79.23	5.83	83.22	16334.420	78.84	0.15	85.80	0.46
19	102	96.14	10.13	79.88	5850.570	92.72	0.20	96.98	0.49
20	91	100.00	4.87	86.75	13370.650	82.90	0.14	81.05	0.41
21	46	100.00	9.75	91.61	6515.860	100.00	0.20	100.00	0.47
22	97	88.58	10.66	54.90	5153.690	80.80	0.23	74.75	0.74
23	65	89.87	7.17	83.69	13923.670	89.87	0.21	95.39	0.62
24	188	89.53	11.98	65.37	20516.060	82.19	0.30	85.06	0.69
25	114	88.11	10.12	34.26	9548.420	83.65	0.29	94.38	0.66
26	50	97.27	7.26	58.03	7996.530	90.94	0.20	35.35	0.41
27	66	95.96	10.41	73.30	6583.600	78.00	0.19	76.95	0.46
28	62	85.36	6.98	68.38	5557.620	76.71	0.32	88.83	0.87
29	61	97.64	5.79	88.33	7168.530	96.00	0.29	97.98	0.69
30	65	94.07	6.53	48.47	12059.590	92.85	0.37	98.10	1.10
31	46	80.29	9.35	80.78	11286.160	80.28	0.32	100.00	0.63

Keterangan :

y = Jumlah Kasus DBD kota Surabaya tahun 2013

x_1 = Persentase rumah / bangunan bebas jentik nyamuk Aedes

x_2 = Persentase rumah tangga miskin

x_3 = Persentase rumah tangga yang berperilaku hidup bersih dan sehat (PHBS)

x_4 = Kepadatan penduduk

x_5 = Rasio tenaga medis (dokter umum)

x_6 = Persentase rumah tangga yang memiliki tempat sampah sehat

x_7 = Rasio sarana kesehatan Puskesmas

x_8 = Persentase rumah sehat

Lampiran 2. Hasil *Standardize Data* untuk Pendeteksian *Outlier*

Y	X1	X2	X3	X4	X5	X6	X7	X8
-1.53	-2.01	-0.40	0.08	-0.67	-0.71	-0.44	0.12	0.44
-1.44	-0.60	-0.27	-1.56	-1.19	-1.04	-0.11	0.67	0.08
-1.35	0.40	-0.07	-0.29	-1.17	-0.66	1.36	-0.65	2.84
-1.06	0.42	-1.02	0.98	-1.28	0.73	-0.33	-2.73	-0.18
-0.83	0.80	0.53	-0.99	0.05	0.73	3.27	-0.69	2.19
-0.80	-0.23	2.26	-1.26	-0.13	-1.46	-0.33	-1.00	-0.91
-0.74	-1.19	-0.24	0.74	-0.01	-0.07	0.91	0.89	-0.14
-0.74	1.46	-0.16	1.47	-0.67	2.00	-0.44	0.89	-0.72
-0.62	1.10	-0.62	-0.79	-0.47	1.05	-0.44	-3.11	-0.94
-0.42	-1.74	0.34	0.64	1.53	-0.92	-0.44	0.14	0.11
-0.30	1.15	-0.89	1.25	-0.58	1.58	0.57	0.76	0.08
-0.27	-0.50	-0.67	-0.10	-0.81	-0.44	0.91	0.20	0.73
-0.21	-0.97	-0.38	-0.68	1.11	-0.66	0.57	0.29	0.19
-0.18	0.10	-0.63	0.94	0.36	0.94	-0.33	0.60	-0.18
-0.18	0.67	-0.75	-1.44	0.10	1.25	1.47	0.77	1.57
-0.15	0.92	-0.04	0.24	-0.66	-0.31	-0.56	-0.54	-0.76
-0.09	-1.24	-0.70	0.62	-0.03	-1.04	0.01	0.70	-0.29
0.02	-0.19	0.46	-0.95	0.23	-0.78	-2.35	0.88	-2.14
0.08	0.72	3.17	-0.43	0.87	-0.10	-0.67	-0.58	-0.32
0.11	-1.04	-0.72	1.00	-1.26	-2.68	-0.11	0.89	-0.47
0.11	1.16	-0.23	1.16	-1.12	0.99	0.57	0.03	1.75
0.20	-1.33	0.57	1.27	1.25	-0.84	-0.44	-1.16	0.08
0.29	0.28	2.51	-0.21	2.77	-0.38	0.12	0.86	0.26
0.40	-1.33	-0.88	0.90	0.70	-0.22	-1.00	0.01	-0.76
0.58	1.46	-1.05	1.14	0.28	0.21	-1.12	-0.29	-0.94
0.76	-0.07	0.00	-1.00	-0.86	-0.01	-0.11	-0.68	0.26
0.90	0.94	-0.09	0.68	-0.77	1.24	-0.44	0.70	-0.65
1.05	1.30	0.41	-0.45	1.65	0.67	-1.34	0.80	-0.87
1.25	-0.13	-0.10	-2.39	-0.25	0.29	0.57	0.54	-0.03
1.72	-0.37	-0.59	-0.27	-0.26	0.50	0.01	0.74	-0.36
3.42	0.06	0.24	-0.30	1.28	0.13	0.68	-0.04	0.08

Lampiran 3. Nilai VIF pada *K-Fold*

Predictor	Coef	SE Coef	T	P	VIF
Constant	-31.55	36.81	-0.86	0.393	
x1	1.1861	0.5000	2.37	0.019	2.466
x2	-1.5603	0.5947	-2.62	0.010	1.891
x3	-0.2861	0.1767	-1.62	0.108	1.225
x4	0.0022609	0.0004221	5.36	0.000	1.631
x5	-0.1181	0.3973	-0.30	0.767	2.551
x6	62.27	50.37	1.24	0.218	3.573
x7	0.3385	0.1521	2.23	0.028	1.077
x8	-39.43	15.83	-2.49	0.014	3.348

Lampiran 4. Nilai MSE *training* Terbaik pada setiap *Node*

Node	MSE
2	87.8
4	5.09
6	8.7
8	9.09
10	3.45
12	0.13
14	2.06
16	10.67
18	4.64
20	3.9

Lampiran 5. Output MLP (8-12-1)

i =

35

ntr =

124

iu =

8

M =

121

bobot_input_hidden_train =

-1.6804	0.3782	-1.9057	0.2847	-0.9153	-1.6423	1.6972	-1.2236
2.0540	0.1003	0.3743	0.7531	-1.0374	-2.1589	1.0986	-1.6467
-0.0393	-0.2355	1.9284	-1.9462	0.4485	1.6016	-2.0585	0.1565
0.9394	1.7662	-1.1092	0.6586	-0.3915	-1.7138	1.7842	1.6285
-0.9919	-1.6667	-2.1268	0.9460	-1.6453	0.8719	1.2524	0.6133
-2.0842	0.2515	-1.3211	1.8210	-0.4538	-0.1861	-2.0582	0.8040
0.6379	0.6978	0.9260	1.8141	-1.9057	-1.6717	0.9882	-1.4652
-0.6936	-2.2881	1.1318	-1.5676	1.6182	-0.6894	-0.0706	-1.4276
0.5575	0.5330	0.4908	1.6828	-1.3120	-1.7781	-1.7429	-1.7330
1.6893	1.6511	-0.1552	-1.4666	0.8767	-1.3740	-1.8826	0.7977
-2.0163	-0.5304	-0.9389	-1.4279	1.2649	1.0668	-2.1413	0.0367
0.4665	-0.3310	0.7474	-2.5964	-2.5628	0.2584	0.5696	-0.0750

bobot_bias_hidden_train =

3.8199
-3.1254
2.4309
-1.7363
1.0418
0.3473
0.3473
-1.0418

```

1.7363
2.4309
-3.1254
3.8199
bobot_hidden_output_train =
Columns 1 through 9
0.7234 0.9310 0.5194 -0.5066 0.3525 0.5186 0.1724 0.6449 0.8537
Columns 10 through 12
0.2751 -0.9621 -0.6313
bobot_bias_output_train =
-0.9465
atrain =
Columns 1 through 9
84.9409 35.0080 45.9369 63.9359 66.0083 96.9835 64.9922 107.0028
49.9848
Columns 10 through 18
74.9644 71.9874 24.9675 113.9710 64.9806 129.8984 187.9505 46.0034
56.9718
Columns 19 through 27
22.0133 43.9917 67.9426 42.9765 61.9365 61.0115 91.0219 74.0093 77.9994
Columns 28 through 36
100.3393 81.0421 74.9826 18.9836 64.9922 107.0028 49.9848 74.9644
71.9874
Columns 37 through 45
24.9675 113.9710 64.9806 129.8984 187.9505 46.0034 56.9718 22.0133
43.9917

```

Columns 46 through 54

67.9426 42.9765 61.9365 61.0115 91.0219 74.0093 77.9994 100.3393
81.0421

Columns 55 through 63

74.9826 18.9836 84.9409 35.0080 45.9369 63.9359 66.0083 96.9835
113.9710

Columns 64 through 72

64.9806 129.8984 187.9505 46.0034 56.9718 22.0133 43.9917 67.9426
42.9765

Columns 73 through 81

61.9365 61.0115 91.0219 74.0093 77.9994 100.3393 81.0421 74.9826
18.9836

Columns 82 through 90

84.9409 35.0080 45.9369 63.9359 66.0083 96.9835 64.9922 107.0028
49.9848

Columns 91 through 99

74.9644 71.9874 24.9675 22.0133 43.9917 67.9426 42.9765 61.9365 61.0115

Columns 100 through 108

91.0219 74.0093 77.9994 100.3393 81.0421 74.9826 18.9836 84.9409
35.0080

Columns 109 through 117

45.9369 63.9359 66.0083 96.9835 64.9922 107.0028 49.9848 74.9644
71.9874

Columns 118 through 124

24.9675 113.9710 64.9806 129.8984 187.9505 46.0034 56.9718

e =

Columns 1 through 16

2.7440	1.0300	-1.8490	-0.0484	-4.0989	-1.5783	-1.1755	-3.2720								
2.4476	1.7039	0.5658	-1.2437	1.0531	-3.1139	-2.0249	-1.9933								
-0.4413	0.2456	0.7153	-2.4486	1.1748	1.4272	-5.3055	-2.9981								
1.2627	-0.5409	1.1277	-1.7731	-2.2399	0.1709	1.5681	0.2859								
bobotakhir_bias_hidden_train															
4.3249															
-2.6622															
5.5441															
-1.1596															
-3.4062															
-1.0729															
-0.2378															
-1.5069															
2.8894															
3.9892															
-0.5410															
4.1103															
bobotakhir_hidden_output_train															
Columns 1 through 9															
0.9614	-0.6053	-1.0387	-2.6104	-0.8855	0.1404	1.0431	-1.5016	1.0919							
Columns 10 through 12															
2.0932	-1.0906	-0.2094													
bobotakhir_bias_output_train															
-1.2839															
artes =															
Columns 1 through 16															
74	78	100	81	75	19	85	35	46	64	66	97	65	107	50	75
Columns 17 through 31															
72	25	114	65	130	188	46	57	22	44	68	43	62	61	91	
r	0.9999														
r2	0.9999														
MSE_ts	0.12														

Lampiran 6 . Syntax *Poisson Regression* dengan *Software R*

```
#FULL MODEL (Data Asli)
data <- read.csv("D:new2.csv", sep=",")
library(MASS)
m1=glm(y~x1+x2+x3+x4+x5+x6+x7+x8,family=poisson,data=data)
summary(m1)
with (m1, cbind(res.deviance=deviance, df=8,
p=pchisq(deviance, 8, lower.tail=FALSE)))
pre <- predict(m1)

# exclude x3,x5,x7
m2 <- update(m1, . ~ . - x3,x5,x7)
m2=glm(y~x1+x2+x4+x6+x8,family=poisson,data=data)

summary(m2)
with (m2, cbind(res.deviance=deviance, df=5,
p=pchisq(deviance, 5, lower.tail=FALSE)))
pre <- predict(m2)
BIC(m2)
```

Lampiran 7 . Syntax *Negative Binomial Regression* dengan *Software R*

```
#FULL MODEL (Data Asli)
data <- read.csv("D:new2.csv", sep=",")
library(MASS)
m1<-glm.nb(y~x1+x2+x3+x4+x5+x6+x7+x8,data=data)
summary(m1)
with (m1, cbind(res.deviance=deviance, df=8,
p=pchisq(deviance, 8, lower.tail=FALSE)))
preM1 <- predict(m1)
# exclude
m2 <- update(m1, . ~ . - x1,x2,x3,x5,x6,x7,x8)
m2=glm.nb(y~x4,data=data)
summary(m2)
with (m2, cbind(res.deviance=deviance, df=1,
p=pchisq(deviance, 1, lower.tail=FALSE)))
preM2 <- predict(m2)
BIC(m2)
```

Lampiran 8 . Syntax *Neural Networks* dengan *Matlab*

```
function dbd4;
clc
data=xlsread('new2.xls');
y=data(:,2);
x=data(:,3:10);
p=x';
t=y';
ptr=p(:,1:25)
ttr=t(:,1:25)
ptes=p(:,26:31)
ttes=t(:,26:31)
h=20;
i=0;
MSE_tropt = 10000;
i_opt_tr = 0;
for i = 0:49;
    i=i+1
    ntr=size((ttr),2) %jml observasi data training (2: mengambil ukuran pada kolom ke-2)
    iu=size((ptr),1) % jmlh input unit
    M=((iu+1)*h)+(h+1) %jumlah parameter
    ntes=size((ttes),2); %jml observasi data testing
    net=newff(ptr,ttr,h,{'logsig','purelin'},'trainlm');
    net.IW{1,1}; %bobot input
    net.b{1,1}; %bobot bias input
    net.LW{2,1}; %bobot hidden
    net.b{2,1}; %bobot bias input
    bobot_input_hidden_train=net.IW{1,1} %bobot input
    bobot_bias_hidden_train=net.b{1,1} %bobot bias input
    bobot_hidden_output_train=net.LW{2,1} %bobot hidden
    bobot_bias_output_train=net.b{2,1} %bobot bias input
    net.trainParam.epochs=1000;
    net.trainParam.goal=0.01;
    net.trainParam.lr=0.1;
    net.trainParam.show=50;
    net.divideFcn="";
    [net,tr]=train(net,ptr,ttr); %proses learning
    bobotakhir_input_hidden_train=net.IW{1,1}; %bobot input akhir
    bobotakhir_bias_hidden_train=net.b{1,1}; %bobot bias input akhir
    bobotakhir_hidden_output_train=net.LW{2,1}; %bobot hidden akhir
    bobotakhir_bias_output_train=net.b{2,1}; %bobot bias input akhir
    %Untuk mendapatkan prediksi data training --
    atrain=sim(net,ptr) %hasil simulasi
    atround=round(atrain)
    e=ttr-atround;
    [m,b,r]=postreg(atround,ttr);
    disp 'r'
    disp(r);
    r2=r*r;
    disp 'r2'
    disp(r2);
```

```

MSE_train=mse(e);
disp 'MSE_train'
disp (MSE_train);
AIC=(ntr*log(MSE_train))+2*M;
BIC=(ntr*(log(MSE_train)))+(M*log(ntr));
disp 'AIC'
disp(AIC);
disp 'BIC'
disp (BIC);
disp 'bobotakhir_input_hidden_train'
disp (bobotakhir_input_hidden_train)
disp 'bobotakhir_bias_hidden_train'
disp (bobotakhir_bias_hidden_train)
disp 'bobotakhir_hidden_output_train'
disp (bobotakhir_hidden_output_train)
disp 'bobotakhir_bias_output_train'
disp (bobotakhir_bias_output_train)
if (MSE_train <= MSE_tropt)
    MSE_tropt = MSE_train;
    i_opt_tr = i;
    IW_opt = net.IW{1,1};    %bobot input
    b1_opt = net.b{1,1};    %bobot bias input
    LW_opt = net.LW{2,1};    %bobot hidden
    b2_opt =net.b{2,1};    %bobot bias input
    %data testing
    ates=sim(net,ptes);
    artes=round(ates)
    etes=ttes-artes;
    [m,b,r]=postreg(artes,ttes);
    disp 'r'
    disp(r);
    r2=r*r;
    disp 'r2'
    disp(r2);
    MSE_ts=mse(etes);
    disp 'MSE_ts'
    disp(MSE_ts);
end;
end;
disp ('-----Hasil Terbaik-----')
disp 'MSE_tropt'
disp (MSE_tropt);
disp 'i_opt_tr'
disp(i_opt_tr);
disp 'IW_opt'
disp (IW_opt);
disp 'b1_opt'
disp (b1_opt);
disp 'LW_opt'
disp (LW_opt);
disp 'b2_opt'
disp (b2_opt);

```

DAFTAR PUSTAKA

Agresti, A. (2007), *An Introduction to Categorical Data Analysis*, 2nd edition, John Wiley & Sons, Inc., New York.

Anders, U., dan Korn, O. (1996), "Model Selection in Neural networks", *ZEW Discussion Papers*, No. 96-21

Aydin, A., dan Tiryaki, S. (2014), "An Artificial Neural Network Model for Predicting Compression Strength of Heat Treated Woods and Comparison with a Multiple Linear Regression Model", *Construction and Building Materials*, No.62, hal. 102-108.

Bishop, C.M. (2005), *Neural networks for Pattern Recognition*, Oxford University Press, Inc., New York.

Caffrey, J.M (2013), *Understanding and Using K-Fold Cross-Validation for Neural Networks*, <https://visualstudiomagazine.com/articles/2013/10/01/understandingand-using-kfold.aspx>.

Castillo, C.C., Korbl, B., Stewart A., Gonzales, J.F., dan Ponce, F. (2011), "Application of Spatial Analysis to the Examination of Dengue Fever in Guayaquil, Ecuador", *Procedia Environmental Sciences*, No.7, hal 188-193

Dinkes Surabaya (2015), *waspadai-demam-berdarah-berantas-sarang nya-muk*, <http://dinkes.surabaya.go.id/portal/index.php/artikelkesehatan>.

Dinkes Jatim (2013), *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*, Dinkes Provinsi Jatim, Surabaya.

Draper, N.R and Smith, H (1992), *Applied Regression Analysis*, 3rd edition, John Wiley & Sons, Inc., New York.

Duddu, V.R, dan Pulugharta, S.S. (2012), "Neural networks to Estimate Crashes at Zona Level for Transportation Planning", *Association for European Transport and Contributors*.

Earnest, A. et.al. (2011), "Comparing Statistical Models to Predict Dengue Fever Notifications ", *Computational and Mathematical Methods in Medicine*, Vol.2012.

Famoye, F., Wulu, J. T., dan Singh, K. P. (2004), "On The Generalized Poisson Regression Model with an Application to Accident Data", *Journal of Data Science*, No. 2, hal. 287-295.

Fatmasari, F. (2014), *Pendekatan negative binomial regression untuk Data Berdistribusi Poisson yang mengalami Overdispersi*, Tugas Akhir Matematika, Universitas Brawijaya, Malang.

Fausett, L. (1994), *Fundamentals of Neural networks: Architectures, Algorithms, and Applications*, Prentice-Hall Inc., New Jersey.

Greene, W. (2008), "Functional forms for the *negative binomial* model for count data". *Economics Letters* 99(3) , hal 585-590.

Hardin, J. W., & Hilbe, J. M. (2007). *Generalized Linear Models and Extensions Second Edition*, : Stata Press., Texas.

Haykin, S., (1999), *Neural networks, A Comprehensive Foundation*, 2nd edition, Prentice-Hall International, Inc., New Jersey.

Hidayanti,U. (2015), *Pemodelan Dan Pemetaan Jumlah Kasus Demam Berdarah Dengue (Dbd) Di Kota Surabaya Dengan Geographically Weighted Negative Binomial Regression (Gwnbr) Dan Flexibly Shaped Spatial Scan Statistic*, Tugas Akhir Statistika, ITS, Surabaya.

Hakim, A (2015), *Jumlah Kasus Demam Berdarah di Surabaya Meningkat*, <http://www.antaraneews.com/berita/477267/jumlah-kasus-demam-berdarah-di-surabaya-meningkat>

Hocking, R. R. (1996). *Methods and applications of linear models: regression and the analysis of variance*, John Wiley and Sons ., New York.

Kementrian Kesehatan RI (2010), *Buletin Jendela Epidemiologi*, Vol.2, Kemen-kes RI, Jakarta

Khormi, M.H., dan Kumar, L. (2011), "Modeling Dengue Fever Risk Based on Socioeconomic Parameters, Nationality and Age Groups: GIS and Remote Sensing Based Case Study", *Science of the Total Environment*, No. 409, hal. 4713-4719.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*, Chapman and Hall., London.

Myers, R.H (1990), *Classical and Modern Regression with Applications*, 2nd edition, Duxbury Advanced Series in Statistics, PWS-Kent Publishing Co. Boston, MA.

Nandasari, B.N. (2014), *Pemodelan Jumlah Kejadian Luar Biasa Difteri di Jawa Timur Menggunakan Geographically Weighted Negative Binomial Regression*, Tugas Akhir Statistika, ITS, Surabaya.

Notoatmodjo, S. (2003). *Ilmu Kesehatan Masyarakat, Prinsip-Prinsip Dasar*. Ja-karta: Rineka Cipta.

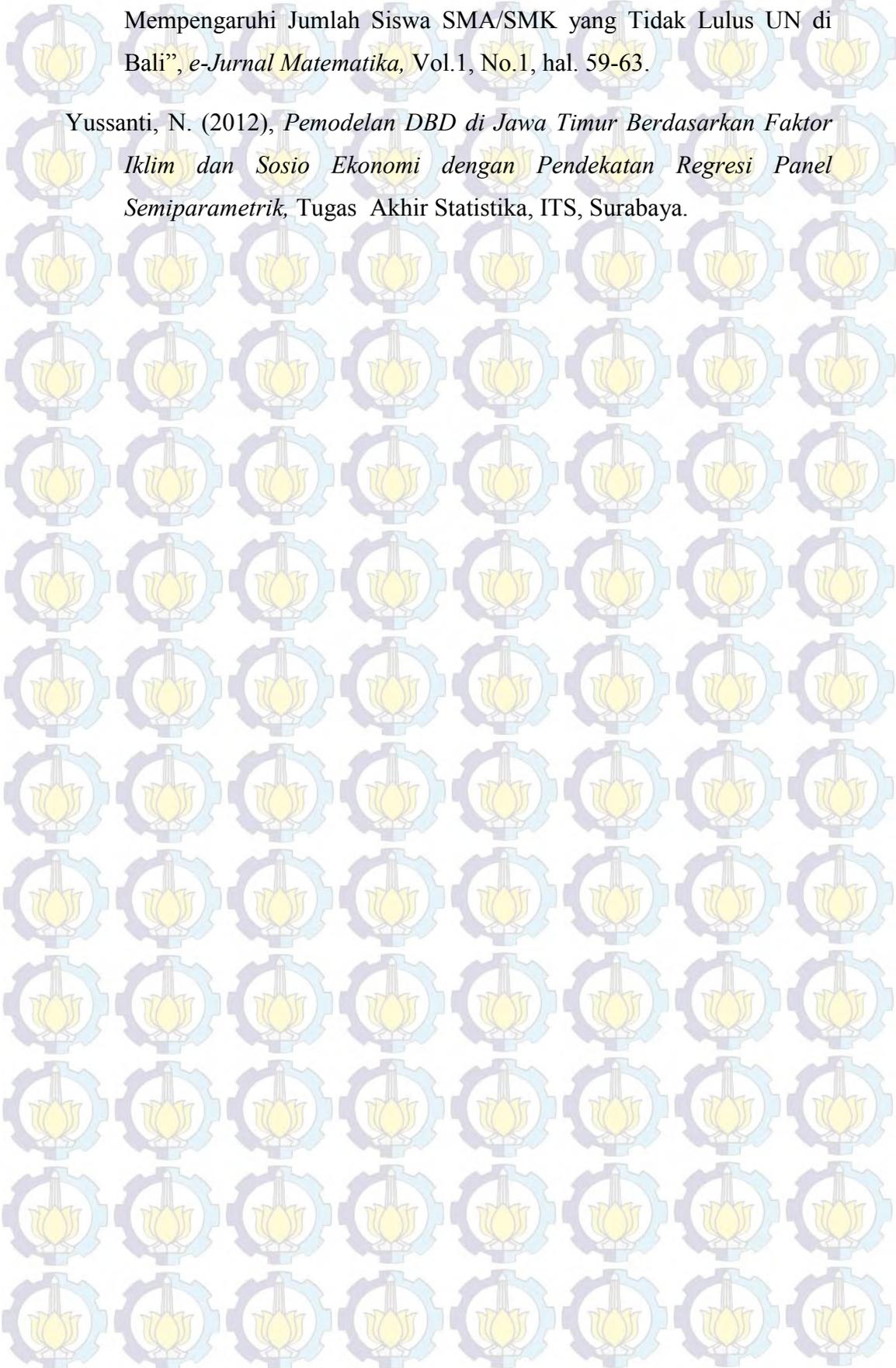
Rodríguez, G. (2007). *Lecture Notes on Generalized Linear Models*. URL: <http://data.princeton.edu/wws509/notes/>

Sundari, I. (2014), *Poisson regression dan Penerapannya untuk Memodelkan Hubungan Usia dan Perilaku Merokok terhadap Jumlah Kematian Penderita Penyakit Kanker Paru-Paru*, Tugas Akhir Matematik, Universitas Andalas, Padang.

Terasvirta, T., Tjostheim, D. dan Granger, C.W.J. (1994). "Aspect Modelling Nonlinear Time Series", dalam *Handbook of Econometrics*, eds. R.F. Engle dan D.L. McFadden, Vol. 4, Bab 48, hal. 2919-2957, Elsevier Science B.V.

Wei, W.W.S. (2006). *Time Series Analysis Univariate and Multivariate Methods*. United States : Pearson Education, Inc.

Yulianingsih, K.A., Sukarsa, K.A., dan Suciptawati, L.P. (2015), "Penerapan Poisson regression untuk Mengetahui Faktor-Faktor yang



Mempengaruhi Jumlah Siswa SMA/SMK yang Tidak Lulus UN di Bali”, *e-Jurnal Matematika*, Vol.1, No.1, hal. 59-63.

Yussanti, N. (2012), *Pemodelan DBD di Jawa Timur Berdasarkan Faktor Iklim dan Sosio Ekonomi dengan Pendekatan Regresi Panel Semiparametrik*, Tugas Akhir Statistika, ITS, Surabaya.

BIODATA PENULIS



“Senyumlah, Karena Pilihan Allah untuk Hambanya Tidak Akan Pernah Salah”

merupakan salah satu motto hidup penulis yang bernama lengkap Fastha Aulia Pradhani. Penulis melanjutkan pendidikannya pada jenjang S-2 di Statistika ITS pada tahun 2014. Selama menjadi mahasiswa Pasca Sarjana, penulis aktif dalam organisasi FORSIMPAS (Forum Silaturahmi Mahasiswa Pasca Sarjana) sebagai bendahara. Penulis juga pernah menjadi asisten dosen pada mata kuliah Analisis Data II. Apabila pembaca ingin berdiskusi mengenai laporan Tesis ini, penulis dapat dihubungi melalui email: fastha13@gmail.com