

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 9, NOMOR 1, JUNI 2017 ISSN 2086 – 4132  
AKREDITASI NOMOR: 747/Akred/P2MI-LIPI/04/2016

Analisis Regresi Tobit Spasial: Studi Kasus Penggunaan Internet di Pulau Jawa

ANDHIE SURYA MUSTARI dan ISMAINI ZAIN

*Generalized Multilevel Linear Model* dengan Pendekatan Bayesian untuk Pemodelan Data  
Pengeluaran Perkapita Rumahtangga

AZKA UBAIDILLAH, ANANG KURNIA dan KUSMAN SADIK

Faktor-Faktor yang Memengaruhi *Foreign Direct Investment* (FDI) di Enam Koridor Ekonomi  
Indonesia: *Market Seeking* atau *Resource Seeking*?

IRIANI TRISNA RAHAYU dan ERNAWATI PASARIBU

Determinan Perilaku Merokok pada Remaja Sekolah di Indonesia

TITIK HARSANTI dan FEBRI WICAKSONO

Persepsi Masyarakat Kelurahan Bukit Duri terhadap Program Normalisasi Kali Ciliwung di  
Jakarta Tahun 2017 serta Variabel-Variabel yang Memengaruhinya

LOVERIA CANDRA PUSPITA dan ACHMAD PRASETYO

*Named Entity Recognition on A Collection of Research Titles*

SITI MARIYAH



UNIT PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT  
SEKOLAH TINGGI ILMU STATISTIK  
(UPPM-STIS)

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

---

Jurnal “Aplikasi Statistika dan Komputasi Statistik” memuat karya ilmiah hasil penelitian dan kajian teori statistik dan komputasi statistik yang diterapkan khususnya pada bidang ekonomi dan sosial kependudukan, serta teknologi informasi yang terbit dua kali dalam setahun setiap bulan Juni dan Desember.

**Penanggung Jawab:** Ketua Sekolah Tinggi Ilmu Statistik

**Dewan Redaksi :**

**Ketua:** Dr. Hardius Usman

**Anggota:** Dr. Nasrudin.

Dr. Ernawati Pasaribu

**Mitra Bestari:**

Prof. Dr. Abuzar Asra

Prof. Dr. Irdam Ahmad

Prof. Nur Iriawan, Ph.D.

Dr. Hari Wijayanto

Dr. Erni Tri Astuti

Setia Pramana, Ph.D.

**Pelaksana Redaksi:**

M. Dokhi, Ph.D.

Dr. Tiodora Hadumaon S.

Dr. I Made Arcana

Dr. M. Ari Anggorowati

Novia Budi Parwanto, Ph.D.

**Alamat Redaksi:**

Sekolah Tinggi Ilmu Statistik

Jl. Otto Iskandardinata 64C

Jakarta Timur 13330

Telp. 021-8191437

Redaksi menerima karya ilmiah atau artikel penelitian mengenai kajian teori statistik dan komputasi statistik pada bidang ekonomi dan sosial kependudukan, serta teknologi informasi. Redaksi berhak menyunting tulisan tanpa mengubah makna substansi tulisan. Isi Jurnal Aplikasi Statistika dan Komputasi Statistik dapat dikutip dengan menyebutkan sumbernya.

## **PENGANTAR REDAKSI**

Puji syukur kehadirat Allah, Tuhan Yang Maha Esa, “Jurnal Aplikasi Statistika dan Komputasi Statistik” Volume 9, Nomor 1, Juni 2017 dapat diterbitkan. Jurnal kampus STIS ini dapat terwujud atas partisipasi semua pihak, internal maupun eksternal STIS yang telah mengirimkan tulisannya, serta mitra bestari.

Semoga artikel dalam jurnal ini dapat menambah pengetahuan para pembaca tentang penggunaan metode statistika serta komputasi statistik pada berbagai jenis data. Redaksi terus menunggu artikel-artikel ilmiah selanjutnya dari Bapak/Ibu guna dapat menghasilkan publikasi yang menjadi salah satu sarana untuk memberikan sosialisasi statistika bagi masyarakat.

Jakarta, Juni 2017

Ketua Dewan Redaksi,

**Hardius Usman**

# JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 9, NOMOR 1, JUNI 2017  
AKREDITASI NOMOR: 747/Akred/P2MI-LIPI/04/2016

## DAFTAR ISI

<b>Pengantar Redaksi</b> .....	iii
<b>Daftar Isi</b> .....	iv
<b>Abstrak</b> .....	v-x
<b>Analisis Regresi Tobit Spasial: Studi Kasus Penggunaan Internet di Pulau Jawa</b> Andhie Surya Mustari dan Ismaini Zain.....	1-16
<b>Generalized Multilevel Linear Model dengan Pendekatan Bayesian untuk Pemodelan Data Pengeluaran Perkapita Rumah tangga</b> Azka Ubaidillah, dkk.....	17-28
<b>Faktor-Faktor yang Memengaruhi <i>Foreign Direct Investment</i> (FDI) di Enam Koridor Ekonomi Indonesia: <i>Market Seeking</i> atau <i>Resource Seeking</i>?</b> Iriani Trisna Rahayu dan Ernawati Pasaribu .....	29-38
<b>Determinan Perilaku Merokok pada Remaja Sekolah di Indonesia</b> Titik Harsanti dan Febri Wicaksono .....	39-46
<b>Persepsi Masyarakat Kelurahan Bukit Duri terhadap Program Normalisasi Kali Ciliwung di Jakarta Tahun 2017 serta Variabel-Variabel yang Memengaruhinya</b> Loveria Candra Puspita dan Achmad Prasetyo .....	47-58
<b><i>Named Entity Recognition on A Collection of Research Titles</i></b> Siti Mariyah.....	59-70

Kata kunci bersumber dari artikel. Lembar abstrak ini boleh diperbanyak tanpa izin dan biaya

DDC: 315.98

Andhie Surya Mustari dan Ismaini Zain

Analisis Regresi Tobit Spasial: Studi Kasus Penggunaan Internet di Pulau Jawa

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 1 – 16

Abstrak

Dibutuhkan metode khusus untuk menganalisis data tersensor yang memiliki korelasi spasial. Jika menggunakan regresi linier, akan menghasilkan estimasi parameter yang tidak valid, tidak terpenuhinya asumsi normalitas dan mengaburkan interpretasi model. Model regresi Tobit spasial digunakan untuk menganalisis data penggunaan internet di Pulau Jawa. Estimasi parameter menggunakan metode MCMC Gibbs sampler dengan pendekatan inferensia Bayesian. Hasilnya, penggunaan internet di Pulau Jawa dipengaruhi oleh persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki telepon genggam, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

Kata kunci: data tersensor, korelasi spasial, Tobit spasial, MCMC, penggunaan internet

DDC: 315.98

Azka Ubaidillah, Anang Kurnia dan Kusman Sadik

*Generalized Multilevel Linear Model* dengan Pendekatan Bayesian untuk Pemodelan Data Pengeluaran Perkapita Rumahtangga

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 17 – 28

Abstrak

Data pengeluaran perkapita rumahtangga merupakan salah satu informasi penting sebagai pendekatan untuk mengukur tingkat kemakmuran dan kesejahteraan di suatu daerah. Data tersebut sangat diperlukan oleh pemerintah baik di pusat maupun daerah dalam merumuskan, melaksanakan dan mengevaluasi pelaksanaan pembangunan. Penelitian ini akan menganalisis model yang tepat untuk pemodelan data pengeluaran perkapita rumahtangga yang memperhitungkan kekhususan data BPS yang memiliki struktur hirarki dan pola distribusi data yang memiliki karakteristik *skewed* kanan. Pemodelan dilakukan dengan menggunakan distribusi Log-normal tiga parameter (LN3P) dan Log-logistik tiga parameter (LL3P) dengan struktur satu tingkat (*unilevel*) dan dua tingkat (*multilevel*). Proses pendugaan parameter dilakukan dengan metode *Markov Chain Monte Carlo* (MCMC) dan algoritma *Gibbs Sampling*. Hasil penelitian menunjukkan bahwa pada model *unilevel*, model LL3P lebih baik dari model LN3P. Sedangkan pada model *multilevel*, model LN3P lebih baik dari model LL3P. Hasil penelitian juga menunjukkan model terbaik untuk pemodelan data pengeluaran perkapita rumahtangga adalah model multilevel LN3P dengan intercept sebagai komponen berhirarki dengan nilai Deviance Information Criterion (DIC) terkecil.

Kata kunci : *Generalized Multilevel Linear Model*, LL3P, LN3P, MCMC, Pengeluaran perkapita rumahtangga.

DDC: 315.98

Iriani Trisna Rahayu dan Ernawati Pasaribu

Faktor-Faktor yang Memengaruhi *Foreign Direct Investment* (FDI) di Enam Koridor Ekonomi Indonesia: *Market Seeking* atau *Resource Seeking*?

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 29 – 38

Abstrak

Pembangunan ekonomi suatu negara sangat tergantung pada besarnya penanaman modal asing langsung atau *Foreign Direct Investment* (FDI), termasuk di enam koridor ekonomi Indonesia. Kesenjangan kondisi di koridor ekonomi diperkirakan akan membawa perbedaan yang mempengaruhi arus masuk FDI ke dalam koridor. Penelitian ini menggunakan regresi data panel untuk menganalisis faktor-faktor di balik arus masuk FDI di setiap koridor ekonomi dan untuk menentukan karakteristik FDI di setiap koridor ekonomi. Hasil penelitian menunjukkan bahwa proporsi belanja modal pemerintah, jumlah angkatan kerja berpendidikan tinggi, keterbukaan perdagangan, dan proporsi ekspor minyak dan mineral hanya mempengaruhi arus masuk FDI di beberapa koridor ekonomi. Lebih jauh lagi dapat diindikasikan bahwa sementara “*market seeking FDI*” terjadi di semua koridor ekonomi Indonesia, “*resource seeking FDI*” hanya ditemukan di koridor ekonomi Sulawesi, Maluku dan Papua.

Kata kunci: *FDI*, Koridor Ekonomi, data panel, *market seeking*, *resource seeking*

DDC: 315.98

Titik Harsanti dan Febri Wicaksono

Determinan Perilaku Merokok Pada Remaja Sekolah di Indonesia

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 39 – 46

Abstrak

Saat ini merokok telah menjadi masalah kesehatan secara global dan menjadi beban ekonomi yang berat. Di Indonesia, tren merokok cenderung semakin meningkat dan kecenderungan ini tidak hanya terjadi pada orang dewasa tetapi juga pada remaja. Penelitian ini bertujuan mengidentifikasi faktor-faktor yang mempengaruhi perilaku merokok bagi kalangan remaja yang bersekolah di Indonesia dengan menggunakan model logistik multivariat biner. Analisis dilakukan dengan menggunakan 5.986 sampel siswa dari *Global Youth Tobacco Survey 2014* (GYTS) 2014. Hasilnya menunjukkan bahwa 25% siswa pernah merokok dan 15% siswa saat ini merokok. Peluang siswa untuk merokok lebih tinggi di kalangan anak laki-laki dibandingkan dengan anak perempuan. Risiko merokok yang lebih tinggi teramati di antara siswa yang memiliki teman dekat yang merokok dibandingkan dengan siswa yang tidak memiliki teman dekat yang merokok. Siswa yang salah satu atau kedua orang tuanya merokok lebih cenderung merokok dibandingkan dengan siswa yang orang tuanya tidak merokok. Siswa yang pernah melihat gurunya merokok atau pernah melihat orang-orang merokok di rumah mereka dan tempat-tempat umum lebih cenderung merokok dibandingkan dengan mereka yang tidak pernah melihat gurunya merokok atau tidak pernah melihat orang merokok di rumah mereka dan tempat umum. Temuan ini menunjukkan bahwa penegakan peraturan untuk mengurangi aksesibilitas rokok diperlukan untuk mengekang penggunaan rokok di kalangan siswa. Selain itu, intervensi dan kampanye pendidikan yang menargetkan siswa sekolah menengah juga diperlukan.

Kata kunci: Tembakau, Merokok, Sekolah, Remaja, Indonesia

DDC: 315.98

Loveria Candra Puspita dan Achmad Prasetyo

Persepsi Masyarakat Kelurahan Bukit Duri terhadap Program Normalisasi Kali Ciliwung di Jakarta Tahun 2017 serta Variabel-Variabel yang Memengaruhinya

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 47 – 58

#### Abstrak

Salah satu cara menangani masalah banjir adalah melakukan program normalisasi sungai. Namun, tidak semua masyarakat menerima program ini. Untuk itu, ingin diketahui persepsi masyarakat terhadap normalisasi Kali Ciliwung dan menganalisa variabel-variabel yang memengaruhinya. Data persepsi diperoleh melalui survei dengan pendekatan rumahtangga di Kelurahan Bukit Duri yang kemudian dianalisis dengan regresi logistik. Hasil penelitian menunjukkan bahwa 28 persen rumahtangga sekitar sungai dan 22 persen rumahtangga bukan sekitar sungai menolak normalisasi. Persepsi rumahtangga sekitar sungai secara signifikan dipengaruhi oleh jenis kelamin, keikutsertaan organisasi, mendapatkan sosialisasi, dan pengeluaran perkapita. Sedangkan persepsi rumah tangga yang tinggal bukan di sekitar sungai dipengaruhi status pekerjaan, keikutsertaan organisasi, dan mendapatkan sosialisasi.

Kata kunci : persepsi, normalisasi sungai, regresi logistik

DDC: 315.98

Siti Mariyah

*Named Entity Recognition on A Collection of Research Titles*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 59 – 70

#### Abstrak

Judul dapat membantu pembaca untuk mendapatkan sudut pandang universal dari artikel tersebut sebagai pemahaman awal sebelum membaca konten secara keseluruhan. Pada penelitian teknis, judul memuat informasi penting. Dalam penelitian ini, kami mengembangkan teknik ekstraksi informasi untuk mengenali dan mengekstrak masalah, metode, dan domain penelitian yang terdapat dalam judul. Kami menerapkan pendekatan supervised learning pada 671 judul penelitian dalam bidang ilmu komputer dari beragam jurnal online dan prosiding seminar internasional. Kami melakukan beberapa percobaan dengan skema yang berbeda untuk mempelajari pengaruh fitur dan kinerja algoritma. Kami menguji fitur kontekstual, fitur sintaksis, dan fitur bag of words menggunakan Naïve Bayes dan Maximum Entropy. Classifier Naïve Bayes yang belajar dari kelompok set fitur pertama berhasil memprediksi kategori masing-masing token dalam dataset judul. Keakuratan dan nilai f1-score untuk setiap kelas lebih dari 0,80 karena kelompok pertama set fitur mempertimbangkan lokasi token dalam sebuah kalimat, memperhatikan token sekitar dan tag POS dari beberapa token sebelum dan sesudah. Sementara classifier Naïve Bayes yang dipelajari dari kelompok kedua dari rangkaian fitur lebih tepat mengklasifikasikan token frase daripada token kata.

Kata kunci: *research titles, named entity recognition, information extraction, contextual features, naïve bayes classifier.*





Kata kunci bersumber dari artikel. Lembar abstrak ini boleh diperbanyak tanpa izin dan biaya

DDC: 315.98

Andhie Surya Mustari dan Ismaini Zain

Analisis Regresi Tobit Spasial: Studi Kasus  
Penggunaan Internet di Pulau Jawa

Jurnal Aplikasi Statistika & Komputasi  
Statistik, Volume 9, Nomor 1, Juni 2017,  
hal 1 – 16

*Abstract*

*Special method is required for analyzing censored data with spatial dependence. Using linear regression will results in invalid parameter estimations, normality assumption violations, and obscure the model interpretation. Spatial Tobit regression model is used to analyze the data of internet usage in Java. MCMC Gibbs sampler method with Bayesian inference approach was used for parameter estimation. As a result, internet usage in Java Island is influenced by the percentage of population living in urban areas, the percentage of population graduated from senior high school, the average length of school, the percentage of households with mobile phones, and the percentage of villages receiving cell phone signal.*

*Keywords: censored data, spatial dependence, spatial Tobit, MCMC, internet usage*

DDC: 315.98

Azka Ubaidillah, Anang Kurnia dan  
Kusman Sadik

*Generalized Multilevel Linear Model dengan Pendekatan Bayesian untuk Pemodelan Data Pengeluaran Perkapita Rumahtangga*

Jurnal Aplikasi Statistika & Komputasi  
Statistik, Volume 9, Nomor 1, Juni 2017,  
hal 17 – 28

*Abstract*

*Household per capita expenditure data is one of the important information as an approach to measure the level of prosperity in an area. Such data is needed by the government, both at the central and regional levels in formulating, implementing and evaluating the implementation of development programs. This research is aimed at modeling the household per capita expenditure data which takes into account the specificity of BPS data which has a hierarchical structure, and data distribution pattern which has the right skewed characteristic. The modeling is done by using the three parameters of Log-normal distribution (LN3P) and the three parameters of Log-logistics (LL3P) with a single level (unilevel) and two levels (multilevel) structure. The parameter estimation process is done by Markov Chain Monte Carlo (MCMC) method and Gibbs Sampling algorithm. The results showed that on the unilevel model, the LL3P model is better than the LN3P model. While in multilevel model, LN3P model is better than LL3P model. The results also show that the best model for modeling household per capita expenditure data is the LN3P multilevel model with the smallest Deviance Information Criterion (DIC) value.*

*Keywords: Generalized Multilevel Linear Model, LL3P, LN3P, MCMC, Household per capita expenditure.*

DDC: 315.98

Iriani Trisna Rahayu dan Ernawati Pasaribu

Faktor-Faktor yang Memengaruhi *Foreign Direct Investment* (FDI) di Enam Koridor Ekonomi Indonesia: *Market Seeking* atau *Resource Seeking*?

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 29 – 38

*Abstract*

*The economic development of a country depends on the amount of foreign direct investment (FDI), including in the Indonesian six economic corridors. The huge gaps of conditions in economic corridors are expected to differences infactors affecting the FDI-inflow into the corridors. This study uses a panel data regression to analyze factors behind the FDI-inflow in each economic corridor and to determine the FDI characteristic in each economic corridor. It shows that the proportion of government capital expenditure, number of highly-educated labor force, trade openness, and the proportion of oil and mineral export affect the FDI-inflow only in some economic corridors. Furthermore, it indicates that, while market seeking FDI occurred in all Indonesian economic corridors, resource seeking FDI was only found in Sulawesi, Maluku and Papua economic corridors..*

*Key words: Foreign Direct Investment (FDI), Indonesian economic corridors, panel data regression, market seeking, resource seeking*

DDC: 315.98

Titik Harsanti dan Febri Wicaksono

Determinan Perilaku Merokok pada Remaja Sekolah di Indonesia

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 39 – 46

*Abstract*

*Smoking is a global public health concern and it imposes a heavy economic burden. However, the trend of smoking in Indonesia seems to be increasing and the magnitude of the problem affects not only adults but also adolescents. This paper identifies cigarette smoking determinants among school adolescents in Indonesia, using a multivariate binary logistic model. The analysis uses 5,986 samples of students from the 2014 Indonesia Global Youth Tobacco Survey (GYTS). The results show that 25% of the students have ever smoked and 15% of students are currently smoking. The students' odds of smoking are higher for boys compared to girls. Higher risk of smoking is observed among the students who have closed-peer smoking compared to students who don't have closed-peer smoking. Students whose one or both parents are smoking are more likely to smoke compared to whose parents are not smoking. Students who have seen their teacher smoking or have seen people smoking in their house and public places are more likely to smoke compared to who haven't ever seen their teacher smoking or haven't ever seen people smoking in their house and public places. These findings suggest that enforcement of legislations to decrease accessibility of cigarettes are necessary to curb the cigarette use among students. Beside that the interventions and education campaigns that target secondary school students are also needed.*

*Keywords: Tobacco, Smoking, School, Adolescent, Indonesia*

DDC: 315.98

Loveria Candra Puspita dan Achmad Prasetyo

Persepsi Masyarakat Kelurahan Bukit Duri terhadap Program Normalisasi Kali Ciliwung di Jakarta Tahun 2017 serta Variabel-Variabel yang Memengaruhinya

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 47 – 58

*Abstract*

*River normalization program is one of the ways to handle flood problems. However, not all communities accept this program. For that, we want to know the public perception towards normalization of Ciliwung River and analyze the variables that influence it. Perception data was obtained through survey with household approach in Bukit Duri Village which then analyzed by logistic regression. The results show that 28 percent of households around the river and 22 percent of households not around the river reject normalization. Household perceptions around the river are significantly influenced by sex, organizational participation, socialization, and per capita expenditure. The non-rivers are influenced by employment status, organizational participation, and socialization.*

*Keywords: perception, river normalization, logistic regression*

DDC: 315.98

Siti Mariyah

*Named Entity Recognition on A Collection of Research Titles*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 9, Nomor 1, Juni 2017, hal 59 – 70

*Abstract*

*The title can help the reader to get the universal point of view of the article as the initial understanding before reading the content as a whole. On technical research papers, the title states essential information. In this study, we aim to develop information extraction techniques to recognize and extract problem, method, and domain of research contained in a title. We apply supervised learning on 671 research titles in computer science from various online journals and international conference proceedings. We conducted some experiments with different schemas to discover the influence of features and the performance of the algorithm. We examined contextual, syntactic, and the bag of words feature sets using Naïve Bayes and Maximum Entropy. The Naïve Bayes classifier learned from the first group of the feature set is successful in predicting category of each token in title dataset. The accuracy and f1-score for each class are more than 0.80 since the first group of feature sets considers the location of a token within a sentence, considers the token and POS tag of some tokens before and after and deliberates the rules of a token. While the Naïve Bayes classifier learned from the second group of the feature set is more appropriate classifying a phrase token than a word token.*

*Keywords: research titles, named entity recognition, information extraction, contextual features, naïve bayes classifier*



# ANALISIS REGRESI TOBIT SPASIAL : Studi Kasus Penggunaan Internet di Pulau Jawa

Andhie Surya Mustari<sup>1</sup>, Ismaini Zain<sup>2</sup>

<sup>1</sup>Badan Pusat Statistik, email: andhie@bps.go.id

<sup>2</sup>Institut Teknologi Sepuluh Nopember,  
email: <sup>1</sup>andhie@bps.go.id, <sup>2</sup>ismaini\_z@statistika.its.ac.id

## Abstrak

Dibutuhkan metode khusus untuk menganalisis data tersensor yang memiliki korelasi spasial. Jika menggunakan regresi linier, akan menghasilkan estimasi parameter yang tidak valid, tidak terpenuhinya asumsi normalitas dan mengaburkan interpretasi model. Model regresi Tobit spasial digunakan untuk menganalisis data penggunaan internet di Pulau Jawa. Estimasi parameter menggunakan metode MCMC Gibbs sampler dengan pendekatan inferensia Bayesian. Hasilnya, penggunaan internet di Pulau Jawa dipengaruhi oleh persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki telepon genggam, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

**Kata kunci:** data tersensor, korelasi spasial, Tobit spasial, MCMC, penggunaan internet

## Abstract

*Special method is required for analyzing censored data with spatial dependence. Using linear regression will results in invalid parameter estimations, normality assumption violations, and obscure the model interpretation. Spatial Tobit regression model is used to analyze the data of internet usage in Java. MCMC Gibbs sampler method with Bayesian inference approach was used for parameter estimation. As a result, internet usage in Java Island is influenced by the percentage of population living in urban areas, the percentage of population graduated from senior high school, the average length of school, the percentage of households with mobile phones, and the percentage of villages receiving cell phone signal.*

**Keywords:** *censored data, spatial dependence, spatial Tobit, MCMC, internet usage*

## PENDAHULUAN

Menggunakan model regresi linier klasik untuk analisis data tersensor yang memiliki korelasi spasial merupakan keputusan yang kurang tepat. Istilah data tersensor digunakan untuk menjelaskan sekelompok data yang memiliki sejumlah nilai yang tidak diketahui pada batas atas atau bawahnya. Long (1997) menjelaskan bahwa jika menggunakan model regresi linier pada keseluruhan data tersensor, akan menghasilkan nilai parameter yang *overestimates* pada *slope* dan *underestimates* pada *intercept*. Sedangkan jika menghilangkan atau memotong observasi yang nilainya tidak diketahui, akan menghasilkan koefisien parameter yang *underestimates* pada *slope* dan *overestimates* pada *intercept*. Data terpotong menyebabkan terjadinya korelasi antara variabel prediktor dengan residual, sehingga menghasilkan estimasi yang tidak konsisten. Efek korelasi spasial dapat muncul pada pembentukan model regresi linier yang menggunakan data kewilayahan (*cross section data*). Hal tersebut mengakibatkan tidak terpenuhinya asumsi error yang independen dan identik berdistribusi normal, sehingga menghasilkan estimasi parameter yang tidak valid dan mengaburkan interpretasi model (Marsh, Mittelhammer, & Huffaker, 2000). Korelasi spasial dapat diamati dari mengelompoknya besaran nilai tertentu pada data yang berasal dari wilayah yang berdekatan, misalnya data tingkat penggunaan internet kabupaten/kota di Pulau Jawa.

Tingginya tingkat penggunaan internet di Pulau Jawa terutama ditemukan di kota-kota besar sebagai pusat jasa pendidikan dan hiburan, seperti DKI Jakarta, Yogyakarta, Bandung, dan Surabaya, kemudian diikuti oleh wilayah kabupaten/kota lain di sekitarnya. Fenomena dependensi spasial ini dapat diaplikasikan untuk analisis data tersensor yang memiliki korelasi spasial, dimana kabupaten/kota dengan penggunaan internet kategori tinggi dapat dianggap sebagai data yang tidak diketahui nilainya. Sebanyak 34 dari 118 kabupaten/kota di Pulau Jawa memiliki persentase pengguna internet yang lebih tinggi daripada 16 persen

(BPS, 2011), suatu nilai yang setara dengan jumlah akun *facebook* yang dibuat oleh penduduk Indonesia pada tahun 2011 (Socialbaker, 2011).

Dibutuhkan metode khusus untuk melakukan analisis penggunaan internet, dengan asumsi bahwa data penggunaan internet di Pulau Jawa merupakan data tersensor yang memiliki korelasi spasial. Fischer dan Getis (2010) mengatakan bahwa pemodelan data tersensor yang melibatkan wilayah sebaiknya menggunakan analisis spasial, metode yang paling sesuai adalah regresi Tobit spasial. Selain itu, Lee (2010) juga menyatakan bahwa pendekatan Tobit spasial lebih disarankan untuk analisis wilayah yang melibatkan data tersensor. Analisis regresi Tobit spasial digunakan apabila variabel respon pada model spasial melibatkan data yang diyakini memiliki nilai tersensor (LeSage & Pace, 2009).

Penelitian ini bermaksud untuk membentuk model regresi Tobit spasial dan mencari metode estimasi parameter dari model regresi Tobit spasial. Data yang digunakan sebagai variabel respon adalah persentase penduduk umur 5 tahun ke atas yang mengakses internet selama tiga bulan terakhir di pulau Jawa pada tahun 2010. Sensor diberikan kepada wilayah kabupaten/kota dengan persentase pengguna internet lebih besar dari 16 persen, yang dianggap sebagai batas minimal persentase pengguna internet yang ingin dicapai oleh suatu kabupaten/kota.

## METODOLOGI

### a. Tinjauan Referensi

#### Model Regresi Tobit

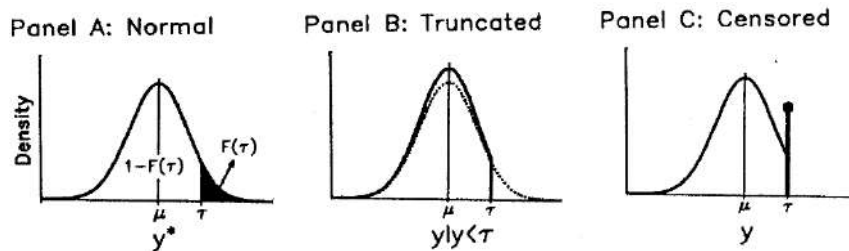
Misalkan  $Y^*$  adalah suatu variabel respon dengan informasi yang lengkap dan  $Y$  adalah data sampel dari  $Y^*$ , maka variabel respon  $Y$  yang tersensor dapat didefinisikan sebagai berikut:

$$y_i = \begin{cases} y_i^* & \text{jika } y_i^* < \tau \\ \tau & \text{jika } y_i^* \geq \tau \end{cases} \quad (1)$$

dimana  $i = 1, 2, \dots, n$ ,  $\tau$  merupakan suatu konstanta batasan, dan  $n$  adalah banyaknya observasi (Tobin, 1958)<sup>1</sup>.

Jika nilai  $y_i^*$  tidak diketahui ketika  $y_i^* \geq \tau$ , maka  $y^*$  mengandung suatu variabel latent yang tidak dapat diamati pada seluruh *range*-nya. Panel A pada Gambar 2.1 memperlihatkan distribusi dari  $y^* \sim N(\mu, \sigma^2)$ , dengan nilai variabel latent

digambarkan sebagai wilayah gelap pada kurva. Jika nilai yang tidak diketahui tersebut dipotong, maka sebagian informasi yang dapat menjelaskan populasi akan hilang dimana kurva distribusi menjadi lebih runcing (panel B). Panel C menggambarkan data tersensor yang mengelompok pada nilai  $\tau$  sehingga tidak merubah informasi berkaitan distribusi populasi.



Gambar 1. Ilustrasi Variabel Latent, Terpotong, dan Tersensor (Long, 1997)

Model Tobit dibentuk dengan terlebih dahulu mengasumsikan adanya hubungan linier antara  $Y^*$  dengan variabel prediktor  $X$  yang dinyatakan dengan:

$$y_i^* = x_i^T \beta + \varepsilon_i \quad (2)$$

dimana  $i = 1, 2, \dots, n$ ,  $\varepsilon_i \sim N(0, \sigma^2)$ ,  $x_i^T = [1 \ x_{1i} \ x_{2i} \ \dots \ x_{pi}]$  adalah vektor variabel prediktor,  $\beta = [\beta_0 \ \beta_1 \ \dots \ \beta_p]^T$  adalah vektor parameter, dan  $p$  merupakan banyaknya variabel, dengan  $y^*$  mengandung variabel latent yang mewakili nilai tersensor<sup>2</sup>. Nilai tersensor tersebut bisa lebih kecil dari suatu batas bawah ( $y_i^* \leq \tau < \bar{y}^*$ ), lebih besar dari batas atas ( $y_i^* \geq \tau > \bar{y}^*$ ), atau keduanya.

Ketika  $y^*$  tersensor pada batas atas  $y_i^* \geq \tau$ , maka model regresi Tobit dinyatakan dengan persamaan sebagai berikut:

$$y_i = \begin{cases} y_i^* = x_i^T \beta + \varepsilon_i & \text{jika } y_i^* < \tau \\ \tau & \text{jika } y_i^* \geq \tau \end{cases} \quad (3)$$

### Model Regresi Spasial

Pada tahun 1988, Anselin mengembangkan bentuk umum dari model regresi spasial (*general spatial model*)

menggunakan data *cross section* sebagai berikut:

$$y = (I - \rho W)^{-1} X \beta + (I - \rho W)^{-1} (I - \lambda W)^{-1} \varepsilon \quad (4)$$

dimana  $\varepsilon \sim N(0, \sigma^2 I)$ ,  $y$  merupakan vektor variabel respon yang memiliki korelasi spasial,  $X$  adalah matriks variabel prediktor, dan  $\beta$  adalah vektor parameter regresi. Adapun  $\rho$  adalah koefisien korelasi spasial lag dari variabel respon,  $\lambda$  merupakan koefisien korelasi spasial error, dan  $W$  merupakan matriks penimbang spasial dengan elemen diagonalnya bernilai nol. Persamaan ini juga biasa disebut sebagai model regresi *spatial autoregressive moving average* (SARMA).

$$\begin{aligned} y &= [y_1 \ y_2 \ \dots \ y_n]^T, \\ \beta &= [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p]^T, \\ \varepsilon &= [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]^T, \\ X &= \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \text{ dan} \end{aligned}$$

<sup>1</sup> Model regresi Tobit pertama kali diperkenalkan oleh James Tobin (1958), yang dijelaskan kembali oleh Long (1997), DeMaris (2004), Greene (2008), dan Lee (2010).

<sup>2</sup> Sebagaimana dijelaskan oleh Long (1997) dan Greene (2008).

$$W = \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ x_{21} & 0 & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix}$$

Persamaan (4) di atas menjadi bentuk model regresi spasial lag ketika  $\lambda = 0$ , yang menjelaskan terjadinya efek korelasi spasial antar lag variabel respon. Ketika  $\rho = 0$ , persamaan (4) menjadi bentuk model regresi spasial error, yang menjelaskan terjadinya efek korelasi spasial antar lag variabel prediktor. Statistik uji Lagrange digunakan untuk menentukan bentuk dari model regresi spasial (Anselin, 1999).

Untuk melihat ada atau tidaknya efek korelasi spasial lag, digunakan hipotesis  $H_0: \rho = 0$  lawan  $H_1: \rho \neq 0$  dengan statistik uji *Lagrange Multiplier* atau LM-lag Test sebagai berikut:

$$LM_{LAG} = \frac{(e^T W y / \hat{\sigma}^2)^2}{D} \quad (5)$$

Untuk melihat adanya atau tidaknya efek spasial error, digunakan hipotesis  $H_0: \lambda = 0$  lawan  $H_1: \lambda \neq 0$  dengan statistik uji LM-err Test sebagai berikut:

$$LM_{ERR} = \frac{(e^T W e / \hat{\sigma}^2)^2}{T} \quad (6)$$

dimana  $\hat{\sigma}^2 = e^T e / n$ ,  $e = y - X\hat{\beta}$ ,  $D = \left[ (WX\hat{\beta})^T (I - X(X^T X)^{-1} X^T) (WX\hat{\beta}) \right] / \sigma^2 + T$

, dan  $T = \text{tr} \{ (W^T + W)W \}$ . Statistik uji ini mengikuti distribusi asimtotik  $\chi^2_{(\alpha,1)}$ , sehingga  $H_0$  ditolak apabila  $LM > \chi^2_{(\alpha,1)}$  atau *p-value* lebih kecil dari nilai  $\alpha$ .

### Markov Chain Monte Carlo (MCMC)

MCMC merupakan suatu teknik metode simulasi yang membangkitkan sejumlah sampel dari distribusi data yang telah diketahui (Chib & Greenberg, 1996). Ide dasar dari teknik MCMC adalah daripada menghitung suatu fungsi kepadatan peluang  $p(\theta|y)$ , lebih baik mengambil sampel random dalam jumlah besar dari  $p(\theta|y)$  untuk mengetahui bentuk probabilitas tersebut secara tepat. Dengan ukuran sampel random yang cukup besar, nilai rata-rata dan standar deviasinya dapat dihitung secara akurat (Casella & George,

1992). LeSage (1999) menjelaskan bahwa algoritma MCMC *Gibbs sampler* akan memberi kemudahan estimasi parameter untuk model regresi Tobit spasial daripada harus memecahkan sejumlah persamaan integral pada metode maksimum likelihood. Metode MCMC *Gibbs sampler* bertujuan untuk mencari nilai estimasi dari  $\theta_i$  menggunakan suatu distribusi posterior bersyarat, dimana nilai  $\theta_{-i}$  lainnya diasumsikan telah diketahui. Distribusi posterior dari parameter  $\theta$  ditentukan melalui prinsip dari teorema Bayes yang dinyatakan oleh:

$$p(\theta|y) = \frac{L(\theta|y)p(\theta)}{m(y)} \propto L(\theta|y)p(\theta) \quad (7)$$

dimana  $L(\theta|y)$  merupakan fungsi likelihood dari  $y$ ,  $m(y)$  merupakan distribusi marginal dari  $y$  yang tidak melibatkan parameter, dan  $p(\theta)$  merupakan distribusi prior dari  $\theta$  yang diperoleh dari penelitian sebelumnya maupun berdasarkan kajian teoritis atas masalah yang sedang diteliti (Casella & Berger, 2002).

Hastings (1970) mengembangkan metode *Metropolis* untuk mencari estimasi parameter  $\theta^{(t)}$  melalui suatu nilai inisiasi awal  $\theta^{(0)}$ , yang didasari oleh distribusi kandidat  $f(\theta^{(*)}|\theta^{(t-1)})$  dimana nilai  $\theta^{(t-1)}$  diketahui. Nilai dari  $\theta^{(*)}$  dibangkitkan dari distribusi kandidat, kemudian terima  $\theta^{(*)}$  sebagai  $\theta^{(t)}$  jika  $p(\theta^{(*)}, \theta^{(t-1)}) > q$ . Dimana  $q \sim \text{Uniform}(0,1)$  dan  $p(\theta^{(*)}, \theta^{(t-1)}) =$

$$\min \left[ 1, \frac{p(\theta^{(*)}|y)f(\theta^{(t-1)}|\theta^{(*)})}{p(\theta^{(t-1)}|y)f(\theta^{(*)}|\theta^{(t-1)})} \right] \quad (8)$$

### Penelitian Sebelumnya

Penelitian yang menggunakan model regresi Tobit spasial di antaranya dilakukan oleh Langyintuo dan Mekuria (2008) yang menggunakan metode maksimum likelihood untuk membentuk model Tobit SARMA pada data petani di Mozambique. Kaliba (2002) mengembangkan model Tobit SARMA menggunakan modul aplikasi Maximum Likelihood 4 dari paket program GAUSS (dikembangkan oleh Aptech Systems, 1995) pada data pedesaan di



Tanzania. Sementara LeSage dan Pace (2009) menggunakan data simulasi yang dibangkitkan oleh Koop untuk membentuk model Tobit spasial menggunakan pendekatan Bayesian MCMC (*Markov Chain Monte Carlo*) dengan algoritma Gibbs Sampling. Adapun penelitian yang menggunakan model Tobit spasial pada data teknologi informasi dan komunikasi (TIK) belum pernah ditemukan.

Pembangunan TIK suatu negara memiliki hubungan yang positif dengan pertumbuhan ekonomi. Artinya, pembangunan TIK akan memberikan efek berantai kepada meningkatnya pertumbuhan ekonomi (Kominfo, 2010). Rao dan Pattnaik (2006) menyatakan bahwa pertumbuhan TIK telah membuka kesempatan bagi masyarakat untuk lebih memanfaatkan fasilitas pembangunan sosial ekonomi dan budaya secara lebih modern. Pembangunan TIK memberikan pengaruh ekonomi yang luas, baik secara langsung maupun tidak langsung, meningkatkan kesejahteraan dan pembangunan fasilitas sosial ekonomi (ITU, 2010). Akses internet merupakan indikator yang paling merepresentasikan tingkat pembangunan TIK suatu negara, selain pertumbuhan ekonomi sektor telekomunikasi, kepemilikan telepon seluler atau penguasaan komputer.

Menggunakan data dari 154 negara, Howard dan Mazaheri (2009) menemukan bahwa kesenjangan penggunaan TIK (telepon seluler, komputer, dan bandwidth internet) dipengaruhi oleh; investasi asing, perdagangan, jumlah penduduk, populasi perkotaan, literacy rate, konsumsi, telepon kabel, serta sembilan variabel lain yang menjelaskan regulasi pemerintah. Andonova dan Serrano (2007) menjelaskan bahwa perkembangan TIK dan pertumbuhan pemanfaatan internet lebih banyak dipengaruhi oleh faktor perhatian pemerintah dan regulasi yang berlaku di wilayah tersebut. Michailidis dkk. mengungkapkan bahwa pengguna internet di pedesaan Yunani dipengaruhi oleh tingkat pendapatan, harga dari akses internet, kepemilikan PC, tempat tinggal, serta variabel sosial demografi seperti; jenis kelamin, jumlah penduduk muda yang

tinggal satu rumah, umur, tingkat pendidikan, dan status pekerjaan (Michailidis, Partalidou, Nastis, Klavdianou, & Charatsari, 2011).

Berdasarkan hasil-hasil pada penelitian terdahulu, model regresi Tobit spasial yang dibangun pada penelitian ini akan menggunakan algoritma MCMC sebagai metode estimasi parameternya. Tingkat penggunaan internet di kabupaten/kota di Pulau Jawa digunakan sebagai variabel respon dengan variabel-variabel prediktor sebagai berikut; persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk usia muda, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki komputer, persentase rumah tangga yang memiliki telepon seluler, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

## **b. Metode Analisis**

### **Sumber Data dan Variabel Penelitian**

Sumber data yang digunakan dalam penelitian ini adalah data olahan hasil Susenas 2010 dan Podes 2008 yang dikumpulkan oleh Badan Pusat Statistik (BPS). Matriks penimbang spasial  $W$  disusun menggunakan metode *queen contiguity*, yaitu daerah kabupaten/kota yang saling berbatasan wilayah akan memiliki korelasi spasial sedangkan yang saling terpisah tidak memiliki korelasi. Nilai  $w_{i,j} = 1$  jika daerah  $i$  dan  $j$  saling berbatasan wilayah, menjadi  $w_{i,j} = 0$  jika tidak saling berbatasan. Adapun peta digital yang digunakan berdasarkan hasil kegiatan *updating* peta Sensus Penduduk 2010.

Objek penelitian yang dijadikan sebagai variabel respon adalah tingkat penggunaan internet di 118 kabupaten/kota di Pulau Jawa, yaitu persentase penduduk usia 5 tahun ke atas yang pernah mengakses internet dalam tiga bulan terakhir. Sensor diberikan kepada wilayah kabupaten/kota dengan persentase penduduk pengguna internet di atas 16 persen dengan menganggap nilai  $y_i = 16$  pada nilai  $y_i^* \geq 16$ . Adapun variabel prediktor yang digunakan adalah sebagai berikut:

- $X_1$  : Persentase penduduk yang tinggal di daerah perkotaan.
- $X_2$  : Persentase penduduk usia muda (13-24 tahun).
- $X_3$  : Persentase penduduk lulusan SMA ke atas.
- $X_4$  : Rata-rata lama sekolah.
- $X_5$  : Persentase rumah tangga yang memiliki komputer.
- $X_6$  : Persentase rumah tangga yang memiliki telepon seluler.
- $X_7$  : Persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

### Model Regresi Tobit Spasial

Model regresi Tobit spasial merupakan penerapan model regresi spasial pada data yang tersensor. Sehingga dengan menggabungkan persamaan (4) ke dalam (1), akan diperoleh suatu model umum regresi Tobit spasial sebagaimana berikut.

$$y_i = \begin{cases} y_i^* = (I - \rho W)^{-1} X\beta \\ \quad + (I - \rho W)^{-1} (I - \lambda W)^{-1} \varepsilon & \text{jika } y_i^* < \tau \\ \tau & \text{jika } y_i^* \geq \tau \end{cases} \quad (9)$$

dimana  $\tau$  merupakan suatu nilai konstanta batasan yang diberikan kepada  $y_i$  jika  $y_i^*$  tersensor,  $i = 1, 2, \dots, n$ . Persamaan  $y^* = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} (I - \lambda W)^{-1} \varepsilon$  merupakan model umum dari regresi spasial yang menjadi spasial lag ketika  $\lambda = 0$  dan menjadi spasial error ketika  $\rho = 0$ , dimana  $\varepsilon \sim N(0, \sigma^2 V)$  dengan  $\text{diag}(V)$  merupakan pengali yang menyatakan heterogenitas dari varians error. Pada kondisi homoskedastisitas, matriks  $V$  menjadi  $V = I$  (LeSage, 2000).

Apabila nilai  $\rho \neq 0$  dan  $\lambda = 0$ , persamaan (9) menjadi model regresi Tobit spasial lag sebagai berikut:

$$y_i = \begin{cases} \rho w_i^T y + x_i^T \beta + \varepsilon_i & \text{jika } y_i < \tau \\ \tau & \text{jika } y_i \geq \tau \end{cases} \quad (10)$$

Apabila nilai  $\rho = 0$  dan  $\lambda \neq 0$ , menjadi model regresi Tobit spasial error:

$$y_i = \begin{cases} \lambda w_i^T y + x_i^T \beta - \lambda (w_i^T X) \beta + \varepsilon_i & \text{jika } y_i < \tau \\ \tau & \text{jika } y_i \geq \tau \end{cases} \quad (11)$$

Pengujian heteroskedastisitas untuk menentukan matrik  $\text{diag}(V)$  menggunakan

statistik uji Breusch-Pagan (BP) dengan hipotesis sebagai berikut:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

(homoskedastisitas)

$H_1$ : minimal ada satu  $\sigma_i^2 \neq \sigma^2$  (heteroskedastisitas)

Nilai dari BP-Test adalah sebagai berikut:

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{f} \quad (12)$$

dengan elemen vektor  $\mathbf{f}$  adalah

$f_i = \left( \frac{e_i^2}{\hat{\sigma}^2} - 1 \right)$ , dimana  $e_i$  adalah residual observasi ke- $i$  hasil regresi linier,  $i = 1, 2, \dots, n$ ,  $\hat{\sigma}^2 = \mathbf{e}^T \mathbf{e} / n$ ,  $\mathbf{X}$  adalah matriks  $n \times (p + 1)$  dari observasi dengan elemen kolom pertama merupakan vektor satu, dan  $p$  adalah jumlah variabel prediktor (Breusch & Pagan, 1979).  $H_0$  ditolak apabila  $BP > \chi^2_{(\alpha, p)}$ .

### Melengkapi Data Tersensor

Estimasi parameter model regresi Tobit spasial dilakukan dengan asumsi awal bahwa variabel respon  $Y$  merupakan data dengan informasi yang lengkap, tidak tersensor, dan memiliki korelasi spasial. Padahal sesuai persamaan (1), data yang memiliki informasi lengkap adalah variabel respon  $Y^*$  yang mengikuti distribusi normal  $y_i^* \sim N(\mu, \sigma^2)$ . Nilai dari  $y_i^*$  ketika  $y_i^* \geq \tau$  merupakan observasi yang tidak diketahui atau dianggap sebagai variabel latent. Sehingga nilainya harus dilengkapi menggunakan suatu nilai  $y_i^* = z_i$ . Variabel respon  $Y$  yang lengkap kemudian didefinisikan sebagai berikut:

$$y_i = \begin{cases} y_i^* & \text{jika } y_i^* < \tau \\ z_i & \text{jika } y_i^* \geq \tau \end{cases} \quad (13)$$

Melengkapi observasi tersensor dilakukan dengan cara membangkitkan variabel random  $Z$  yang berdistribusi normal  $\mathbf{z} \sim N(\mu, \Sigma)$ . Nilai  $z_i$  merupakan elemen ke- $i$  dari vektor  $\mathbf{z} = [z_1 \ z_2 \ \dots \ z_n]^T$  yang digunakan untuk mengganti observasi

tersensor ketika  $z_i \geq \tau$ . Menurut LeSage (1999), nilai rata-rata dari  $\mathbf{z}$ <sup>3</sup>:

$$\boldsymbol{\mu} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} \quad (14)$$

dengan nilai varians dari  $\mathbf{z}$  adalah:

$$\boldsymbol{\Sigma} = (\mathbf{I} - \rho\mathbf{W})^{-1} (\mathbf{I} - \lambda\mathbf{W})^{-1} \sigma_\varepsilon^2 \mathbf{V} \times [(\mathbf{I} - \rho\mathbf{W})^{-1}]^T [(\mathbf{I} - \lambda\mathbf{W})^{-1}]^T \quad (15)$$

### Estimasi Parameter

Sesuai dengan asumsi awal bahwa variabel respon  $Y$  merupakan data dengan informasi yang lengkap, tidak tersensor dan memiliki korelasi spasial, maka hubungannya dengan variabel prediktor  $X$  diwakili oleh model regresi spasial sesuai dengan persamaan (4) di atas. LeSage (2000) dan Lacombe (2008) merumuskan distribusi posterior bersyarat dari masing-masing parameter adalah sebagai berikut:

$$\left(\frac{1}{\sigma^2} \boldsymbol{\varepsilon}^T \mathbf{V}^{-1} \boldsymbol{\varepsilon}\right) \sim \chi^2_{(n+4)} \quad (16)$$

$$\left\{\frac{1}{v_{ii}} \left(\frac{\varepsilon_i^2}{\sigma^2} + r\right)\right\} \sim \chi^2_{(r+1)} \quad (17)$$

$$p(\boldsymbol{\beta} | \sigma^2, \mathbf{V}, \rho, \lambda) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2} [\mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]^T \mathbf{V}^{-1} [\mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]\right\} \quad (18)$$

$$p(\rho, \lambda | \sigma^2, \mathbf{V}, \boldsymbol{\beta}) \propto |\mathbf{I} - \rho\mathbf{W}| |\mathbf{I} - \lambda\mathbf{W}| \times \exp\left\{-\frac{1}{2\sigma^2} \boldsymbol{\varepsilon}^T \mathbf{V}^{-1} \boldsymbol{\varepsilon}\right\} \quad (19)$$

dimana  $\mathbf{A} = (\mathbf{I} - \rho\mathbf{W})$ ,  $\mathbf{B} = (\mathbf{I} - \lambda\mathbf{W})$ , dan  $\boldsymbol{\varepsilon} = \mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

Estimasi parameter metode MCMC *Gibbs sampler* dilakukan dengan cara membangkitkan angka random yang mengikuti distribusi posterior bersyarat dari masing-masing parameter sebanyak jumlah iterasi yang diinginkan. Adapun algoritma *Metropolis within Gibbs* digunakan pada bentuk distribusi posterior yang tidak standar, yaitu untuk parameter  $\rho$  atau parameter  $\lambda$  (LeSage, 2000). Untuk menentukan layak atau tidaknya suatu

variabel prediktor dimasukkan ke dalam model, digunakan statistik uji Wald dengan hipotesis sebagai berikut:

$$H_0: \boldsymbol{\beta}_k = \mathbf{0}$$

$$H_1: \boldsymbol{\beta}_k \neq \mathbf{0}, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah:

$$W_k = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (20)$$

dimana  $\hat{\beta}_k$  merupakan parameter hasil estimasi, dan  $SE(\hat{\beta}_k) = \sqrt{\widehat{Var}(\hat{\beta}_k)}$ .

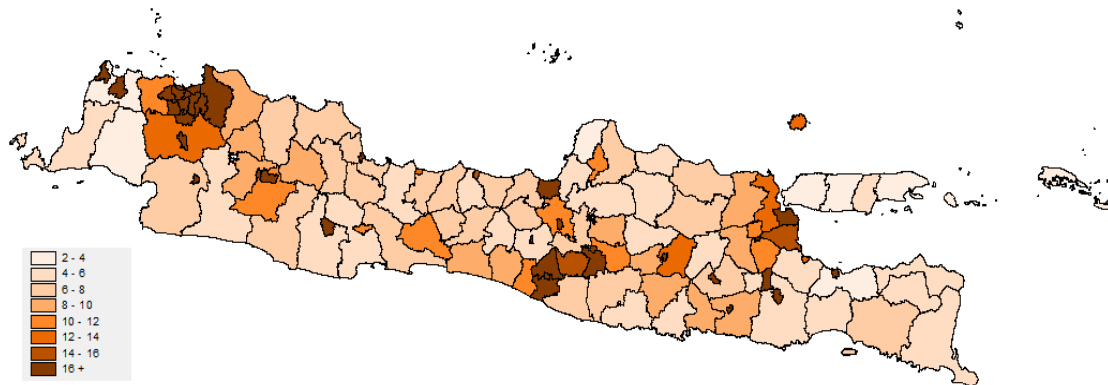
Keputusan untuk menolak  $H_0$  diambil apabila nilai  $|W_k| > Z_{\alpha/2}$  atau p-value lebih kecil dari nilai  $\alpha$ .

## HASIL DAN PEMBAHASAN

### Penggunaan Internet di Pulau Jawa

Tingkat penggunaan internet dilihat dari nilai persentase penduduk usia lima tahun ke atas yang pernah mengakses internet dalam dalam tiga bulan terakhir. Penggunaan internet terbesar pada tahun 2010 dapat ditemukan di kota pelajar Yogyakarta yaitu sebanyak 36,20 persen, disusul oleh Jakarta Selatan, Kabupaten Sleman, dan kota industri Tangerang Selatan, masing-masing sebanyak 33,20 persen, 29,52 persen dan 29,29 persen. Rata-rata penggunaan internet per kabupaten/kota di Pulau Jawa sebesar 12,02 persen. Wilayah dengan nilai penggunaan internet di sekitar rata-rata adalah Kabupaten Mojokerto, Bandung, Madiun, dan Kota Pasuruan. Sementara wilayah dengan persentase pengguna internet terkecil ada di Kabupaten Sampang dan Bangkalan, masing-masing sebesar 2,23 persen dan 2,54 persen.

<sup>3</sup> Informasi nilai rata-rata dan varians dari variabel random latent  $\mathbf{Z}$  ini dijelaskan pula dalam LeSage (2000) serta LeSage dan Pace (2009).



**Gambar 2. Penggunaan Internet Kabupaten/Kota di Pulau Jawa Tahun 2010**

Peta tematik penggunaan internet pada Gambar 2 memperlihatkan bahwa penggunaan internet kategori tinggi banyak terdapat pada wilayah perkotaan seperti DKI Jakarta dan sekitarnya, deretan Yogyakarta ke arah Surakarta, serta wilayah Bandung, Semarang, Surabaya dan sekitarnya. Tingginya penggunaan internet di kota-kota pusat pelajar, industri, dan bisnis tersebut kemudian diikuti oleh wilayah lain di sekitarnya. Wilayah yang bersinggungan langsung dengan kota-kota tersebut memiliki persentase pengguna internet yang sedikit lebih rendah, sementara wilayah berikutnya yang bersinggungan secara tidak langsung memiliki nilai yang lebih rendah lagi.

Untuk pemodelan regresi Tobit spasial, nilai persentase pengguna internet dari 34 kabupaten/kota dianggap tidak diketahui. Mereka adalah wilayah dengan tingkat penggunaan internet yang lebih besar dari 16 persen, yaitu sebagai batas minimal persentase pengguna internet yang ingin dicapai oleh suatu kabupaten/kota. Nilai-

nilai tingkat penggunaan internet yang tidak diketahui tersebut dianggap sama dengan 16 persen, sehingga diperoleh variabel persentase pengguna internet sebagai data tersensor. Hal ini sesuai dengan konsep pada persamaan (1) di atas.

Deskriptif data tingkat penggunaan internet sebagai variabel respon yang tersensor dapat dilihat pada Tabel 1 di bawah ini. Nilai maksimum dari persentase pengguna internet per kabupaten/kota sama dengan 16 persen, dengan rata-rata dan standar deviasi masing-masing sebesar 9,97 persen dan 4,63. Variabel yang memiliki variasi nilai terbesar adalah persentase penduduk perkotaan dengan standar deviasi sebesar 30,84 dan panjang range data dari 9,27 persen hingga 100 persen. Variabel rata-rata lama sekolah memiliki variasi terkecil dengan standar deviasi sebesar 1,52. Hal itu karena satuan dari variabel tersebut dalam ukuran tahun, sementara variabel lainnya dalam satuan persentase.

**Tabel 1. Deskriptif Variabel Penelitian**

Variabel Penelitian	Deskripsi	Minimum	Maksimum	Rata-rata	Standar Deviasi
(1)	(2)	(3)	(4)	(5)	(6)
$Y$	Persentase pengguna internet	2,23	16,00	9,968	4,626
$X_1$	Persentase penduduk perkotaan	9,27	100,00	57,991	30,837
$X_2$	Persentase penduduk usia muda	12,44	25,00	17,395	2,567
$X_3$	Persentase penduduk lulusan SMA ke atas	5,30	50,26	21,349	11,105
$X_4$	Rata-rata lama sekolah	4,21	11,55	8,052	1,518
$X_5$	Persentase rumah tangga yang memiliki komputer	2,20	40,22	11,433	9,127
$X_6$	Persentase rumah tangga memiliki telepon seluler	44,87	94,89	71,923	12,050
$X_7$	Persentase desa/kelurahan mendapat sinyal telepon	50,00	100,00	88,892	10,026

Pemodelan regresi linier berganda dengan metode *ordinary least squared* (OLS) dilakukan untuk menjelaskan hubungan antara variabel prediktor terhadap tingkat penggunaan internet di Pulau Jawa. Pada derajat kepercayaan 95 persen, hasil uji parameter hanya menghasilkan dua dari tujuh variabel prediktor yang berpengaruh terhadap variabel respon. Nilai *variance inflation factor* (VIF) yang sangat tinggi pada variabel  $X_3$  dan  $X_4$  mengindikasikan

adanya kondisi multikolinieritas antar variabel prediktor. Meskipun menghasilkan model yang fit dengan nilai koefisien determinasi yang sangat tinggi, model regresi linier berganda yang diperoleh kurang tepat untuk digunakan sebagai dasar analisis. Hal ini karena tidak terpenuhinya asumsi non multikolinieritas dan banyaknya informasi dari variabel prediktor yang terbuang.

**Tabel 2. Model Regresi Linier Berganda dan Nilai *Variance Inflation Factor* (VIF)**

Parameter	Koefisien	Uji Parameter		Statistik VIF
		Statistik Uji t	<i>p-Value</i>	
(1)	(2)	(3)	(4)	(5)
$\hat{\beta}_0$	-13,729	-4,158	0,000	–
$\hat{\beta}_1$	0,022	1,960	0,053	5,705
$\hat{\beta}_2$	-0,022	-0,298	0,766	1,701
$\hat{\beta}_3$	0,107	1,585	0,116	27,368
$\hat{\beta}_4$	0,797	1,936	0,055	18,952
$\hat{\beta}_5$	-0,026	-0,531	0,597	9,827
$\hat{\beta}_6$	0,083	2,671	0,009	6,872
$\hat{\beta}_7$	0,094	4,582	0,000	2,072
<i>Analysis of Variance</i> (uji $F$ )	–	132,47	0,000	–
Koefisien Determinasi ( $R^2$ )	0,894	–	–	–

Menurut Gujarati (2004), kondisi multikolinieritas dapat ditangani dengan cara melakukan pemilihan variabel, baik dengan menambah variabel prediktor baru maupun dengan mengurangi yang sudah ada. Cara lain yang dapat dilakukan adalah dengan menggunakan model lain yang lebih sesuai untuk menjelaskan hubungan antara variabel prediktor terhadap respon. Model regresi Tobit spasial lebih tepat digunakan untuk menjelaskan faktor-faktor yang mempengaruhi tinggi serta keragaman rendahnya penggunaan internet antar kabupaten/kota di Pulau Jawa yang dipengaruhi secara spasial oleh daerah-daerah di sekitarnya.

### Pemodelan Regresi Tobit Spasial

Berdasarkan persamaan (9) di atas, model regresi Tobit spasial untuk penggunaan internet kabupaten/kota di Pulau Jawa adalah sebagai berikut:

$$y_i = \begin{cases} \rho w_i^T y + \lambda w_i^T y + (\rho W^T \lambda W)^T y \\ + x_i^T \beta - \lambda (w_i^T X \beta) + \varepsilon_i & \text{jika } y_i < 16 \\ 16 & \text{jika } y_i \geq 16 \end{cases} \quad (21)$$

Pemodelan Tobit spasial dilakukan dengan terlebih dahulu menguji efek korelasi spasial dan heteroskedastisitas, untuk menentukan bentuk korelasi spasial yang ada serta kondisi heteroskedastisitas yang akan diperhitungkan dalam algoritma MCMC. Metode *backward elimination* dilakukan untuk menghasilkan model regresi Tobit spasial terbaik dengan cara menggugurkan

satu persatu variabel yang tidak berpengaruh secara nyata (Draper & Smith, 1998).

Hasil uji efek korelasi spasial dan heteroskedastisitas memperlihatkan bahwa bentuk model yang akan dibentuk adalah regresi Tobit spasial lag dalam kondisi homoskedastisitas. Menggunakan hipotesis  $H_0: \rho = 0$ , nilai statistik uji LM-lag memperlihatkan kesimpulan untuk menolak  $H_0$  pada derajat kepercayaan 95 persen. Sementara nilai statistik uji LM-err memperlihatkan bahwa  $H_0: \lambda = 0$  gagal ditolak pada nilai  $\alpha = 0,05$ . Adapun hasil uji heteroskedastisitas yang menggunakan statistik uji Breusch Pagan menyimpulkan untuk tidak menolak hipotesis  $H_0$ : homoskedastisitas pada derajat kepercayaan 95 persen. Kondisi tersebut berlaku pada setiap tahapan pembentukan model menggunakan metode *backward elimination*.

Estimasi parameter dilakukan dengan terlebih dahulu membangkitkan sejumlah angka random  $z \sim N(\mu, \Sigma)$ , di mana nilai  $z_i$  ketika  $z_i \geq 16$  digunakan untuk mengganti data yang tersensor. Algoritma MCMC digunakan pada data yang telah lengkap untuk estimasi parameter model regresi Tobit spasial lag, yaitu *Gibbs sampler* untuk estimasi  $\hat{\sigma}^2$  dan  $\hat{\beta}$  serta *Metropolis within Gibbs* untuk estimasi  $\hat{\rho}$ . Nilai  $V = I$  digunakan karena kondisi homoskedastisitas yang terpenuhi. Eliminasi variabel diputuskan berdasarkan p-value dari statistik uji Wald yang lebih besar daripada nilai  $\alpha = 0,05$ .

**Tabel 3. Pengujian Efek Korelasi Spasial, Heteroskedastisitas, dan Hasil Estimasi Parameter untuk Pemodelan Regresi Tobit Spasial menggunakan Metode *Backward Elimination***

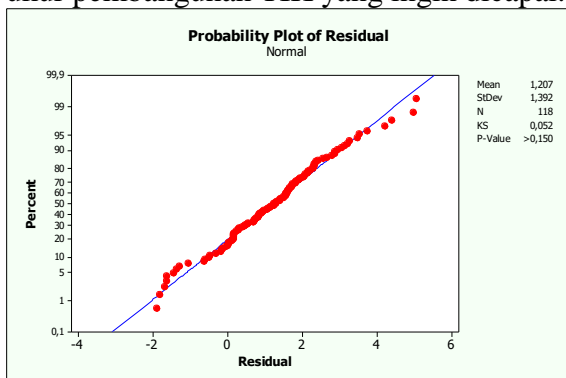
Statistik Uji/ Parameter	Tahap I (7 Variabel)		Tahap II (6 Variabel)		Tahap III (5 Variabel)	
	Nilai	p-value	Nilai	p-value	Nilai	p-value
(1)	(2)	(3)	(4)	(5)	(6)	(7)
LM-lag	12,1942	0,0000	12,2410	0,0000	11,8122	0,0000
LM-err	1,5595	<b>0,2117</b>	1,5765	<b>0,2093</b>	1,4396	<b>0,2302</b>
Breusch Pagan	9,7859	<b>0,2010</b>	9,4602	<b>0,1493</b>	7,9076	<b>0,1614</b>
$\hat{\rho}$	-0,1662	0,0000	-0,1693	0,0000	-0,1700	0,0000
$\hat{\beta}_0$	-13,9281	0,0000	-13,8073	0,0000	-13,7268	0,0000
$\hat{\beta}_1$	0,0115	0,0000	0,0122	0,0000	0,0117	0,0000
$\hat{\beta}_2$	0,0107	<b>0,2779</b>				
$\hat{\beta}_3$	0,1549	0,0000	0,1529	0,0000	0,1435	0,0000

$\hat{\beta}_4$	0,7565	0,0000	0,7574	0,0000	0,7548	0,0000
$\hat{\beta}_5$	-0,0143	<b>0,2102</b>	-0,0139	<b>0,2067</b>		
$\hat{\beta}_6$	0,0918	0,0000	0,0929	0,0000	0,0931	0,0000
$\hat{\beta}_7$	0,0831	0,0000	0,0828	0,0000	0,0827	0,0000
$\hat{\sigma}^2$	1,9561	–	1,9671	–	1,9580	–
$R^2$	0,8420	–	0,8403	–	0,8394	–

Variabel prediktor yang tersisih dari model adalah  $X_2$  (persentase penduduk usia muda) dan  $X_5$  (persentase rumah tangga yang memiliki komputer). Pada tahap akhir *backward elimination* diperoleh lima variabel prediktor yang secara nyata mempengaruhi keragaman penggunaan internet di Pulau Jawa. Berdasarkan persamaan (21) dan hasil estimasi parameter yang disajikan pada tabel 3 di atas, model regresi Tobit spasial lag yang terbentuk adalah:

$$\hat{y}_i = \begin{cases} -13,727 - 0,17 \sum_{j=1, j \neq i}^n w_{ij} y_j + 0,012 x_{1i} \\ + 0,144 x_{3i} + 0,755 x_{4i} + 0,093 x_{6i} + 0,083 x_{7i} ; y_i < 16 \\ 16 ; y_i \geq 16 \end{cases} \quad (22)$$

Model ini digunakan untuk menjelaskan faktor-faktor dan daerah di sekitarnya yang mempengaruhi tingkat penggunaan internet di suatu kabupaten/kota, ketika nilainya lebih kecil dari 16 persen. Adapun untuk kabupaten/kota dengan tingkat penggunaan internet yang tinggi dianggap sebagai tolok ukur pembangunan TIK yang ingin dicapai.



Gambar 3. Grafik Plot Probabilitas Normal dari Residual Model Regresi Tobit Spasial Lag

Grafik plot probabilitas normal dari residual digunakan untuk menguji asumsi normalitas dari error model. Plot residual terlihat berada di sekitar garis probabilitas

normal, sehingga dapat disimpulkan bahwa asumsi normalitas dari error model terpenuhi secara nyata. Asumsi berikutnya terkait masalah varians error yang homogen dianggap telah terpenuhi, mengingat hasil uji Breusch Pagan sebelumnya yang memperlihatkan tidak terpenuhinya kondisi heteroskedastisitas serta proses simulasi MCMC yang didasari oleh kondisi homoskedastisitas. Adapun asumsi tidak adanya autokorelasi dalam error juga telah terpenuhi, karena model yang terbentuk bukanlah model regresi Tobit spasial error.

### Interpretasi Model

Persamaan (22) di atas menjelaskan bahwa untuk kabupaten/kota dengan persentase pengguna internet yang kurang dari 16 persen, tingkat penggunaan internet di wilayah tersebut dipengaruhi oleh daerah lain di sekitarnya serta variabel-variabel sebagai berikut:

- Persentase penduduk yang tinggal di perkotaan ( $X_1$ ). Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase penduduk yang tinggal di perkotaan sebesar satu persen akan mengakibatkan kenaikan persentase pengguna internet sebesar 0,012 persen.
- Persentase penduduk lulusan SMA ke atas ( $X_3$ ). Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase penduduk lulusan SMA ke atas sebesar satu persen akan mengakibatkan kenaikan persentase pengguna internet sebesar 0,144 persen.
- Rata-rata lama sekolah ( $X_4$ ). Apabila variabel lain dianggap konstan, maka setiap kenaikan rata-rata lama sekolah selama satu tahun akan mengakibatkan kenaikan penggunaan internet suatu kabupaten/kota sebesar 0,755 persen.

- Persentase rumah tangga yang memiliki telepon genggam ( $X_6$ ). Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase rumah tangga yang memiliki telepon genggam sebesar satu persen akan mengakibatkan kenaikan persentase pengguna internet sebesar 0,093 persen.
- Persentase desa/kelurahan yang mendapat sinyal telepon seluler ( $X_7$ ).

Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase desa/kelurahan yang mendapat sinyal telepon seluler sebesar satu persen akan menyebabkan kenaikan penggunaan internet sebesar 0,083 persen.

Nilai koefisien determinasi memperlihatkan bahwa sebesar 83,94 persen variasi penggunaan internet di Pulau Jawa dijelaskan oleh lima variabel prediktor dalam model, sisanya oleh variabel lain. Pengaruh spasial lag dari daerah lain yang saling bersinggungan wilayah dapat terlihat pada model regresi Tobit spasial lag masing-masing kabupaten/kota, yaitu model  $\hat{y}_i$  ketika  $y_i < 16$ . Adapun untuk kabupaten/kota dengan kategori penggunaan internet yang tinggi, maka nilai persentase penduduk umur 5 tahun ke atas yang pernah mengakses internet dalam tiga bulan terakhir dianggap sama dengan 16 persen atau  $\hat{y}_i = 16$  ketika  $y_i \geq 16$ .

Persentase penduduk yang tinggal di perkotaan mengindikasikan tingkat kemajuan dan kelengkapan fasilitas umum di daerah tersebut. Adapun presentase penduduk lulusan SMA ke atas dan rata-rata lama sekolah mencerminkan kualitas sumber daya manusia di daerah tersebut. Dengan demikian, meningkatkan persentase pengguna internet dapat dilakukan melalui upaya peningkatan kualitas sumber daya manusia dari aspek pendidikan. Selain itu, pembangunan kelengkapan fasilitas umum di daerah pedesaan juga dapat mendorong tingkat penggunaan internet di kabupaten/kota.

Variabel karakteristik perangkat dan jaringan di daerah mengindikasikan pentingnya perkembangan teknologi telepon seluler bagi pertumbuhan internet. Berbagai

kemudahan akses internet yang disediakan melalui perangkat telepon genggam dan keluasan jaringan telepon seluler telah secara nyata mendorong peningkatan tingkat penggunaan internet. Di sisi lain, penggunaan internet ternyata tidak secara nyata diakses melalui komputer atau didominasi oleh penduduk usia muda. Internet dapat diakses oleh siapapun dan melalui media apapun, terutama telepon seluler.

Selain dipengaruhi oleh kelima variabel di atas, tingkat penggunaan internet kabupaten/kota di Pulau Jawa juga dipengaruhi oleh daerah lain yang bersinggungan wilayah. Sebagai contoh, tingkat penggunaan internet di Kabupaten Kepulauan Seribu dapat dijelaskan melalui model regresi Tobit spasial lag berikut ini:

$$\hat{y}_{3101} = -0.085(y_{3175} + y_{3603}) + x_{3101}^T \beta \quad (23)$$

dimana

$$x_{3101}^T = [1 \quad x_{1_{3101}} \quad x_{3_{3101}} \quad x_{4_{3101}} \quad x_{6_{3101}} \quad x_{7_{3101}}]$$

adalah vektor variabel prediktor dari Kepulauan Seribu dan

$$\beta = [-13,727 \quad 0,012 \quad 0,144 \quad 0,755 \quad 0,093 \quad 0,083]^T$$

adalah vektor parameter. Tingkat penggunaan internet di Kepulauan Seribu dipengaruhi pula oleh penggunaan internet di Kota Jakarta Utara dan Kabupaten Tangerang. Jika variabel lain dianggap konstan, maka penggunaan internet di Kepulauan Seribu adalah sebesar -0,085 kali dari gabungan penggunaan internet di Tangerang ( $y_{3603}$ ) dan Kota Jakarta Utara ( $y_{3175}$ ). Secara detail, masing-masing 84 model regresi Tobit spasial lag ketika nilai persentase pengguna internet kurang dari 16 persen dapat dilihat pada lampiran.

## KESIMPULAN DAN SARAN

Model regresi Tobit spasial merupakan suatu model regresi spasial yang diterapkan pada data tersensor, dengan bentuk model umum dari regresi Tobit spasial adalah:



$$y_i = \begin{cases} \rho w_i^T y + \lambda w_i^T y + (\rho W^T \lambda W)_i^T y \\ + x_i^T \beta - \lambda (w_i^T X \beta) + \varepsilon_i & \text{jika } y_i < \tau \\ \tau & \text{jika } y_i \geq \tau \end{cases} \quad (24)$$

Metode estimasi parameter yang dapat digunakan adalah *Markov Chain Monte Carlo* (MCMC) yang dilengkapi dengan algoritma *Gibbs sampler* dan *Metropolis within Gibbs*. Metode ini lebih mengedepankan teknik simulasi komputasi untuk membangkitkan sejumlah besar variabel random menggunakan pendekatan inferensia Bayesian.

Menggunakan penggunaan internet di Pulau Jawa sebagai studi kasus, diketahui bahwa model regresi Tobit Spasial lag menghasilkan informasi yang lebih kaya daripada model regresi linier berganda. Faktor-faktor yang mempengaruhi penggunaan internet kabupaten/kota di Pulau Jawa adalah persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki telepon genggam, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler. Selain dipengaruhi oleh kelima variabel tersebut, tingkat penggunaan internet kabupaten/kota di Pulau Jawa juga dipengaruhi oleh daerah lainnya yang bersinggungan wilayah.

Berdasarkan hasil penelitian yang telah diperoleh, pengembangan lebih lanjut dapat dilakukan dengan menggunakan *highest posterior density* (HPD) dan Bayes Faktor sebagai metode pengujian parameter dan model. Penelitian ini masih menggunakan matriks penimbang *queen contiguity*, sehingga pada penelitian selanjutnya dapat dikembangkan menggunakan matriks penimbang lain misalnya jarak. Lebih lanjut, metode MCMC *Gibbs sampler* untuk pemodelan regresi Tobit spasial ini dapat digunakan untuk data dan kasus lain yang lebih aplikatif.

#### DAFTAR PUSTAKA

Andonova, V., & Serrano, L. D. 2007. *Political Institutions and the*

*Development of Telecommunications*. Bonn: IZA Discussion Paper.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. 1999. *Spatial Econometrics*. Dallas: University of Texas.

BPS. 2011. *Sensus Penduduk 2010*. Hämtat från Sensus Penduduk 2010: <http://sp2010.bps.go.id/index.php/site/index> (diakses 4 November 2011)

BPS. 2011. *Statistik Komunikasi dan Teknologi Informasi Tahun 2010*. Jakarta: Badan Pusat Statistik.

Breusch, T., & Pagan, A. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, Vol. 47, No. 5, 1287-1294.

Casella, G. dan Berger, R. 2002. *Statistical Inference*. Duxbury, Thomson Learning.

Casella, G. dan George, E. I. 1992. Explaining the Gibbs Sampler. *The American Statistician*, Vol. 46, No. 3, 167-335.

Chib, S. dan Greenberg, E. 1996. Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometrics Theory*, Vol. 12, 409-431.

DeMaris, A. 2004. *Regression with Social Data: Modelling Continuous and Limited Response Variable*. New Jersey: John Wiley and Sons, Inc.

Draper, N. R. dan Smith, H. 1998. *Applied Regression Analysis*. New York: John Wiley and Sons, Inc.

Fischer, M. M. dan Getis, A. 2010. *Handbook of Applied Spatial Analysis: Software Tools, Methods, and Application*. New York: Springer.

Greene, W. H. 2008. *Econometric Analysis, Sixth Edition*. New York: Pearson - Prentice Hall.

Hastings, W. 1970. Monte Carlo Sampling Methods using Markov Chains and Their Applications. *Biometrika*, Vol. 57, No. 1, 97-109.

Howard, P. N. dan Mazaheri, N. 2009. Telecommunications Reform, Internet Use, and Mobile Phone Adoption in

- Developing World. *World Development*, Vol. 37, No. 7, 1159-1169.
- ITU. 2010. *Measuring the Information Society*. Geneva: International Telecommunications Union.
- Kaliba, A. R. 2002. Participatory Evaluation of Community Based Water and Sanitation Programmes: The Case of Central Tanzania. *Dissertation*. Mahattan: Kansas State University.
- Kominfo. 2010. *Buku Putih Komunikasi dan Informatika Indonesia*. Jakarta: Pusat Data Kementerian Komunikasi dan Informatika.
- Lacombe, D. J. (2008, Juli 24). *An Introduction to Bayesian Inference in Spatial Econometrics*. Hämtat från <http://ssrn.com/abstract=1244261>. (diakses 13 November 2011).
- Langyintuo, A. S. dan Mekuria, M. 2008. Assessing the Influence of Neighborhood Effects on the Adoption of Improved Agricultural Technologies in Developing Agriculture. *AfJARE*, Vol. 2, No. 2, 151-169.
- Lee, M. J. 2010. *Micro-Econometrics: Methods of Moments and Limited Dependent Variables, Second Edition*. New York: Springer.
- LeSage, J. P. 1999. *The Theory and Practice of Spatial Econometrics*. Ohio: University of Toledo.
- LeSage, J. P. 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis*, Vol. 32, No. 1, 19-35.
- LeSage, J. dan Pace, R. K. 2009. *Introduction to Spatial Econometrics*. New York: CRC Press.
- Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. California: Sage Publications, Inc.
- Marsh, T. L., Mittelhammer, R. C., & Huffaker, R. G. 2000. Probit with Spatial Correlation by Field Plot: Potato Leafroll Virus Net Necrosis in Potatoes. *Journal of Agricultural, Biological, and Environmental Statistics*, Volume 5, Number 1, Pages 22-36.
- Michailidis, A., Partalidou, M., Nastis, S. A., Klavdianou, A. P. dan Charatsari, C. 2011. Who Goes Online? Evidence of Internet Use Patterns from Rural Greece. *Telecommunications Policy*, Vol. 35, 333-343.
- Rao, J. G. dan Pattnaik, S. 2006. Technology for Rural Development Role of Telecommunication Media in India. *Indian Media Studies Journal*, Vol. 1, No. 1, 85-92.
- Socialbaker. (2011, Agustus 17) . *Facebook Statistics by Country*. Hämtat från [www.socialbaker.com: http://www.socialbakers.com/facebook-statistics/?interval=last-3-months#chart-intervals](http://www.socialbakers.com/facebook-statistics/?interval=last-3-months#chart-intervals).
- Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, Vol. 26, No. 1, 24-36.

## LAMPIRAN

Model regresi Tobit spasial lag untuk 84 kabupaten/kota dengan nilai  $y_i < 16$ :

1. Kepulauan Seribu :  $\hat{y}_{3101} = -0,085(y_{3175} + y_{3603}) + x_{3101}^T \beta$
2. Bogor :  $\hat{y}_{3201} = -0,015(y_{3202} + y_{3203} + y_{3214} + y_{3215} + y_{3216} + y_{3271} + y_{3275} + y_{3276} + y_{3602} + y_{3603} + y_{3674}) + x_{3201}^T \beta$
3. Sukabumi :  $\hat{y}_{3202} = -0,043(y_{3201} + y_{3203} + y_{3272} + y_{3602}) + x_{3202}^T \beta$
4. Cianjur :  $\hat{y}_{3203} = -0,028(y_{3201} + y_{3202} + y_{3214} + y_{3217} + y_{3274} + y_{3605}) + x_{3203}^T \beta$
5. Bandung :  $\hat{y}_{3204} = -0,024(y_{3203} + y_{3205} + y_{3211} + y_{3213} + y_{3217} + y_{3273} + y_{3277}) + x_{3204}^T \beta$
6. Garut :  $\hat{y}_{3205} = -0,043(y_{3203} + y_{3204} + y_{3206} + y_{3211}) + x_{3205}^T \beta$

7. Tasikmala ya :  $\hat{y}_{3206} = -0,034(y_{3205} + y_{3207} + y_{3210} + y_{3211} + y_{3278}) + x_{3206}^T \beta$
8. Ciamis :  $\hat{y}_{3207} = -0,028(y_{3206} + y_{3208} + y_{3210} + y_{3278} + y_{3279} + y_{3301}) + x_{3207}^T \beta$
9. Kuningan :  $\hat{y}_{3208} = -0,034(y_{3207} + y_{3209} + y_{3210} + y_{3301} + y_{3329}) + x_{3208}^T \beta$
10. Cirebon :  $\hat{y}_{3209} = -0,034(y_{3208} + y_{3210} + y_{3212} + y_{3274} + y_{3329}) + x_{3209}^T \beta$
11. Majalengka :  $\hat{y}_{3210} = -0,028(y_{3206} + y_{3207} + y_{3208} + y_{3209} + y_{3211} + y_{3212}) + x_{3210}^T \beta$
12. Sumedang :  $\hat{y}_{3211} = -0,028(y_{3204} + y_{3205} + y_{3206} + y_{3210} + y_{3212} + y_{3213}) + x_{3211}^T \beta$
13. Indramayu :  $\hat{y}_{3212} = -0,043(y_{3209} + y_{3210} + y_{3211} + y_{3213}) + x_{3212}^T \beta$
14. Subang :  $\hat{y}_{3213} = -0,028(y_{3204} + y_{3211} + y_{3212} + y_{3214} + y_{3215} + y_{3217}) + x_{3213}^T \beta$
15. Purwakarta :  $\hat{y}_{3214} = -0,034(y_{3201} + y_{3203} + y_{3213} + y_{3215} + y_{3217}) + x_{3214}^T \beta$
16. Karawang :  $\hat{y}_{3215} = -0,043(y_{3201} + y_{3213} + y_{3214} + y_{3216}) + x_{3215}^T \beta$
17. Bandung Barat :  $\hat{y}_{3217} = -0,028(y_{3203} + y_{3204} + y_{3213} + y_{3214} + y_{3273} + y_{3277}) + x_{3217}^T \beta$
18. Kota Banjar :  $\hat{y}_{3279} = -0,085(y_{3207} + y_{3301}) + x_{3279}^T \beta$
19. Cilacap :  $\hat{y}_{3301} = -0,028(y_{3207} + y_{3208} + y_{3279} + y_{3302} + y_{3305} + y_{3329}) + x_{3301}^T \beta$
20. Banyumas :  $\hat{y}_{3302} = -0,024(y_{3301} + y_{3303} + y_{3304} + y_{3305} + y_{3327} + y_{3328} + y_{3329}) + x_{3302}^T \beta$
21. Purbalingga :  $\hat{y}_{3303} = -0,043(y_{3302} + y_{3304} + y_{3326} + y_{3327}) + x_{3303}^T \beta$
22. Banjarnegara :  $\hat{y}_{3304} = -0,028(y_{3302} + y_{3303} + y_{3305} + y_{3307} + y_{3325} + y_{3326}) + x_{3304}^T \beta$
23. Kebumen :  $\hat{y}_{3305} = -0,034(y_{3301} + y_{3302} + y_{3304} + y_{3306} + y_{3307}) + x_{3305}^T \beta$
24. Purworejo :  $\hat{y}_{3306} = -0,043(y_{3305} + y_{3307} + y_{3308} + y_{3401}) + x_{3306}^T \beta$
25. Wonosobo :  $\hat{y}_{3307} = -0,024(y_{3304} + y_{3305} + y_{3306} + y_{3308} + y_{3323} + y_{3324} + y_{3325}) + x_{3307}^T \beta$
26. Magelang :  $\hat{y}_{3308} = -0,019(y_{3306} + y_{3307} + y_{3309} + y_{3310} + y_{3322} + y_{3323} + y_{3371} + y_{3401} + y_{3404}) + x_{3308}^T \beta$
27. Boyolali :  $\hat{y}_{3309} = -0,019(y_{3308} + y_{3310} + y_{3311} + y_{3313} + y_{3314} + y_{3315} + y_{3322} + y_{3372} + y_{3404}) + x_{3309}^T \beta$
28. Klaten :  $\hat{y}_{3310} = -0,034(y_{3308} + y_{3309} + y_{3311} + y_{3403} + y_{3404}) + x_{3310}^T \beta$
29. Wonogiri :  $\hat{y}_{3312} = -0,028(y_{3311} + y_{3313} + y_{3403} + y_{3501} + y_{3502} + y_{3520}) + x_{3312}^T \beta$
30. Karanganyar :  $\hat{y}_{3313} = -0,024(y_{3309} + y_{3311} + y_{3312} + y_{3314} + y_{3372} + y_{3520} + y_{3521}) + x_{3313}^T \beta$
31. Sragen :  $\hat{y}_{3314} = -0,043(y_{3309} + y_{3313} + y_{3315} + y_{3521}) + x_{3314}^T \beta$
32. Grobogan :  $\hat{y}_{3315} = -0,021(y_{3309} + y_{3314} + y_{3316} + y_{3318} + y_{3319} + y_{3321} + y_{3322} + y_{3521}) + x_{3315}^T \beta$

- 33 Blora :  $\hat{y}_{3316} = -0,028(y_{3315} + y_{3317} + y_{3318} + y_{3521} + y_{3522} + y_{3523}) + \mathbf{x}_{3316}^T \boldsymbol{\beta}$   
 .
- 34 Rembang :  $\hat{y}_{3317} = -0,057(y_{3316} + y_{3318} + y_{3523}) + \mathbf{x}_{3317}^T \boldsymbol{\beta}$   
 .
- 35 Pati :  $\hat{y}_{3318} = -0,034(y_{3315} + y_{3316} + y_{3317} + y_{3319} + y_{3320}) + \mathbf{x}_{3318}^T \boldsymbol{\beta}$   
 .
- 36 Kudus :  $\hat{y}_{3319} = -0,043(y_{3315} + y_{3318} + y_{3320} + y_{3321}) + \mathbf{x}_{3319}^T \boldsymbol{\beta}$   
 .
- 37 Jepara :  $\hat{y}_{3320} = -0,057(y_{3318} + y_{3319} + y_{3321}) + \mathbf{x}_{3320}^T \boldsymbol{\beta}$   
 .
- 38 Demak :  $\hat{y}_{3321} = -0,034(y_{3315} + y_{3319} + y_{3320} + y_{3322} + y_{3374}) + \mathbf{x}_{3321}^T \boldsymbol{\beta}$   
 .
- 39 Semarang :  $\hat{y}_{3322} = -0,021(y_{3308} + y_{3309} + y_{3315} + y_{3321} + y_{3323} + y_{3324} + y_{3373} + y_{3374}) + \mathbf{x}_{3322}^T \boldsymbol{\beta}$   
 .
- 40 Temanggung :  $\hat{y}_{3323} = -0,043(y_{3307} + y_{3308} + y_{3322} + y_{3324}) + \mathbf{x}_{3323}^T \boldsymbol{\beta}$   
 .  
 ng
- 41 Kendal :  $\hat{y}_{3324} = -0,034(y_{3307} + y_{3322} + y_{3323} + y_{3325} + y_{3374}) + \mathbf{x}_{3324}^T \boldsymbol{\beta}$   
 .
- 42 Batang :  $\hat{y}_{3325} = -0,034(y_{3304} + y_{3307} + y_{3324} + y_{3326} + y_{3375}) + \mathbf{x}_{3325}^T \boldsymbol{\beta}$   
 .
- 43 Pekalonga :  $\hat{y}_{3326} = -0,034(y_{3303} + y_{3304} + y_{3325} + y_{3327} + y_{3375}) + \mathbf{x}_{3326}^T \boldsymbol{\beta}$   
 .  
 n
- 44 Pemasang :  $\hat{y}_{3327} = -0,043(y_{3302} + y_{3303} + y_{3326} + y_{3328}) + \mathbf{x}_{3327}^T \boldsymbol{\beta}$   
 .
- 45 Tegal :  $\hat{y}_{3328} = -0,043(y_{3302} + y_{3327} + y_{3329} + y_{3376}) + \mathbf{x}_{3328}^T \boldsymbol{\beta}$   
 .
- 46 Brebes :  $\hat{y}_{3329} = -0,028(y_{3208} + y_{3209} + y_{3301} + y_{3302} + y_{3328} + y_{3376}) + \mathbf{x}_{3329}^T \boldsymbol{\beta}$   
 .
- 47 Kota :  $\hat{y}_{3375} = -0,085(y_{3325} + y_{3326}) + \mathbf{x}_{3375}^T \boldsymbol{\beta}$   
 .  
 Pekalonga  
 n
- 48 Kota :  $\hat{y}_{3376} = -0,085(y_{3328} + y_{3329}) + \mathbf{x}_{3376}^T \boldsymbol{\beta}$   
 .  
 Tegal
- 49 Kulon :  $\hat{y}_{3401} = -0,043(y_{3306} + y_{3308} + y_{3402} + y_{3404}) + \mathbf{x}_{3401}^T \boldsymbol{\beta}$   
 .  
 Progo
- 50 Gunung :  $\hat{y}_{3403} = -0,034(y_{3310} + y_{3311} + y_{3312} + y_{3402} + y_{3404}) + \mathbf{x}_{3403}^T \boldsymbol{\beta}$   
 .  
 Kidul
- 51 Pacitan :  $\hat{y}_{3501} = -0,057(y_{3312} + y_{3502} + y_{3503}) + \mathbf{x}_{3501}^T \boldsymbol{\beta}$   
 .
- 52 Ponorogo :  $\hat{y}_{3502} = -0,024(y_{3312} + y_{3501} + y_{3503} + y_{3504} + y_{3318} + y_{3519} + y_{3520}) + \mathbf{x}_{3502}^T \boldsymbol{\beta}$   
 .
- 53 Trenggalek :  $\hat{y}_{3503} = -0,057(y_{3501} + y_{3502} + y_{3504}) + \mathbf{x}_{3503}^T \boldsymbol{\beta}$   
 .  
 k
- 54 Tulungagung :  $\hat{y}_{3504} = -0,034(y_{3502} + y_{3503} + y_{3505} + y_{3506} + y_{3518}) + \mathbf{x}_{3504}^T \boldsymbol{\beta}$   
 .  
 ng
- 55 Blitar :  $\hat{y}_{3505} = -0,043(y_{3504} + y_{3506} + y_{3507} + y_{3572}) + \mathbf{x}_{3505}^T \boldsymbol{\beta}$   
 .
- 56 Kediri :  $\hat{y}_{3506} = -0,028(y_{3504} + y_{3505} + y_{3507} + y_{3517} + y_{3518} + y_{3571}) + \mathbf{x}_{3506}^T \boldsymbol{\beta}$   
 .
- 57 Malang :  $\hat{y}_{3507} = -0,019(y_{3505} + y_{3506} + y_{3508} + y_{3513} + y_{3514} + y_{3516} + y_{3517} + y_{3573} + y_{3579}) + \mathbf{x}_{3507}^T \boldsymbol{\beta}$   
 .

- 58 Lumajang :  $\hat{y}_{3508} = -0,057(y_{3507} + y_{3509} + y_{3513}) + x_{3508}^T \beta$   
 .
- 59 Jember :  $\hat{y}_{3509} = -0,043(y_{3508} + y_{3510} + y_{3511} + y_{3513}) + x_{3509}^T \beta$   
 .
- 60 Banyuwani :  $\hat{y}_{3510} = -0,057(y_{3509} + y_{3511} + y_{3512}) + x_{3510}^T \beta$   
 .
- 61 Bondowoso :  $\hat{y}_{3511} = -0,043(y_{3509} + y_{3510} + y_{3512} + y_{3513}) + x_{3511}^T \beta$   
 .
- 62 Situbondo :  $\hat{y}_{3512} = -0,057(y_{3510} + y_{3511} + y_{3513}) + x_{3512}^T \beta$   
 .
- 63 Probolinggo :  $\hat{y}_{3513} = -0,024(y_{3507} + y_{3508} + y_{3509} + y_{3511} + y_{3512} + y_{3514} + y_{3574}) + x_{3513}^T \beta$   
 .
- 64 Pasuruan :  $\hat{y}_{3514} = -0,028(y_{3507} + y_{3513} + y_{3515} + y_{3516} + y_{3575} + y_{3579}) + x_{3514}^T \beta$   
 .
- 65 Sidoarjo :  $\hat{y}_{3515} = -0,034(y_{3514} + y_{3516} + y_{3525} + y_{3576} + y_{3578}) + x_{3515}^T \beta$   
 .
- 66 Mojokerto :  $\hat{y}_{3516} = -0,021(y_{3507} + y_{3514} + y_{3515} + y_{3517} + y_{3524} + y_{3525} + y_{3576} + y_{3579}) + x_{3516}^T \beta$   
 .
- 67 Jombang :  $\hat{y}_{3517} = -0,028(y_{3506} + y_{3507} + y_{3516} + y_{3518} + y_{3522} + y_{3524}) + x_{3517}^T \beta$   
 .
- 68 Nganjuk :  $\hat{y}_{3518} = -0,028(y_{3502} + y_{3504} + y_{3506} + y_{3517} + y_{3519} + y_{3522}) + x_{3518}^T \beta$   
 .
- 69 Madiun :  $\hat{y}_{3519} = -0,028(y_{3502} + y_{3518} + y_{3520} + y_{3521} + y_{3522} + y_{3577}) + x_{3519}^T \beta$   
 .
- 70 Magetan :  $\hat{y}_{3520} = -0,028(y_{3212} + y_{3213} + y_{3502} + y_{3519} + y_{3521} + y_{3577}) + x_{3520}^T \beta$   
 .
- 71 Ngawi :  $\hat{y}_{3521} = -0,024(y_{3213} + y_{3214} + y_{3215} + y_{3216} + y_{3519} + y_{3520} + y_{3522}) + x_{3521}^T \beta$   
 .
- 72 Bojonegoro :  $\hat{y}_{3522} = -0,024(y_{3216} + y_{3517} + y_{3518} + y_{3519} + y_{3521} + y_{3523} + y_{3524}) + x_{3522}^T \beta$   
 .
- 73 Tuban :  $\hat{y}_{3523} = -0,043(y_{3216} + y_{3217} + y_{3522} + y_{3524}) + x_{3523}^T \beta$   
 .
- 74 Lamongan :  $\hat{y}_{3524} = -0,034(y_{3516} + y_{3517} + y_{3522} + y_{3523} + y_{3525}) + x_{3524}^T \beta$   
 .
- 75 Gresik :  $\hat{y}_{3525} = -0,034(y_{3515} + y_{3516} + y_{3524} + y_{3526} + y_{3578}) + x_{3525}^T \beta$   
 .
- 76 Bangkalan :  $\hat{y}_{3526} = -0,057(y_{3525} + y_{3527} + y_{3578}) + x_{3526}^T \beta$   
 .
- 77 Sampang :  $\hat{y}_{3527} = -0,085(y_{3526} + y_{3528}) + x_{3527}^T \beta$   
 .
- 78 Pamekasan :  $\hat{y}_{3528} = -0,085(y_{3527} + y_{3529}) + x_{3528}^T \beta$   
 .
- 79 Sumenep :  $\hat{y}_{3529} = -0,170 y_{3528} + x_{3529}^T \beta$   
 .
- 80 Kota Pasuruan :  $\hat{y}_{3575} = -0,170 y_{3514} + x_{3575}^T \beta$   
 .
- 81 Pandeglang :  $\hat{y}_{3601} = -0,085(y_{3602} + y_{3604}) + x_{3601}^T \beta$   
 .
- 82 Lebak :  $\hat{y}_{3602} = -0,034(y_{3201} + y_{3202} + y_{3601} + y_{3603} + y_{3604}) + x_{3602}^T \beta$   
 .

83 Tangerang :  $\hat{y}_{3603} = -0,021(y_{3101} + y_{3174} + y_{3175} + y_{3201} + y_{3602} + y_{3604} + y_{3671} + y_{3674}) + \mathbf{x}_{3603}^T \boldsymbol{\beta}$   
.

84 Serang :  $\hat{y}_{3604} = -0,034(y_{3601} + y_{3602} + y_{3603} + y_{3672} + y_{3673}) + \mathbf{x}_{3604}^T \boldsymbol{\beta}$   
.

# **GENERALIZED MULTILEVEL LINEAR MODEL DENGAN PENDEKATAN BAYESIAN UNTUK PEMODELAN DATA PENGELUARAN PERKAPITA RUMAHTANGGA**

**Azka Ubaidillah<sup>1</sup>, Anang Kurnia<sup>2</sup>, Kusman Sadik<sup>2</sup>**

<sup>1</sup>Politeknik Statistika STIS, Jakarta

<sup>2</sup>Departemen Statistika, Institut Pertanian Bogor, Bogor

e-mail: <sup>1</sup>azka@stis.ac.id, <sup>2</sup>akstk29@gmail.com, <sup>2</sup>kusmansadik@gmail.com

## **Abstrak**

Data pengeluaran perkapita rumahtangga merupakan salah satu informasi penting sebagai pendekatan untuk mengukur tingkat kemakmuran dan kesejahteraan di suatu daerah. Data tersebut sangat diperlukan oleh pemerintah baik di pusat maupun daerah dalam merumuskan, melaksanakan dan mengevaluasi pelaksanaan pembangunan. Penelitian ini akan menganalisis model yang tepat untuk pemodelan data pengeluaran perkapita rumahtangga yang memperhitungkan kekhususan data BPS yang memiliki struktur hirarki dan pola distribusi data yang memiliki karakteristik skewed kanan. Pemodelan dilakukan dengan menggunakan distribusi Log-normal tiga parameter (LN3P) dan Log-logistik tiga parameter (LL3P) dengan struktur satu tingkat (unilevel) dan dua tingkat (multilevel). Proses pendugaan parameter dilakukan dengan metode Markov Chain Monte Carlo (MCMC) dan algoritma Gibbs Sampling. Hasil penelitian menunjukkan bahwa pada model unilevel, model LL3P lebih baik dari model LN3P. Sedangkan pada model multilevel, model LN3P lebih baik dari model LL3P. Hasil penelitian juga menunjukkan model terbaik untuk pemodelan data pengeluaran perkapita rumahtangga adalah model multilevel LN3P dengan intercept sebagai komponen berhirarki dengan nilai Deviance Information Criterion (DIC) terkecil.

**Kata kunci** : *Generalized Multilevel Linear Model*, LL3P, LN3P, MCMC, Pengeluaran perkapita rumahtangga.

## **Abstract**

*Household per capita expenditure data is one of the important information as an approach to measure the level of prosperity in an area. Such data is needed by the government, both at the central and regional levels in formulating, implementing and evaluating the implementation of development programs. This research is aimed at modeling the household per capita expenditure data which takes into account the specificity of BPS data which has a hierarchical structure, and data distribution pattern which has the right skewed characteristic. The modeling is done by using the three parameters of Log-normal distribution (LN3P) and the three parameters of Log-logistics (LL3P) with a single level (unilevel) and two levels (multilevel) structure. The parameter estimation process is done by Markov Chain Monte Carlo (MCMC) method and Gibbs Sampling algorithm. The results showed that on the unilevel model, the LL3P model is better than the LN3P model. While in multilevel model, LN3P model is better than LL3P model. The results also show that the best model for modeling household per capita expenditure data is the LN3P multilevel model with the smallest Deviance Information Criterion (DIC) value.*

**Keywords**: *Generalized Multilevel Linear Model*, LL3P, LN3P, MCMC, Household per capita expenditure.

## PENDAHULUAN

Salah satu tujuan Negara Republik Indonesia yang sekaligus menjadi amanat konstitusi sebagaimana termaktub dalam Pembukaan Undang-Undang Dasar 1945 adalah memajukan kesejahteraan umum. Untuk mewujudkan tujuan tersebut maka pemerintah melakukan perumusan perencanaan, melaksanakan pembangunan dan secara berkesinambungan melakukan evaluasi atas pelaksanaan pembangunan. Keberhasilan pembangunan secara umum diukur dari tingkat kemakmuran dan kesejahteraan rakyat. Salah satu pendekatan pengukuran tingkat kemakmuran dan kesejahteraan rakyat adalah pengeluaran perkapita rumahtangga.

Pada dasarnya faktor yang berpengaruh dalam masalah kemakmuran dan kesejahteraan dapat dikategorikan dalam dua hal pokok yaitu paradigma perilaku dan paradigma kebijakan (Atika dan Pirmansyah, 2011). Paradigma perilaku terkait dengan upaya masing-masing individu dan rumahtangga dalam mencapai kesejahteraan. Sedangkan paradigma kebijakan terkait dengan kondisi ekonomi, politik dan kebijakan pemerintah. Hal ini menunjukkan bahwa kondisi di luar rumahtangga itu sendiri juga mempengaruhi perbedaan tingkat kesejahteraan. Dengan demikian tingkat kesejahteraan rumahtangga di suatu wilayah dipengaruhi oleh faktor internal dan eksternal rumahtangga tersebut.

Pada umumnya, data di bidang sosial seperti data pengeluaran perkapita rumahtangga memiliki struktur data yang berhirarki dimana unit-unit di tingkat yang lebih rendah, yaitu rumahtangga tersarang (*nested*) atau terkelompok dalam unit-unit di tingkat yang lebih tinggi yaitu wilayah (desa/kelurahan/dsb). Oleh karena itu pemodelan pengeluaran perkapita rumahtangga hendaknya mempertimbangkan kombinasi antara karakteristik rumahtangga dan karakteristik wilayahnya.

Keterbatasan analisis statistik klasik untuk data berstruktur hirarki adalah tidak diperhitungkannya struktur hirarki data.

Teknik yang biasa digunakan adalah teknik agregasi, disagregasi dan pemodelan regresi menurut kelompok (Goldstein, 1995; Raudenbush dan Byrk, 2002; De Leeuw dan Meijer, 2008). Analisis dengan metode klasik hanya dilakukan pada salah satu tingkatan data yaitu tingkat individu dengan menggunakan teknik disagregasi, atau di tingkat kelompok dengan menggunakan teknik agregasi. Namun demikian, teknik klasik tersebut akan sangat berpengaruh dari sisi metodologi dan statistiknya (De Leeuw dan Meijer, 2008).

Menurut Goldstein (1995) dan Hox (1995), penggunaan model *multilevel* untuk data bestruktur hirarki memiliki beberapa kelebihan. Model *multilevel* dapat digunakan untuk menganalisis informasi dari beberapa tingkatan yang berbeda menjadi satu analisis statistik. Model *multilevel* memperhitungkan pengaruh variasi setiap tingkat data terhadap variasi respon. Hal ini memungkinkan peneliti untuk mengetahui variasi di setiap tingkatan data terhadap variasi respon.

Data pengeluaran rumahtangga memiliki karakteristik khusus dengan nilai yang selalu positif dan memiliki frekuensi tinggi untuk pengeluaran perkapita rumahtangga golongan menengah ke bawah, sedangkan untuk golongan rumahtangga menengah ke atas memiliki frekuensi yang relatif rendah. Distribusi yang sesuai untuk pola data tersebut dan banyak digunakan untuk analisis di bidang sosial ekonomi adalah distribusi Log-normal dan distribusi Log-logistik (Johnson, Kotz dan Balakrishnan, 1995b). Sesuai dengan karakteristik pengeluaran perkapita rumahtangga yang tidak pernah nol, maka digunakan distribusi Log-normal dan Log-logistik yang diperluas dengan penambahan satu parameter yang selanjutnya dikenal dengan distribusi Log-normal tiga parameter dan distribusi Log-logistik tiga parameter (Ismartini dkk, 2012).

Beberapa penelitian yang menggunakan konsep *multilevel* dalam analisis penelitiannya antara lain dilakukan oleh Ha dkk (2005) yang melakukan penelitian mengenai pemodelan dengan menggunakan metode *multilevel mixed*



linear untuk data survival CGD (Chronic Granulomatous Deseas). Zimmer dkk (2010) menggunakan model hirarki linier untuk memperkirakan perbedaan pedesaan/perkotaan dalam status fungsional transisi pada masyarakat Cina yang berumur 55 tahun ke atas selama 2 tahun dan memperkirakan derajat dimana tingkat sosial ekonomi individu dan komunitas merupakan penentu dalam menjelaskan perbedaan status tersebut. Sementara Anderson dan Wells (2010) menggunakan pendekatan Bayesian hirarki regresi pada pemodelan hirarki untuk data kelompok longitudinal penelitian hukum empiris.

Keterbatasan model hirarki sederhana seperti hirarki linier klasik dengan pendekatan *Maximum Likelihood* adalah apabila jumlah sampel yang kecil dan tidak seimbang terdapat pada model dengan tingkat yang lebih tinggi, maka inferensia statistiknya ada kemungkinan menjadi tidak dapat dipercaya (Raudenbush dan Bryk, 2002). Sehingga untuk mengatasi permasalahan tersebut maka digunakan pendekatan Bayesian pada model hirarki atau *Hierarchical Bayesian* (HB) dimana menurut Raudenbush dan Bryk (2002), model HB memiliki keuntungan yaitu mampu mengatasi permasalahan pemodelan hirarki untuk jumlah data yang sedikit dan tidak seimbang baik pada tingkat bawah maupun pada tingkat yang lebih tinggi.

Pada tahun 2012, Ismartini dkk mengembangkan model linier hirarki dengan pendekatan Bayesian untuk pemodelan data pengeluaran perkapita rumahtangga berbasis distribusi Log-normal tiga parameter (LN3P) dan distribusi Log-logistik tiga parameter (LL3P). Namun penelitian yang dilakukan oleh Ismartini belum dilakukan kajian yang lebih detail dari sisi konsep *Generalized Linear Model* (GLM).

Tujuan makalah ini yaitu memodelkan pengeluaran perkapita rumahtangga di Kota Jambi dengan menggunakan model *unilevel* dan *multilevel* berbasis pada distribusi LN3P dan distribusi LL3P. Proses pemodelan dimulai dengan membentuk model paling sederhana, yaitu model *unilevel* dengan tanpa kovariat, sampai model yang

kompleks yaitu model *multilevel*. Kemudian model-model tersebut dibandingkan untuk diperoleh model terbaik dengan menggunakan kriteria *Deviance Information Criterion* (DIC).

## TINJAUAN PUSTAKA

### 1. Model *Multilevel* Linier

Model *multilevel* merupakan model regresi yang mengakomodasi adanya struktur data hirarki atau data bersarang. Dalam struktur data hirarki ini, variabel respon diukur pada level mikro saja, sedangkan variabel prediktor diukur baik di level mikro maupun di level makro (Goldstein, 1995; Hox, 2010). Sesuai dengan konsep hirarki, maka model *multilevel* menghasilkan persamaan regresi bertingkat, yaitu koefisien regresi di tingkat lebih rendah diregresikan lagi di tingkat yang lebih tinggi. Ismartini dkk. (2012) pada penelitiannya tentang pemodelan pengeluaran per kapita rumahtangga, menjelaskan model *multilevel* pada level mikro menggambarkan hubungan antara variabel respon (pengeluaran per kapita rumahtangga) dengan beberapa variabel prediktor yang merupakan karakteristik rumahtangga. Sedangkan model pada level makro menjelaskan hubungan antara koefisien model pada level mikro dengan karakteristik wilayah.

Persamaan model mikro untuk setiap kelompok adalah sebagai berikut:

$$Y_{ik} = \beta_{0k} + \beta_{1k}X_{1ik} + \dots + \beta_{pk}X_{pik} + e_{ik} \quad (1)$$

dimana  $i = 1, 2, \dots, n_k$  dan  $k = 1, 2, \dots, K$ , atau jika dinyatakan dalam bentuk matriks menjadi:

$$\mathbf{y}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \mathbf{e}_k \quad (2)$$

dengan:

$$\mathbf{y}_k = [y_{1k} \ y_{2k} \ \dots \ y_{n_k k}]^T,$$

$$\mathbf{X}_k = \begin{bmatrix} 1 & x_{11k} & x_{21k} & \dots & x_{p1k} \\ 1 & x_{12k} & x_{22k} & \dots & x_{p2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_k k} & x_{2n_k k} & \dots & x_{pn_k k} \end{bmatrix},$$

$$\boldsymbol{\beta}_k = [\beta_{0k} \ \beta_{1k} \ \dots \ \beta_{pk}]^T,$$

$$\mathbf{e}_k = [e_{1k} \ e_{2k} \ \dots \ e_{n_k k}]^T,$$

Pembentukan model makro dilakukan dengan menjadikan koefisien regresi pada model mikro,  $\beta_{rk}$ ,  $r = 0, 1, 2, \dots, P$  dalam persamaan (2) sebagai variabel respon yang dijelaskan oleh model makro. Bentuk persamaan makro adalah sebagai berikut:

$$\beta_{rk} = \gamma_{0r} + \gamma_{1r}Z_{1k} + \dots + \gamma_{Lr}Z_{Lk} + u_{rk},$$

atau jika dinyatakan dalam bentuk matriks adalah

$$\boldsymbol{\beta}_r = \mathbf{Z}\boldsymbol{\gamma}_r + \mathbf{u}_r \quad (3)$$

dengan:

$$\boldsymbol{\beta}_r = [\beta_{r1} \ \beta_{r2} \ \dots \ \beta_{rK}]^T,$$

$$\mathbf{Z} = \begin{bmatrix} 1 & z_{11} & z_{21} & \dots & z_{L1} \\ 1 & z_{12} & z_{22} & \dots & z_{L2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1K} & z_{2K} & \dots & z_{LK} \end{bmatrix},$$

$$\boldsymbol{\gamma}_r = [\gamma_{0r} \ \gamma_{1r} \ \dots \ \gamma_{Lr}]^T,$$

$$\mathbf{u}_r = [u_{r1} \ u_{r2} \ \dots \ u_{rK}]^T.$$

Asumsi model hirarki dua tingkat tersebut adalah: (Hox, 1995; De Leeuw dan Kreft, 2006)

a. Level mikro

i. Residual bersifat independen atau  $Cov(e_{ik}, e_{i^*k}) = 0, i \neq i^*$

ii.  $e_{ik}$  berdistribusi normal

iii.  $E(\mathbf{e}_k) = \mathbf{0}$

iv.  $Var(\mathbf{e}_k) = \sigma_{[e]k}^2 \mathbf{I}_{n_k}$

dengan  $\sigma_{[e]k}^2$  adalah *varians* residual model mikro kelompok ke- $k$ .

b. Level makro

i. Residual bersifat independen atau  $Cov(u_{rk}, u_{r^*k^*}) = 0, r \neq r^*$  dan  $k \neq k^*$

ii.  $u_{rk}$  berdistribusi normal

iii.  $E(\mathbf{u}_r) = \mathbf{0}$

iv.  $Var(\mathbf{u}_r) = \mathbf{T}_r$ , dengan:

$$\mathbf{T}_r = \begin{bmatrix} \sigma_{[u]r11}^2 & \sigma_{[u]r21}^2 & \dots & \sigma_{[u]rK1}^2 \\ \sigma_{[u]r12}^2 & \sigma_{[u]r22}^2 & \dots & \sigma_{[u]rK2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{[u]r1K}^2 & \sigma_{[u]r2K}^2 & \dots & \sigma_{[u]rKK}^2 \end{bmatrix}$$

dengan  $\sigma_{[u]rkk}^2$  adalah *varians* residual model makro untuk koefisien regresi ke- $r$  pada kelompok ke- $k$ .

c. Antar level

Residual level mikro bersifat independen terhadap residual level makro atau  $Cov(e_{ik}, u_{rk}) = 0$ .

Persamaan gabungan dari persamaan (2) dan (3) adalah: (Raudenbush dan Byrk, 2002; De Leeuw dan Kreft, 2006; Goldstein, 1995)

$$\mathbf{y}_k = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma} + \mathbf{X}_k \mathbf{u}_k + \mathbf{e}_k \quad (4)$$

dengan:

$\mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma}$  adalah komponen tetap (deterministik) dan  $\mathbf{X}_k \mathbf{u}_k + \mathbf{e}_k$  adalah komponen random (stokastik) dalam model *multilevel*,

$$E(\mathbf{y}_k) = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma},$$

$$Var(\mathbf{y}_k) = \mathbf{X}_k \mathbf{T} \mathbf{X}_k^T + \sigma_k^2 \mathbf{I}_{n_k}.$$

## 2. Generalized Linear Model

*Generalized Linear Model* (GLM) merupakan perluasan dari *Linear Model*. Terdapat tiga komponen dalam GLM yaitu komponen random, komponen sistematis dan komponen fungsi penghubung. Komponen random mencakup variabel yang mempunyai distribusi yang termasuk dalam keluarga eksponensial, misalnya binomial, poisson, normal, gamma, dsb. Komponen sistematis adalah kombinasi linier dari fungsi nilai harapan komponen acak dengan kovariatnya. Sedangkan fungsi penghubung merupakan bentuk dari fungsi parameter natural. Formula distribusi keluarga eksponensial adalah (Dobson, 2002):

$$f(y; \theta) = a(y)b(\theta)\exp[c(\theta)d(y)] \quad (5)$$

Pada penelitian ini digunakan asumsi pola distribusi LN3P dan distribusi LL3P. Jika  $Y$  adalah variabel random yang berdistribusi Log-normal 3 parameter (LN3P) yang dinotasikan  $Y \sim LN3(\mu_{[y]}, \sigma_{[y]}^2, \lambda)$ , maka fungsi kepadatan peluang (pdf) dari  $Y$  dapat ditulis sebagai berikut (Johnson dkk., 1995a dan Aitchison, 1957):

$$f(y|\mu, \sigma^2, \lambda) = \frac{1}{(y-\lambda)\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log(y-\lambda) - \mu)^2\right], \quad (6)$$

dimana  $y > \lambda$ ,  $\lambda > 0$ ,  $\mu > 0$  dan  $\sigma > 0$ , dengan  $\mu$  adalah parameter lokasi (*location*),  $\sigma$  adalah parameter skala (*scale*) dan  $\lambda$  adalah parameter batas (*threshold*).

Persamaan (6) di atas selanjutnya dapat ditulis kembali dalam bentuk persamaan:

$$f(y|\mu, \sigma^2, \lambda) = \exp\left[\frac{\mu}{\sigma^2} \log(y-\lambda) - \frac{\mu}{2\sigma^2} - \frac{1}{2\sigma^2}(\log(y-\lambda))^2 - \log(y-\lambda) - \frac{1}{2} \log \sqrt{2\pi\sigma^2}\right] \quad (7)$$

Dari persamaan (7) tersebut terlihat bahwa distribusi LN3P termasuk keluarga eksponensial. Dengan cara yang sama untuk distribusi LL3P juga dapat dituliskan persamaan distribusi keluarga eksponensial.

Selanjutnya, nilai harapan dan varians model dapat dinyatakan sebagai berikut:

$$E(\ln(y-\lambda)) = \mu \quad (8)$$

$$Var(\ln(y-\lambda)) = \sigma^2 \quad (9)$$

Fungsi nilai harapan model merupakan fungsi identitas dengan persamaan:

$$f(\mu) = \mu = \mathbf{X}\boldsymbol{\beta} \quad (10)$$

Kemudian, untuk pemodelan GLM dengan struktur data berhirarki, maka GLM diperluas menjadi *Generalized Multilevel Linear Model* (GMLM) dengan formula fungsi nilai harapan sebagai berikut:

$$f(\mu) = \mu = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma} + \mathbf{X}_k \mathbf{u}_k \quad (11)$$

### 3. Pengujian Distribusi Data Respon

Pada penelitian ini digunakan uji Anderson-Darling untuk pengujian distribusi data variabel respon. Salah satu alasan digunakannya uji Anderson-Darling adalah bahwa uji Anderson-Darling lebih fleksibel daripada uji Kolmogorov-Smirnov (Anderson dan Darling, 1952). Hal ini karena uji Anderson-Darling merupakan modifikasi dari uji Kolmogorov-Smirnov dimana dilakukan penggabungan fungsi bobot sehingga uji Anderson-Darling menjadi lebih fleksibel.

Formula hipotesis uji Anderson-Darling adalah sebagai berikut:

$H_0$  : Data mengikuti suatu pola fungsi distribusi tertentu

$H_1$  : Data tidak mengikuti suatu pola fungsi distribusi tertentu

Menurut Anderson-Darling (1954), misal  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  dimana  $n$  adalah banyaknya pengamatan, maka statistik ujinya adalah sebagai berikut:

$$w_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\log u_i + \log(1-u_{n-i+1})], \quad (12)$$

dimana  $u_i = F(x_i)$  adalah fungsi distribusi kumulatif. Nilai kritis dari uji Anderson-Darling dirumuskan sebagai berikut:

$$CV = \frac{0,752}{1 + \frac{0,75}{n} + \frac{2,25}{n^2}}, \quad (13)$$

dimana  $CV$  (*Critical Value*) adalah nilai kritis.  $H_0$  ditolak jika  $w_n^2 > CV$ .

### 4. Analisis Bayesian

Metode Bayesian mengacu nama ilmuwan Thomas Bayes (1702-1761) yang menemukan perlakuan matematika untuk masalah *non trivial* dari inferensi Bayesian. Thomas Bayes menemukan suatu penyelesaian untuk kasus khusus yang kemudian dikenal dengan nama Teorema Bayesian. Selanjutnya Teorema Bayesian dipopulerkan oleh Matematikawan asal Prancis, Peirre-Simon Laplace dengan istilah peluang Bayesian.

Berbeda dengan teori statistika klasik (*frequentist*), analisis bayesian memperlakukan semua parameter yang tidak diketahui sebagai variabel random dan memiliki distribusi (Boldstad, 2007). Teorema bayesian didasarkan pada distribusi *posterior* yang merupakan perpaduan antara distribusi *prior* (informasi masa lalu sebelum dilakukan observasi) dan data observasi yang digunakan untuk menyusun fungsi *Likelihood* (Box dan Tiao, 1973). Hubungan distribusi *posterior* dengan distribusi *prior* dan *Likelihood* dapat ditulis sebagai berikut.

$$\text{Distribusi } posterior \propto \text{likelihood} \times \text{Distribusi } prior$$

Pada teorema Bayes, apabila terdapat parameter  $\theta$  yang diberikan oleh data observasi  $y$ , maka distribusi probabilitas untuk *posterior*  $\theta$  pada data  $y$  akan proporsional dengan perkalian antara distribusi *prior*  $\theta$  dan fungsi *Likelihood*  $\theta$  yang diberikan oleh data  $y$ . Secara matematis dapat ditulis sebagai berikut:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)}$$

$$f(\theta|y) \propto f(y|\theta)f(\theta) \quad (14)$$

dimana  $f(\theta|y)$  merupakan distribusi *posterior* yang proporsional dengan perkalian antara fungsi *Likelihood*  $f(y|\theta)$  dan distribusi *prior*  $f(\theta)$ .

#### 4.1 Markov Chain Monte Carlo

Untuk mendapatkan pendugaan parameter dari distribusi *posterior* melalui proses integrasi seringkali sulit dilakukan apabila melibatkan persamaan integral yang sangat kompleks. Oleh karena itu penyelesaian perhitungan pendugaan parameter seringkali dilakukan secara numerik, salah satunya adalah teknik *Markov Chain Monte Carlo* (MCMC). Menurut Carlin (1992) pendekatan MCMC sangat efektif untuk mengurangi beban komputasi dalam menyelesaikan persamaan integrasi yang kompleks. Scollnik (2011) mengemukakan bahwa metode ini

memungkinkan proses simulasi dengan mengambil sampel acak dari model stokastik yang sangat rumit.

Ide dasar dari MCMC adalah membangkitkan data sampel dari distribusi *posterior* sesuai proses *Markov Chain* dengan menggunakan simulasi *Monte Carlo* secara iteratif sehingga diperoleh kondisi yang konvergen terhadap *posterior* (Ntzoufras, 2009). Kondisi tersebut harus memenuhi sifat-sifat *Markov Chain* yang *strongly ergodic* (Boldstad, 2010; Tailor dan Carlin, 1998) yaitu:

- Irreducible*, artinya sampel parameter yang dibangkitkan melalui proses MCMC adalah bersifat random.
- Aperiodic*, artinya sampel parameter yang dibangkitkan tersebut tidak memiliki pola yang periodik dalam domain nilai tertentu.
- Recurrent*, artinya perubahan sampel parameter terjadi secara stabil dalam domain nilai tertentu.

#### 5. Hierarchical Bayesian (HB)

Raudenbush dan Byrk (2002) menyatakan bahwa secara umum proses pembentukan model HB dua tingkat diawali dengan membentuk model mikro sesuai persamaan (2) sebagai *Likelihood* dari data observasi yang memiliki parameter  $\beta$  dan  $\Omega$ , dengan  $\Omega = \text{Var}(y)$  sehingga fungsi *Likelihood* adalah  $f(y|\beta, \Omega)$ . Selanjutnya ditentukan prior dari parameter-parameter yang tidak diketahui dan dilakukan secara bertingkat yaitu *two stage prior* (untuk model hirarki dua tingkat). *Stage-1 prior* berdasarkan model makro sesuai persamaan (2) yang dinyatakan dalam notasi  $P_1(\beta|Y, T)$  dengan  $Y$  adalah matriks koefisien regresi model makro dan  $T$  adalah matriks varians kovarians dari  $u_{rk}$ .

Tahap selanjutnya adalah menentukan *stage-2 prior* yang dinyatakan dalam notasi  $P_2(\gamma, \Omega, T)$ . Dengan demikian distribusi *posterior* model HB adalah fungsi yang proporsional terhadap perkalian *Likelihood*, *stage-1 prior* dan *stage-2 prior* yang dinyatakan dalam notasi sebagai berikut:

$$P(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \mathbf{T} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\Omega}) P_1(\boldsymbol{\beta} | \boldsymbol{\gamma}, \mathbf{T}) P_2(\boldsymbol{\gamma}, \boldsymbol{\Omega}, \mathbf{T}) \quad (15)$$

Distribusi *prior* yang digunakan untuk masing-masing elemen vektor parameter model HB berdasarkan distribusi LN3 adalah sebagai berikut:

$$\begin{aligned} \lambda_k &\sim N(\mu_{[\lambda]k}, \sigma_{[\lambda]k}^2), \\ \beta_{rk} &\sim N(\mu_{[\beta]rk}, \tau_{[\beta]rk}), \\ \tau_{[\gamma]k} &\sim \text{Gamma}(a_{\tau_{[\gamma]k}}, b_{\tau_{[\gamma]k}}) \\ \gamma_{ir} &\sim N(\mu_{\gamma_{ir}}, \sigma_{\gamma_{ir}}^2). \end{aligned} \quad (16)$$

Penentuan nilai parameter distribusi *prior* dilakukan dengan cara gabungan antara *conjugate prior* dan *informative prior* (berdasarkan data). Proses penentuan nilai parameter tersebut dilakukan secara berulang-ulang dimana hasil (*posterior*) pada setiap percobaan dijadikan informasi untuk memperbaiki *prior* model sehingga diperoleh hasil estimasi yang konvergen dan memenuhi sifat-sifat *Markov Chain* yang *strongly ergodic* (Boldstad, 2010; Tailor dan Carlin, 1998), proses perbaikan *prior* berulang tersebut dikenal dengan *two-step* MCMC (Iriawan, 2012).

## 6. Pemodelan

Pemodelan data pengeluaran perkapita rumah tangga dengan distribusi LN3P dan distribusi LL3P dalam penelitian ini meliputi:

a. Model *unilevel*, yang meliputi:

a.1 GLM\_min, yaitu model GLM minimal (parameter hanya berupa intersep). Persamaan modelnya adalah:

$$f(\mu_i) = \beta_o \quad (17)$$

a.2 GLM\_mod, yaitu model GLM dengan kovariat ( $X_1, D_2, X_3, D_4$ ). Persamaan modelnya adalah:

$$f(\mu_i) = \beta_o + \beta_1 X_{1i} + \dots + \beta_4 D_{4i} \quad (18)$$

b. Model *multilevel*, yang meliputi:

b.1 GLMM\_Int, yaitu model GLM dengan parameter intersep yang bervariasi menurut wilayah dan

tanpa kovariat di model level kedua. Persamaan modelnya adalah:

$$f(\mu_{ik}) = \beta_o k + \beta_1 X_{1ik} + \dots + \beta_p X_{pik} + u_k \quad (19)$$

b.2 GLMM\_X, yaitu model GLM dengan parameter intersep dan koefisien kovariat di model level pertama yang bervariasi menurut wilayah dan tanpa kovariat di model level kedua. Persamaan modelnya adalah:

$$f(\mu_k) = \mathbf{X}_k \boldsymbol{\beta}_k + \mathbf{u}_k \quad (20)$$

b.3 GMLM\_Int, yaitu model GMLM dengan struktur hirarki pada parameter intersep (terdapat kovariat di model level kedua). Persamaan modelnya adalah:

$$f(\mu_{ik}) = \gamma_{00} + \gamma_{01} Z_{1k} + \dots + \gamma_{0l} Z_{lk} + \beta_1 X_{1ik} + \dots + \beta_p X_{pik} + u_k \quad (21)$$

b.4 Model GMLM\_X, yaitu model GMLM dengan struktur hirarki pada intersep dan kovariat di model level pertama dan terdapat kovariat di model level kedua. Persamaan modelnya adalah:

$$f(\mu_k) = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma} + \mathbf{X}_k \mathbf{u}_k \quad (22)$$

## 7. Pemilihan Model Terbaik

Strategi pemilihan model terbaik dalam penelitian ini menggunakan *Deviance Information Criterion* (DIC) yang dirumuskan sebagai berikut:

$$DIC = D(\bar{\boldsymbol{\theta}}_m, m) + 2p_m \quad (23)$$

dimana  $p_m$  adalah jumlah parameter efektif dari model. Menurut Ntzoufras (2009), DIC merupakan kriteria seleksi model terbaik yang hampir sama dengan *Akaike's Information Criterion* (AIC).

## METODOLOGI

## 1. Sumber Data

Data yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS), yaitu:

Data karakteristik individu dan rumahtangga di Kota Jambi yang berasal dari Susenas 2011 dan data karakteristik desa/kelurahan di Kota Jambi yang berasal dari Podes 2011.

## 2. Variabel Penelitian

Variabel respon yang digunakan yaitu data pengeluaran perkapita perbulan yang diperoleh dengan cara menghitung pengeluaran rumahtangga per bulan dibagi dengan jumlah anggota rumahtangganya. Sedangkan variabel prediktor dapat dilihat pada Tabel 1 dan Tabel 2 berikut ini.

**Tabel 1. Variabel Prediktor pada Model Mikro (Karakteristik rumahtangga)**

Var	Keterangan
X <sub>1</sub>	Jumlah anggota rumahtangga (ART)
D <sub>2</sub>	=1, jenis lantai terluas dari keramik/ubin/tegel/teraso =0, jenis lantai terluas dari lainnya
X <sub>3</sub>	Luas lantai perkapita
D <sub>4</sub>	=1, sumber energi untuk memasak dari LPG/gas =0, sumber energi untuk memasak dari lainnya

**Tabel 2. Variabel Prediktor pada Model Makro (Karakteristik wilayah Kelurahan)**

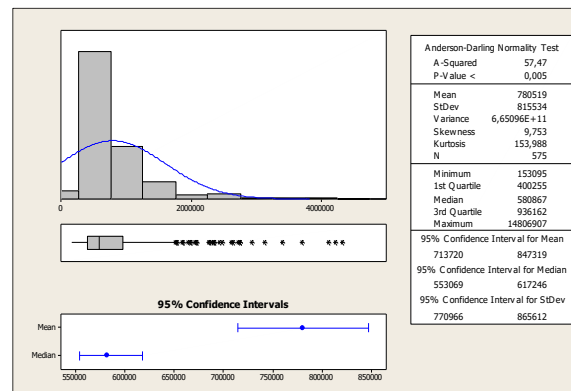
Var	Keterangan
Z <sub>1</sub>	Rasio SLTA per 10.000 penduduk
Z <sub>2</sub>	Rasio Puskesmas per 10.000 penduduk

## HASIL DAN PEMBAHASAN

### 1. Karakteristik Variabel Respon

Karakteristik variabel respon perlu diketahui untuk menentukan bentuk distribusi *likelihood* yang nantinya akan diterapkan pada pemodelan dengan pendekatan Bayesian. Gambar 1 berikut

dapat menjelaskan karakteristik variabel respon (pengeluaran perkapita rumahtangga perbulan).



**Gambar 1. Karakteristik data pengeluaran perkapita rumahtangga perbulan di Kota Jambi**

Dari Gambar 1 di atas terlihat bahwa terdapat dua ciri pola data pengeluaran perkapita rumah tangga perbulan yaitu nilai datanya positif dan memiliki ekor yang menceng ke kanan. Beberapa jenis distribusi yang memiliki kedua ciri tersebut diantaranya adalah distribusi Log-normal 2 parameter, Log-normal 3 Parameter, Log-logistik 3 parameter dan Weibull 3 parameter.

### 2. Distribusi Variabel Respon

Merujuk pada penelitian yang dilakukan oleh Ismartini dkk. (2012), maka pada penelitian ini dilakukan pengujian data pengeluaran perkapita rumahtangga perbulan di Kota Jambi dengan asumsi awal bahwa data mengikuti pola distribusi Log-normal tiga parameter (LN3P) dan distribusi Log-logistik tiga parameter (LL3P). Dari Tabel 3 terlihat bahwa distribusi data pengeluaran perkapita rumahtangga dapat dimodelkan dengan distribusi LN3P dan distribusi LL3P.

**Tabel 3. Hasil Uji Anderson Darling**

Distribusi	Banyaknya amatan	Nilai AD	Critical value
LN3P	575	0,44956	2,5018
LL3P	575	0,63205	2,5018

Selanjutnya, distribusi LN3P dan distribusi LL3P dimodelkan dengan metode *Generalized Linear Model* (GLM) karena kedua distribusi tersebut dapat dikategorikan sebagai keluarga eksponensial.

### 3. Pemodelan Pengeluaran Perkapita Rumahtangga

Proses pendugaan parameter model LN3P dan LL3P dengan pendekatan Bayesian dilakukan dengan menggunakan MCMC dan algoritma *Gibbs Sampling* dengan iterasi sebanyak 10.000, *thin* 10 dan *burn-in* sebanyak 1000 iterasi.

#### ➤ Pemodelan *Unilevel*

Hasil pemodelan *unilevel* ditampilkan pada Tabel 4 sebagai berikut:

**Tabel 4. Hasil Pendugaan Parameter Model *Unilevel* Distribusi LN3P dan LL3P**

Parameter	Model Berdistribusi LN3P		Model Berdistribusi LL3P	
	GLM_min	GLM_mod	GLM_min	GLM_mod
$\lambda$ (Threshold)	125000 (100.2)	124900 (100.7)	124900 (100.3)	124900.0 (102.1)
$\beta_0$	13.06 (0.03159)	12.99 (0.01987)	13.05 (0.03188)	12.99 (0.01989)
$\beta_1$	-	-0.09684 (0.004649)	-	0.3091 (0.01707)
$\beta_2$	-	0.3109 (0.01699)	-	0.01077 (0.00058)
$\beta_3$	-	0.01075 (0.000589)	-	0.2862 (0.0191)
$\beta_4$	-	0.2879 (0.01987)	-	0.3091 (0.01707)
DIC	16369.3	16081.0	16372.2	16061.0

Dari Tabel 4 terlihat bahwa model *unilevel* yang terbaik adalah model GLM\_mod berdistribusi LL3P, yaitu model dengan variabel  $X_1$ ,  $D_2$ ,  $X_3$ , dan  $D_4$  sebagai kovariat.

Model GLM\_mod dengan empat kovariat ( $X_1$ ,  $D_2$ ,  $X_3$ , dan  $D_4$ ) selanjutnya digunakan sebagai basis pemodelan *multilevel*.

#### ➤ Pemodelan *Multilevel*

Kondisi berbeda terjadi pada model *multilevel*, dimana model LN3P secara

umum lebih baik daripada model LL3P jika dilihat dari nilai DIC sebagaimana ditampilkan pada Tabel 5 berikut ini.

**Tabel 5. Nilai *Deviance* dan DIC untuk Pemodelan *Multilevel***

Distribusi	Model	<i>Deviance</i>	DIC
LN3P	GLMM_Int	15950	15975,5
	GLMM_X	15940	15966,4
	GMLM_Int*	<b>15870</b>	<b>15922.5</b>
	GMLM_X*	16630	16654,4
LL3P	GLMM_Int	15950	15992,6
	GLMM_X	15940	15979,9
	GMLM_Int*	<b>15870</b>	<b>15935.9</b>
	GMLM_X*	16660	16694,4

Keterangan :

\*) Model GMLM menggunakan empat kovariat di model level pertama ( $X_1$ ,  $D_2$ ,  $X_3$ , dan  $D_4$ ) dan dua kovariat di model level kedua ( $Z_1$  dan  $Z_2$ ). Penggunaan kovariat tersebut didasarkan pada signifikansi parameter dan kebaikan model

Tabel 5 menunjukkan bahwa model GMLM\_Int berdistribusi LN3P mempunyai nilai DIC yang paling kecil di antara model *multilevel* lainnya. Selain itu, jika dibandingkan dengan model *unilevel* terbaik (GLM\_mod2 berdistribusi LL3P), model GMLM\_Int berdistribusi LN3P mempunyai nilai DIC yang lebih kecil. Dengan demikian, model terbaik di antara semua model yang dihasilkan pada penelitian ini adalah model *multilevel* berdistribusi LN3P dengan struktur hirarki pada parameter intersep (GMLM\_Int berdistribusi LN3P).

### 4. Model Terbaik

Uraian berikut ini menjelaskan lebih detail model terbaik (GMLM\_Int berdistribusi LN3P) yang dihasilkan, meliputi pemilihan distribusi prior, hasil pendugaan parameter dan diagnosa model.

Distribusi prior yang digunakan untuk model GMLM\_Int LN3P adalah:

$$\begin{aligned}
\lambda &\sim N(124950, 1.0E-07), \\
\beta_{0k} &\sim N(\mu_{[\beta]0k}, \tau_{[\beta]0k}), \\
\tau_{[\gamma]k} &\sim \text{Gamma}(6.296, 0.687) \\
, \\
\tau_{[\beta]0k} &\sim \text{Gamma}(10, 10), \\
\gamma_{00} &\sim N(13.18, 167.3126021) \\
, \\
\gamma_{10} &\sim N(0.065936, 1721321.967)
\end{aligned} \tag{24}$$

Adapun hasil pendugaan parameter model terbaik ditampilkan pada Tabel 6 berikut ini:

**Tabel 6. Hasil Pendugaan Parameter Model GMLM\_Int Berdistribusi LN3P**

Parameter	Mean	sd	2.5%	97.5%	Sample
$\lambda$ (Threshold)	121100	3008.0	115100	1.27E+5	9001
$\gamma_{00}$	13.13	0.07217	12.99	13.27	9001
$\gamma_{01}$	0.06592	7.6E-4	0.06441	0.06741	9001
$\gamma_{02}$	0.06593	7.599E-4	0.06445	0.0674	9001
$\beta_1$	-0.0921	0.005058	-0.1022	-0.0824	9001
$\beta_2$	-	-	-	-	-
$\beta_3$	0.3136	0.01699	0.281	0.3469	9001
$\beta_4$	0.01114	6.07E-4	0.00992	0.0123	9001
$\beta_5$	0.279	0.01974	0.2399	0.3175	9001

Tabel 6 menunjukkan bahwa semua parameter model dapat diduga secara signifikan. Hal ini terlihat dari nilai *Credible Interval* (antara 2.5% sampai dengan 97.5%) yang tidak mengandung nilai nol. Secara matematis dapat ditulis dalam persamaan sebagai berikut:

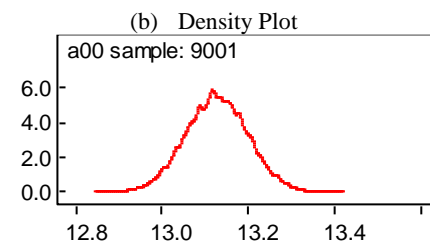
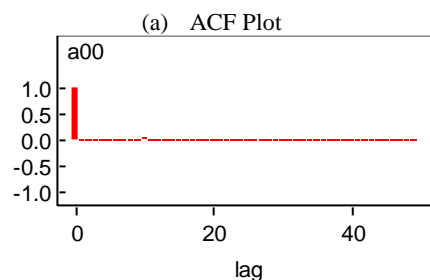
$$\begin{aligned}
f(\mu_{ik}) &= 13.13 + 0.066Z_{4k} + 0.066Z_{5k} \\
&\quad - 0.0921X_{1ik} + 0.3136D_{3ik} + 0.0111X_{4ik} \\
&\quad + 0.279D_{5ik} + u_k
\end{aligned}$$

Dari persamaan model terlihat bahwa kovariat model di level kedua (proporsi banyaknya SMU terhadap jumlah penduduk dan proporsi puskesmas terhadap jumlah penduduk) berpengaruh positif terhadap pengeluaran perkapita rumah tangga masing-masing sebesar  $\exp(0.066)=1.068$  kali. Adapun pengaruh kovariat model di level pertama terlihat bahwa variabel jumlah anggota rumahtangga berpengaruh negatif terhadap pengeluaran perkapita rumahtangga sebesar  $\exp(-0.0921)=0.91$  kali. Selanjutnya, variabel jenis lantai terluas dari keramik/ keramik/ ubin/ tegel/ teraso

berpengaruh terhadap pengeluaran perkapita rumahtangga sebesar  $\exp(0.3136)=1.368$  kali, kemudian variabel luas lantai perkapita berpengaruh terhadap pengeluaran perkapita rumahtangga sebesar  $\exp(0.0111)=1.011$  kali, dan variabel sumber energi untuk memasak berpengaruh terhadap pengeluaran perkapita rumahtangga sebesar  $\exp(0.279)=1.322$  kali.

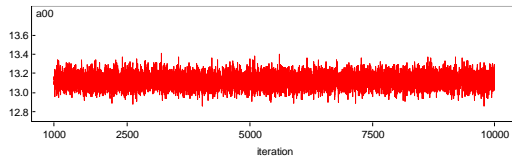
### ➤ Diagnosa Model

Diagnosa model terbaik ditampilkan dalam Gambar 2. Dari Gambar 2 terlihat bahwa sampel parameter yang dibangkitkan dari distribusi posterior untuk menduga parameter sudah memenuhi persyaratan pemodelan Bayesian. Hal ini terlihat dari plot ACF yang menunjukkan kondisi *Irreducible* (sampel parameter yang dibangkitkan melalui proses MCMC bersifat random) dan *Aperiodic* (tidak memiliki pola periodik). Kemudian, dari *Density* plot terlihat bahwa sampel parameter yang dibangkitkan melalui proses MCMC sudah sesuai dengan distribusi normal dan dari *Serial* plot terlihat bahwa pergerakan sampel parameter tercapai kondisi *recurrent* (stabil dalam domain nilai tertentu). Dengan terpenuhinya kondisi *Irreducible*, *Aperiodic* dan *Recurrent* maka dapat disimpulkan hasil pendugaan parameter model telah memenuhi sifat-sifat *Markov Chain* yang *strongly ergodic* sehingga model layak digunakan untuk keperluan analisis statistik.



(c) Serial Plot

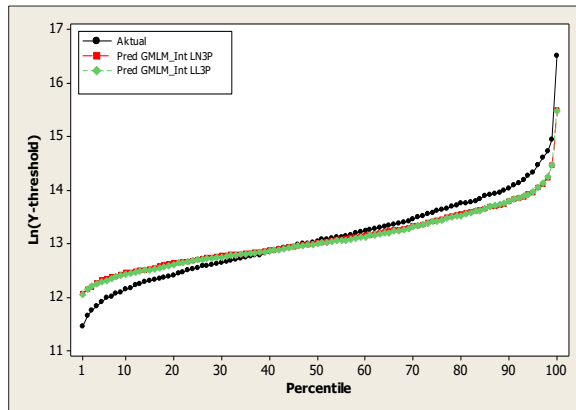




**Gambar 2. Diagnostic plot model GMLM\_Int distribusi LN3P untuk parameter  $\gamma_{00}$**

➤ **Prediksi Model**

Prediksi model GMLM\_Int berdistribusi LN3P dan distribusi LL3P ditampilkan dalam Gambar 3 sebagai berikut:



**Gambar 3 Plot persentil data aktual dan hasil prediksi model GMLM berdistribusi LN3P dan LL3P**

Dari Gambar 4.3 terlihat bahwa hasil prediksi antara model GMLM berdistribusi LN3P dan model GMLM berdistribusi LL3P mempunyai kemiripan dimana plot prediksi kedua model tersebut berhimpitan. Namun masih terlihat bahwa kemampuan prediksi kedua model masih perlu ditingkatkan, khususnya untuk persentil bawah dan persentil atas yang terdapat gap antara data aktual dengan hasil prediksi model.

**KESIMPULAN DAN SARAN**

**1. Kesimpulan**

Dari hasil analisa yang sudah diuraikan pada bab sebelumnya, maka dapat diperoleh kesimpulan sebagai berikut:

1. Model LL3P memberikan hasil pemodelan yang lebih baik daripada model LN3P untuk pemodelan *unilevel*. Sedangkan untuk pemodelan *multilevel*, model LN3P memberikan hasil pemodelan yang lebih baik dari model LL3P.

2. Model GMLM\_Int LN3P (model *multilevel* berdistribusi LN3P dengan struktur hirarki pada parameter intersep) adalah model terbaik untuk pemodelan data pengeluaran perkapita rumah tangga dengan nilai DIC terkecil.

**2. Saran**

Dari plot antara data aktual dan hasil prediksi model ternyata masih didapatkan gap yang semakin besar untuk persentil menengah ke bawah dan persentil menengah ke atas. Selain itu, nilai Deviance dan DIC yang diperoleh dari pemodelan GMLM masih cukup tinggi. Oleh karena itu disarankan untuk penelitian selanjutnya agar menambah level model menjadi tiga level. Harapannya dengan menambah level model akan memperkecil nilai Deviance dan mengurangi besarnya gap antara data aktual dan prediksi model.

**DAFTAR PUSTAKA**

Anderson, T.W. dan Darling, D.A. 1954. A test Goodness of Fit. *Journal of American Statistical Association*, Volume 49, Issue 268, hal. 765-769.

Anderson, T.W. dan Darling, D.A. 1952. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on stochastic Process. *The Annals of Mathematical Statistics*, Vol. 23, No. 2, hal. 193-212.

Anderson, W., dan Wells, M.T. 2010. A Bayesian Hierarchical Regression Approach to Clustered and Longitudinal Data in Empirical Legal Studies. *Journal of Empirical Legal Studies*, Volume 7, Issue 4, hal. 634-663.

Boldstad, W.M. 2007. *Introduction to Bayesian Statistics. 2nd Edition*. Wiley, New Jersey.

Box, G.E.P. dan Tiao, G.C. 1973. *Bayesian Inference in Statistical Analysis*. John Willey And Sons, Inc : New York.

- BPS Kota Jambi. 2011. *Kota Jambi Dalam Angka 2011*. Badan Pusat Statistik Kota Jambi, Provinsi Jambi.
- Carlin, B.P. 1992. A Simple Monte Carlo Approach to Bayesian Graduation. *Transactions of the Society of Actuaries XLIV*, hal. 55–76.
- Carlin, B. P. dan Chib, S. 1995. Bayesian model choice via Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society, Ser. B*. Vol.57 No.3, hal.473–484.
- De Leeuw, J. dan Kreft, I. 2006. Random Coefficient Models for Multilevel Analysis. *Departement of Statistics Paper*. Departement of Statistics, UCLA, Los Angeles. [http://preprints.stat.ucla.edu/496/dLK\\_jes.pdf](http://preprints.stat.ucla.edu/496/dLK_jes.pdf) (19 Juli 2010).
- Dobson, A.J. 2002. *An Introduction to Generalized Linear Models, Second Edition*. Chapman & Hall, London
- Gelman, A., dan Hill, J. 2006. *Data Analysis Using Regression and Multilevel / Hierarchical Models*. Cambridge University Press.
- Goldstein, H. 1995. *Multilevel Statistical Models*. 2<sup>nd</sup> edition. Arnold, London. ISBN-10: 111995682X, hal. 382.
- Ha, I.D., dan Lee. Y. 2005. Multilevel Mixed Linear Models for Survival Data. *Lifetime Data Analysis*, 11, hal. 131-142.
- Hox, J.J. 1995. *Applied Multilevel Analysis*, 1<sup>st</sup> edition, TT-Publikaties, Amsterdam, hal. 119.
- Iriawan, N. 2012. *Pemodelan dan Analisis Data Driven*. ITS Press : Surabaya.
- Ismartini, P., Iriawan, N., Setiawan, dan Ulama, B.S.S. 2012. Toward a Hierarchical Bayesian Framework for Modelling the Effect of Regional Diversity on Household Expenditure. *Journal of Mathematics and Statistics*, Vol.8, No.2, hal. 283-291.
- Kaashoek, J.F. dan Van Dijk, H.K. 2002. Neural Network Pruning Applied to Real Exchange Rate Analysis. *Journal of Forecasting*, 21, pp. 559-577.
- McCullagh, P., dan Nelder, J.A. 1989. *Generalized Linear Models, Second Edition*. Chapman & Hall, London
- Raudenbush, S.W. dan A.S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2<sup>nd</sup> edition. Sage Publications, Thousand Oaks. ISBN-10: 076191904X, hal. 485.
- Scollnik, D. P. M. 2011. *An Introduction To Markov Chain Monte Carlo Methods And Their Actuarial Applications*. Handbook of Markov Chain Monte Carlo Chapter 1, hal.114-165 : Chapman & Hall/CRC Handbooks of Modern Statistical Method.
- Steenbergen, M. R., dan Jones, B. S. 2002. Modelling Multilevel Data Structure. *American Journal of Political Science*, Vol. 46, No. 1, hal. 218-237.
- Taylor, H.M., dan Carlin, S. 1998. *An Introduction to Stochastic Modelling* 3<sup>rd</sup> edition. Academic Press, San Diego.
- Zimmer, Z., Wen, M., dan Kaneda, T. 2010. A Multi-Level Analysis of Urban/Rural and Socioeconomic Differences in Functional Health Status Transition among Older Chinese. *Social Science & Medicine* 71, hal. 559-567.

# FAKTOR-FAKTOR YANG MEMENGARUHI *FOREIGN DIRECT INVESTMENT* (FDI) DI ENAM KORIDOR EKONOMI INDONESIA: *MARKET SEEKING* ATAU *RESOURCE SEEKING*?

Iriani Trisna Rahayu<sup>1</sup> dan Ernawati Pasaribu<sup>2</sup>

<sup>1</sup>Badan Pusat Statistik Propinsi Papua Barat, Jakarta

<sup>2</sup>Dosen Statistika, Sekolah Tinggi Ilmu Statistik, Jakarta

e-mail: <sup>2</sup>ernapasaribu@stis.ac.id

## Abstrak

Pembangunan ekonomi suatu negara sangat tergantung pada besarnya penanaman modal asing langsung atau *Foreign Direct Investment* (FDI), termasuk di enam koridor ekonomi Indonesia. Kesenjangan kondisi di koridor ekonomi diperkirakan akan membawa perbedaan yang mempengaruhi arus masuk FDI ke dalam koridor. Penelitian ini menggunakan regresi data panel untuk menganalisis faktor-faktor di balik arus masuk FDI di setiap koridor ekonomi dan untuk menentukan karakteristik FDI di setiap koridor ekonomi. Hasil penelitian menunjukkan bahwa proporsi belanja modal pemerintah, jumlah angkatan kerja berpendidikan tinggi, keterbukaan perdagangan, dan proporsi ekspor minyak dan mineral hanya mempengaruhi arus masuk FDI di beberapa koridor ekonomi. Lebih jauh lagi dapat diindikasikan bahwa sementara "*market seeking FDI*" terjadi di semua koridor ekonomi Indonesia, "*resource seeking FDI*" hanya ditemukan di koridor ekonomi Sulawesi, Maluku dan Papua.

**Kata kunci:** *FDI*, Koridor Ekonomi, data panel, *market seeking*, *resource seeking*

## Abstract

*The economic development of a country depends on the amount of foreign direct investment (FDI), including in the Indonesian six economic corridors. The huge gaps of conditions in economic corridors are expected to differences infactors affecting the FDI-inflow into the corridors. This study uses a panel data regression to analyze factors behind the FDI-inflow in each economic corridor and to determine the FDI characteristic in each economic corridor. It shows that the proportion of government capital expenditure, number of highly-educated labor force, trade openness, and the proportion of oil and mineral export affect the FDI-inflow only in some economic corridors. Furthermore, it indicates that, while market seeking FDI occurred in all Indonesian economic corridors, resource seeking FDI was only found in Sulawesi, Maluku and Papua economic corridors..*

**Keywords:** *Foreign Direct Investment (FDI), Indonesian economic corridors, panel data regression, market seeking, resource seeking*

## PENDAHULUAN

Pasca krisis tahun 1997, pemerintah mulai memperbaiki diri, terutama untuk membangkitkan perekonomian yang terpuruk. Hasil dari pembangunan yang terus digalakkan pemerintah, baik di pusat maupun di daerah tersebut, dapat terlihat dari perkembangan ekonomi Indonesia yang pada umumnya mengalami pertumbuhan yang positif.

Akan tetapi, mulai tahun 2007, tingkat pertumbuhan ekonomi Indonesia semakin melambat. Hal ini menunjukkan bahwa program pembangunan ekonomi yang selama ini dijalankan oleh pemerintah kurang memberikan hasil yang maksimal untuk percepatan pembangunan ekonomi Indonesia dalam jangka panjang. Oleh karena itu, pemerintah mencoba melakukan transformasi ekonomi, yaitu pengalihan perekonomian Indonesia yang awalnya didasari oleh konsumsi menjadi perekonomian yang didasari oleh investasi (Bappenas, 2011).

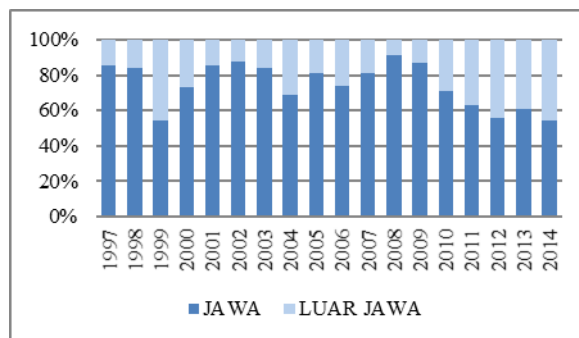
Di sisi lain, dalam upaya menjadikan Indonesia sebagai negara maju dan sejahtera tahun 2025, pemerintah menyusun rencana pembangunan yang difokuskan pada pengembangan sejumlah kegiatan ekonomi utama masing-masing wilayah Indonesia. Rencana tersebut tertuang dalam Masterplan Percepatan dan Perluasan Pembangunan Ekonomi Indonesia (MP3EI) melalui Peraturan Presiden Nomor 32 tahun 2011. MP3EI tersebut membagi wilayah Indonesia menjadi koridor-koridor ekonomi berdasarkan pada keunggulan dan potensi sektoral masing-masing wilayah. Sistem koridor ini mengacu pada kesuksesan sistem koridor ekonomi yang diterapkan negara-negara Asia lainnya. Indonesia menetapkan enam koridor ekonomi, yaitu Koridor Sumatera, Koridor Jawa, Koridor Kalimantan, Koridor Bali dan Nusa Tenggara, Koridor Sulawesi, serta Koridor Maluku dan Papua. Dengan adanya koridor ini, para investor akan lebih mudah dalam menentukan jenis investasi yang akan dilakukan, yaitu pada sektor ekonomi apa dan di mana.

Transformasi ekonomi yang digunakan sebagai alat untuk merealisasikan

MP3EI tersebut tentu membutuhkan sumber daya dan dana yang tidak sedikit. Sumber daya alam yang berlimpah serta jumlah penduduk yang sangat besar mungkin bukan merupakan masalah. Namun dana yang dapat disediakan oleh pemerintah sangatlah terbatas. Oleh sebab itu, dibutuhkan suntikan dana dari swasta, baik dari dalam negeri maupun dari luar negeri.

Salah satu sumber dana yang besar dan dapat segera digunakan adalah Investasi Asing Langsung atau *Foreign Direct Investment* (FDI). Investasi jenis ini lebih menguntungkan dibandingkan dengan investasi yang tidak langsung. Berbeda dengan investasi tak langsung, seperti investasi portofolio yang bisa tiba-tiba ditarik oleh investor, FDI lebih bersifat komitmen jangka panjang, sehingga dianggap lebih bernilai bagi suatu negara.

Namun, realisasi FDI di Indonesia ternyata belum merata di semua koridor ekonomi. Selaras dengan pembangunan yang berpusat di Jawa, FDI yang masuk ke Indonesia pun sebagian besar masih berada di koridor Jawa. Bahkan persentase (*share*) FDI di koridor Jawa dalam periode 1997-2014 selalu lebih dari 50% dari total FDI Indonesia.



Sumber: Badan Koordinasi Penanaman Modal (BKPM, diolah)

**Gambar 1. Persentase Realisasi FDI di Indonesia menurut Jawa dan Luar Jawa, 1997-2014**

Oleh karena itu, penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi FDI di masing-masing koridor ekonomi Indonesia. Penelitian ini juga bertujuan untuk menganalisis motivasi apakah yang melatarbelakangi investor asing untuk menanamkan modalnya di

Indonesia, apakah *market seeking* atau *resource seeking*.

## KAJIAN PUSTAKA

Menurut Todaro & Smith (2003), FDI merupakan investasi yang dilakukan oleh pihak swasta asing, yang mana dana-dana investasinya langsung digunakan untuk menjalankan kegiatan bisnis seperti mendatangkan mesin-mesin dan membeli bahan baku produksi. Teori elektikal Dunning (1993) menyebutkan bahwa keunggulan yang harus dimiliki oleh suatu negara untuk menarik investasi asing adalah antara lain biaya produksi rendah (termasuk transportasi), tersedianya pasar yang besar (*market size*), sumber daya yang melimpah (meliputi SDA dan tenaga kerja berupah rendah yang terampil), serta minimnya hambatan perdagangan dan tingkat keterbukaan ekonomi.

Dunning (1993) kemudian menyimpulkan bahwa beberapa hal yang melatarbelakangi investor asing untuk menanamkan modalnya di suatu negara dapat dikelompokkan ke dalam alasan *market seeking* dan atau *resource seeking*. *Market seeking* merupakan jenis investasi asing yang ditujukan untuk mencari pasar baru atau mempertahankan pasar lama. Sedangkan *resource seeking* adalah jenis investasi yang dilakukan untuk mendapatkan faktor-faktor produksi yang berbentuk sumber daya alam atau tenaga kerja yang lebih efisien di negara lain dibandingkan dengan di negara sendiri.

Penelitian terdahulu menunjukkan adanya beberapa faktor penting yang berpengaruh terhadap masuknya FDI ke suatu negara, yaitu *market size*, ketersediaan Sumber Daya Alam (SDA) dan Sumber Daya Manusia (SDM), infrastruktur, dan kebijakan perdagangan serta kondisi perekonomian negara tujuan (Jadhav, 2012; Hsiao & Hsiao, 2004; Asiedu, 2006, Rohmana, 2011; Sarwedi, 2002; Setiawan, 2002). Besarnya pasar atau *market size* akan memengaruhi keputusan investor untuk menanamkan modalnya ke negara tujuan, khususnya bagi investor yang ingin mengembangkan jangkauan pasarnya

(Asiedu, 2002). Dengan demikian, Indonesia yang merupakan salah satu negara dengan jumlah penduduk terbanyak di dunia merupakan sebuah pasar besar bagi investasi asing.

Meskipun PDB per kapita Indonesia tidak termasuk ke dalam golongan pendapatan tinggi, namun pola konsumsi masyarakat Indonesia masih cenderung konsumtif. Hal tersebut terbukti dari besarnya konsumsi masyarakat dilihat dari unsur PDB pengeluaran, dibandingkan dengan pengeluaran untuk investasi. Rendahnya investasi dan tingginya konsumsi inilah yang menjadi salah satu penyebab mengapa pembangunan ekonomi di Indonesia tidak bisa berlangsung pesat.

Tingginya pola konsumsi masyarakat Indonesia dan ditambah dengan upah yang masih rendah menjadi faktor lain yang dapat menarik FDI ke Indonesia. Perusahaan Multinasional bisa saja memberikan upah yang sedikit lebih tinggi dibandingkan dengan upah dari perusahaan dalam negeri, ataupun perusahaan pesaing, yang mana sebenarnya jika dibandingkan dengan upah di negara asalnya masih jauh lebih rendah. Tenaga kerjanya pun mempunyai *skill* yang cukup bagus jika dibandingkan dengan negara berkembang lainnya.

Selain akses terhadap SDM, akses terhadap Sumber Daya Alam (SDA) di Indonesia juga cukup mudah karena masih melimpahnya SDA di Indonesia. Bagi negara sedang berkembang, seperti Indonesia, ketersediaan SDA merupakan daya tarik bagi investor asing, terutama yang bergerak di industri primer (Jadhav, 2012).

Sebaliknya, langkanya SDA di negara-negara maju memberikan peluang bagi produk ekspor negara sedang berkembang yang bernilai ekonomis. (Asiedu & Lien, 2010). Investor asing akan lebih memilih untuk terjun langsung sebagai pengekspor di negara tujuan karena akan memberikan keuntungan yang lebih besar dibandingkan dengan mengimpor di negaranya sendiri.

Faktor lain yang menjadi penarik investor asing adalah kemudahan akses, melalui ketersediaan fasilitas umum dan infrastruktur. Kemudahan akses ini menjadi faktor penting bagi investor, khususnya di

Indonesia, terbukti dari realisasi FDI yang lebih dari 50% di koridor Jawa yang disebabkan karena selama ini pembangunan infrastruktur yang masih berpusat di koridor Jawa. Kemudahan akses serta tersedianya fasilitas dan infrastruktur akan memudahkan investor, baik dalam proses produksi maupun dalam pendistribusian produk.

Akan tetapi, masuknya investasi asing, khususnya FDI juga sangat tergantung pada kebijakan makroekonomi dan kebijakan perdagangan di negara tujuan. Suatu negara yang menganut sistem ekonomi tertutup tidak akan membuka peluang bagi investor asing untuk menanamkan modal di negara tersebut. Semakin terbuka sebuah negara dalam perdagangan internasional (ekspor dan impor), maka akan semakin banyak FDI yang masuk ke negara tersebut (Jadhav, 2012).

## DATA DAN METODE

### Data

Data yang digunakan adalah data tahunan 2006-2014 di 32 propinsi di Indonesia yang bersumber dari Badan Koordinator Penanaman Modal (BKPM), Badan Pusat Statistik (BPS), dan Direktorat Jenderal Perimbangan Keuangan Kementerian Keuangan (DJPK Kemenkeu).

Variabel dependen yang digunakan yaitu realisasi FDI (triliyun rupiah), sedangkan variabel independen adalah proporsi pengeluaran belanja modal terhadap total belanja pemerintah daerah (dalam persen), jumlah angkatan kerja berpendidikan SMA keatas (dalam logaritma natural), tingkat keterbukaan perdagangan (dalam persen), proporsi ekspor migas dan mineral terhadap total ekspor (dalam persen), dan PDRB riil tahun dasar 2000 (dalam triliun rupiah).

### Metode

Penelitian ini menggunakan metode regresi data panel untuk masing-masing koridor ekonomi Indonesia.

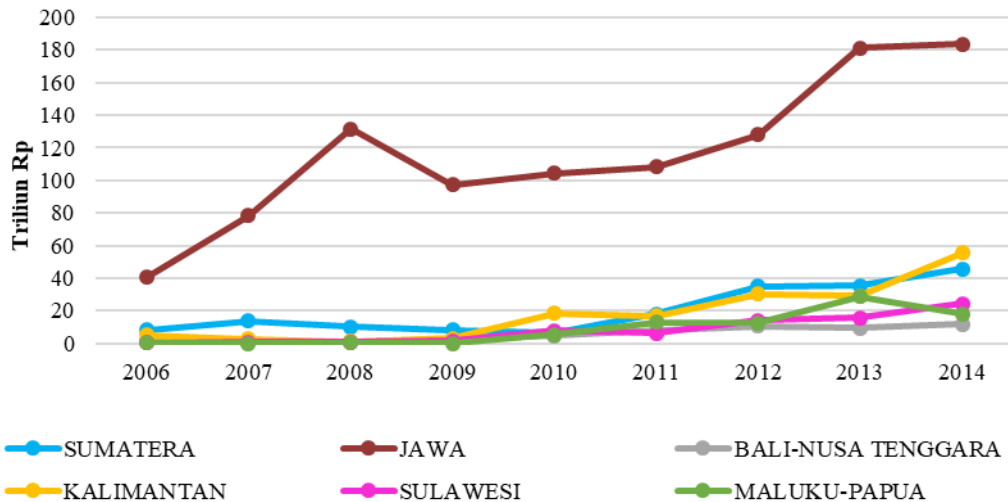
$$FDI_{ait} = \alpha_0 + \beta_1 BM_{ait} + \beta_2 \ln ANGKER_{ait} + \beta_3 OPEN_{ait} + \beta_4 XMIGAS_{ait} + \beta_5 PDRB_{ait} + \varepsilon_{ait}$$

di mana:

$\alpha_0$	: intersep
$\beta_1, \dots, \beta_5$	: koefisien regresi variabel independen
$FDI_{ait}$	: nilai realisasi FDI pada koridor a propinsi i tahun t (triliun rupiah)
$BM_{ait}$	: proporsi nilai pengeluaran belanja modal pemerintah daerah dibagi dengan nilai total pengeluaran belanja pemerintah daerah pada koridor a propinsi i tahun t (persen)
$ANGKER_{ait}$	: jumlah angkatan kerja berpendidikan tinggi pada koridor a propinsi i tahun t (orang)
$OPEN_{ait}$	: tingkat keterbukaan perdagangan pada koridor a propinsi i tahun t (persen)
$XMIGAS_{ait}$	: proporsi nilai ekspor minyak, gas, dan mineral terhadap nilai total ekspor pada koridor a propinsi i tahun t (persen rupiah)
$PDRB_{ait}$	: nilai Produk Domestik Regional Bruto riil pada koridor a propinsi i tahun t (triliun rupiah)
$\varepsilon_{ait}$	: <i>error term</i> untuk koridor a propinsi i tahun t
a	: koridor 1,2,...,6 (1=koridor Sumatera, 2=koridor Jawa, 3=koridor Bali-Nusa Tenggara, 4=koridor Kalimantan, 5=koridor Sulawesi, 6=koridor Maluku-Papua)
i	: Aceh,..., Papua
t	: 2006,..., 2014

## HASIL DAN PEMBAHASAN

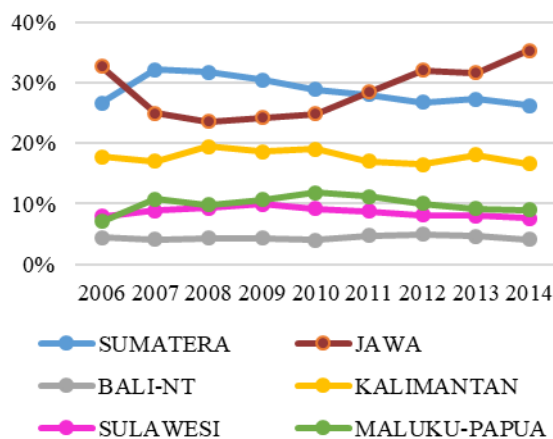
Total FDI yang direalisasikan di masing-masing koridor ekonomi dari tahun ke tahun terus mengalami peningkatan. Pada koridor Jawa, peningkatan tersebut diiringi dengan *share* FDI yang semakin menurun. Artinya, penanaman modal asing pada koridor di luar Jawa semakin berkembang dan merata.



Sumber: Badan Koordinasi Penanaman Modal (BKPM, diolah)

**Gambar 2. Total FDI di Masing-Masing Koridor Ekonomi Indonesia Tahun 2006-2014**

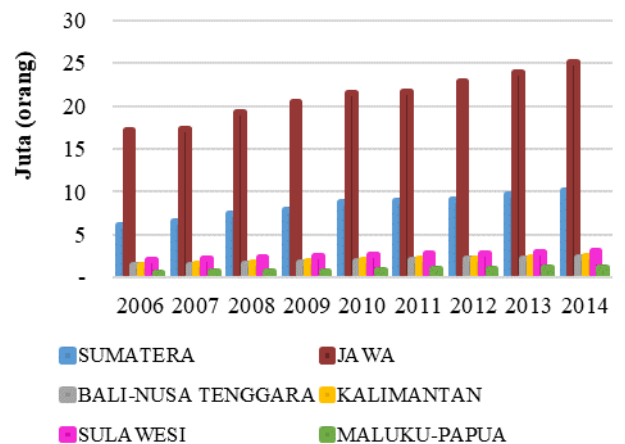
Proporsi belanja modal terbesar berada di koridor Sumatera dan koridor Jawa. Akan tetapi, pemerintah daerah cenderung mengalokasikan belanja modal dengan proporsi yang sama setiap tahun. Beberapa koridor malah memiliki proporsi belanja modal yang cenderung menurun. Hanya koridor Jawa yang selama lima tahun terakhir memiliki proporsi belanja modal yang cenderung semakin membesar.



Sumber: Direktorat Jendral Perimbangan Keuangan Kementerian Keuangan (DJPK Kemenkeu, diolah)

**Gambar 3. Proporsi Belanja Modal terhadap Total Belanja Pemerintah Daerah di Masing-Masing Koridor Ekonomi Indonesia Tahun 2006-2014 (%)**

Jumlah tenaga kerja yang memiliki pendidikan tinggi di Indonesia dari tahun 2006-2014 terus mengalami peningkatan. Dari tahun ke tahun koridor Jawa selalu memiliki jumlah terbesar dibandingkan dengan koridor-koridor lain. Koridor Sumatera juga mempunyai jumlah angkatan kerja berpendidikan tinggi yang cukup banyak. Sebaliknya, koridor Kalimantan, Koridor Bali-Nusa Tenggara, Koridor Sulawesi, dan Koridor Maluku-Papua memiliki jumlah angkatan kerja berpendidikan tinggi yang masih sedikit.

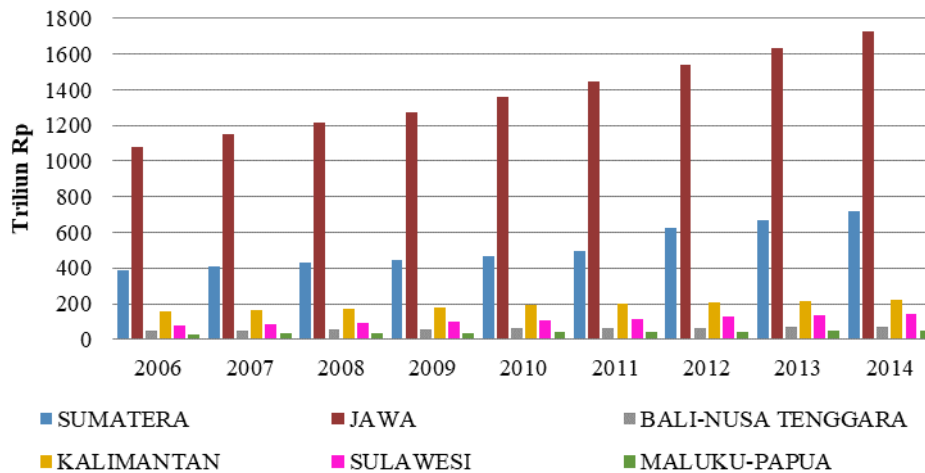


Sumber: BPS (diolah)

**Gambar 4. Jumlah Angkatan Kerja Berpendidikan Tinggi di Masing-Masing Koridor Ekonomi Indonesia Tahun 2006-2014**

Dari tahun 2006 hingga tahun 2014 jumlah nilai PDRB riil tertinggi adalah di koridor Jawa, yang pada tahun 2006 sebesar Rp1.081,4 triliun, meningkat hingga menjadi Rp1.729,4 triliun di tahun 2014. Sebaliknya, PDRB riil terkecil berada di koridor Maluku-Papua, yang mana pada

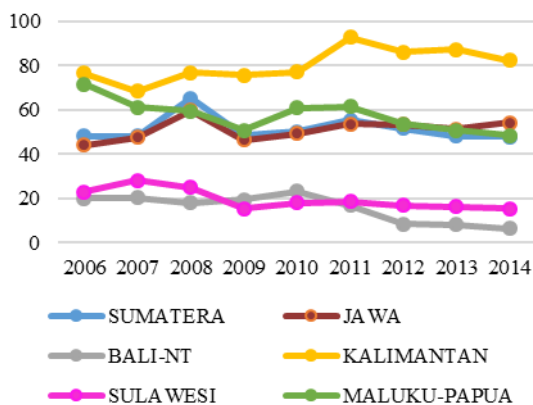
tahun 2006 sebesar Rp29,75 triliun menjadi Rp51,98 triliun di tahun 2014. Dalam kurun waktu 9 tahun, PDRB koridor Maluku-Papua hanya meningkatkan share PDRB-nya dari sebesar 1,66% menjadi 1,78% dari total PDB nasional.



Sumber: BPS (diolah)

**Gambar 5. Nilai PDRB Atas Dasar Harga Konstan Tahun 2000 di Masing-Masing Koridor Ekonomi Indonesia Tahun 2006-2014**

Tingkat keterbukaan perdagangan koridor Kalimantan merupakan koridor yang paling tinggi, yaitu mencapai 0,82 di tahun 2014, sedangkan koridor Bali-Nusa Tenggara dan koridor Sulawesi merupakan koridor yang cenderung paling tertutup. Koridor Sumatera, koridor Jawa, dan koridor Maluku-Papua memiliki tingkat keterbukaan yang hampir sama, yaitu masih berfluktuatif sekitar 0,5.



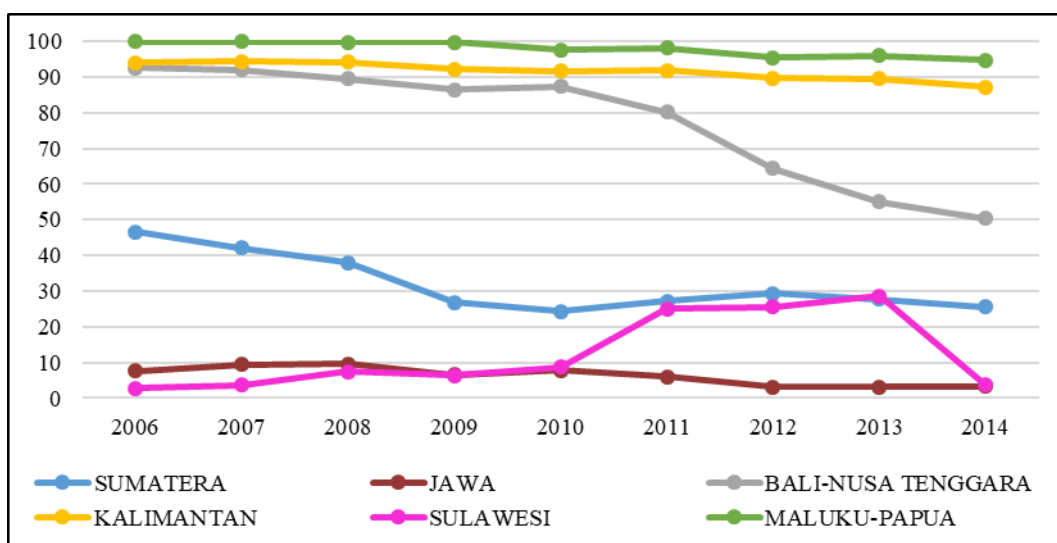
**Gambar 6. Tingkat Keterbukaan Perdagangan (Trade Openness) di Masing-Masing Koridor Ekonomi Indonesia Tahun 2006-2014**

Secara umum, koridor Maluku-Papua adalah koridor yang memiliki proporsi ekspor migas dan mineral tertinggi dibandingkan dengan koridor lain, sedangkan koridor Jawa adalah koridor dengan proporsi ekspor migas dan mineral terendah. Proporsi ekspor migas dan mineral koridor Maluku-Papua terhadap total ekspor koridor ini terus menurun, walaupun selalu lebih dari 99% hingga tahun 2009. Di tahun 2014, proporsi ekspor migas dan mineral koridor ini masih sangat besar, yaitu 0,94 (atau 94%).

Analisis inferensia dengan model regresi data panel menunjukkan bahwa model yang terpilih untuk koridor Sumatera yaitu *Fixed Effect Model* dengan *Seemingly Unrelated Regression* (FEM-SUR), begitu juga untuk model koridor Jawa dan Sulawesi. Sedangkan untuk koridor Bali-Nusa Tenggara hanya menggunakan *Fixed Effect Model*. Untuk koridor Kalimantan menggunakan *Fixed Effect Model* dengan *Weighted Least Square* (FEM-WLS). Sedangkan pada koridor Maluku-Papua terpilih model *Common*



*Effect Model dengan Weighted Least Square (CEM-WLS).*



**Gambar 7. Proporsi Ekspor Migas dan Mineral terhadap Total Ekspor di Masing-Masing Koridor Ekonomi Indonesia Tahun 2006-2014**

Analisis inferensia dengan model regresi data panel menunjukkan bahwa model yang terpilih untuk koridor Sumatera yaitu *Fixed Effect Model* dengan *Seemingly Unrelated Regression (FEM-SUR)*, begitu juga untuk model koridor Jawa dan Sulawesi. Sedangkan untuk koridor Bali-Nusa Tenggara hanya menggunakan

*Fixed Effect Model*. Untuk koridor Kalimantan menggunakan *Fixed Effect Model* dengan *Weighted Least Square (FEM-WLS)*. Sedangkan pada koridor Maluku-Papua terpilih model *Common Effect Model* dengan *Weighted Least Square (CEM-WLS)*.

**Tabel 1. Hasil estimasi untuk masing-masing koridor ekonomi**

Koridor \ Variabel	Sumatera	Jawa	Bali-Nusa Tenggara	Kalimantan	Sulawesi	Maluku-Papua
<b>R-square</b>	0.54128	0.83090	0.83964	0.74482	0.73076	0.32286
<b>F-statistik</b>	0.47534	0.83410	0.83964	0.79079	0.65064	0.32329
<b>Prob.</b>	0.00000	0.00000	0.00000	0.00000	0.00000	0.00431
<b>Koefisien:</b>						
• <b>Kontanta</b>	-19.78570	-214.68580	6.59222	-35.58441	-31.86938	22.41631
• <b>BM</b>	<b>-0.04195*</b> (0,0205)	<b>-0.53512*</b> (0,0143)	-0.03111 (0,5204)	0.01272 (0,8501)	<b>-0.03072*</b> (0,0756)	-0.07743 (0,2800)
• <b>LNANKER</b>	<b>1.87009*</b> (0,0039)	<b>14.62105*</b> (0,0587)	-0.66160 (0,7486)	0.41216 (0,9413)	<b>2.52456*</b> (0,0130)	-1.75210 (0,3818)
• <b>OPEN</b>	<b>-0.03082*</b> (0,0611)	0.03064 (0,8894)	<b>-0.16459*</b> (0,0000)	0.00598 (0,2373)	<b>-0.11929*</b> (0,0026)	<b>-0.16917*</b> (0,0015)
• <b>XMIGAS</b>	<b>-0.02881*</b> (0,0463)	0.10944 (0,3430)	-0.00702 (0,5924)	<b>-0.06619*</b> (0,0123)	<b>0.02729*</b> (0,0105)	<b>0.06167*</b> (0,0490)
• <b>PDRB</b>	<b>0.00892*</b> (0,0032)	<b>0.10077*</b> (0,0000)	<b>0.42783*</b> (0,0015)	<b>0.78313*</b> (0,0008)	<b>0.12673*</b> (0,0001)	<b>0.46480*</b> (0,0010)

\*signifikan pada level 5%

Berdasarkan hasil analisis inferensia, proporsi belanja modal, jumlah angkatan kerja berpendidikan tinggi, dan tingkat keterbukaan perdagangan mempunyai pengaruh yang berbeda-beda di masing-masing koridor ekonomi Indonesia.

Proporsi belanja modal pemerintah daerah hanya berpengaruh signifikan terhadap FDI di koridor Sumatera, koridor Jawa, dan koridor Sulawesi. Sedangkan di koridor Bali dan Nusa Tenggara, koridor Kalimantan, dan koridor Maluku dan Papua, proporsi belanja modal pemerintah daerah tidak berpengaruh signifikan terhadap FDI.

Sebaliknya, jumlah angkatan kerja berpendidikan tinggi berpengaruh signifikan dan positif terhadap FDI di koridor Sumatera, koridor Jawa, dan koridor Sulawesi. Hal ini kemungkinan terjadi karena koridor Jawa dan koridor Sumatera memiliki jumlah angkatan kerja berpendidikan tinggi terbanyak. Sedangkan koridor Sulawesi memiliki proporsi angkatan kerja berpendidikan tinggi terbesar dibandingkan dengan koridor-koridor lain.

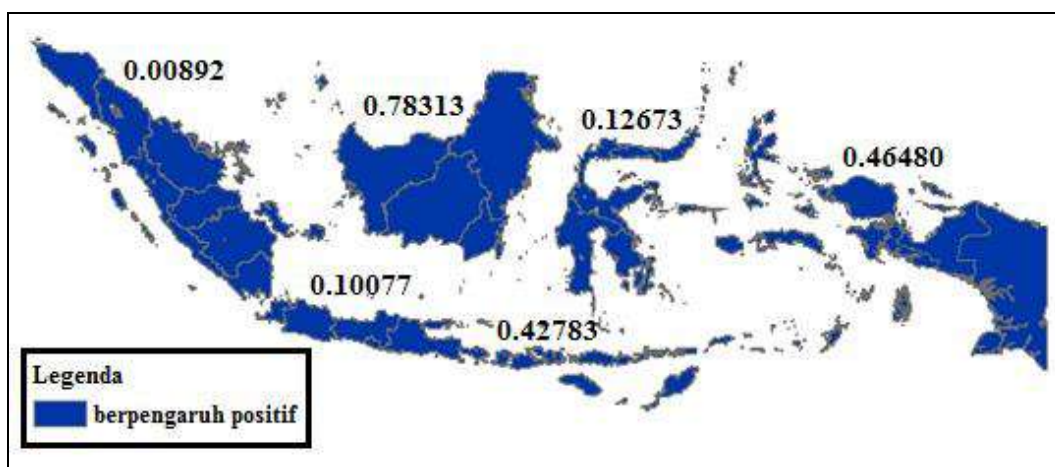
Di samping itu, seperti halnya variabel proporsi belanja modal, jumlah angkatan kerja berpendidikan tinggi juga tidak berpengaruh signifikan terhadap FDI di koridor Bali dan Nusa Tenggara, koridor Kalimantan, dan koridor Maluku dan Papua.

Tingkat keterbukaan perdagangan berpengaruh negatif dan signifikan terhadap FDI di koridor Sumatera, koridor Bali dan Nusa Tenggara, koridor Sulawesi, dan

koridor Maluku dan Papua. Sedangkan di koridor Jawa dan koridor Kalimantan, tingkat keterbukaan perdagangan tidak berpengaruh signifikan terhadap FDI.

Pengaruh tingkat keterbukaan perdagangan terhadap FDI tergantung pada jenis FDI yang masuk ke wilayah tersebut. Pada market seeking FDI, tingkat keterbukaan perdagangan akan cenderung untuk berpengaruh negatif terhadap FDI. Hal tersebut biasa terjadi pada negara-negara maju, yang mana cenderung memiliki tingkat keterbukaan yang lebih kecil dibandingkan dengan negara-negara berkembang (Briguglio, 2016). Sebaliknya, pada resource seeking FDI, tingkat keterbukaan perdagangan cenderung berpengaruh positif terhadap FDI. Akan tetapi, tingkat keterbukaan bisa berpengaruh terhadap FDI bertipe market seeking apabila pasar yang dicari oleh investor bukan hanya berada di dalam wilayah tersebut, melainkan juga wilayah disekitarnya.

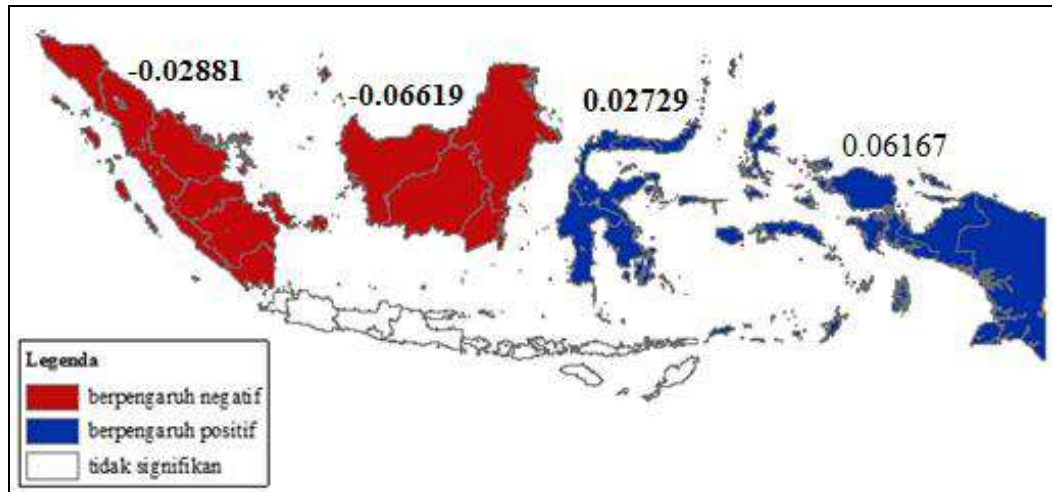
Motivasi market seeking dan resource seeking di setiap koridor ekonomi Indonesia disimpulkan berdasarkan signifikansi dan pengaruh variabel-variabel proksi terhadap FDI. Adanya motivasi market seeking dan resource seeking adalah apabila variabel-variabel proksi untuk market size dan natural resource availability, yaitu PDRB dan proporsi ekspor migas dan mineral, berpengaruh signifikan dan positif terhadap FDI.



**Gambar 9.** Nilai *Slope* dan Pengaruh PDRB Terhadap FDI di Seluruh Koridoe Ekonomi Indonesia

Dari hasil pengujian signifikansi variabel untuk masing-masing koridor, variabel proksi untuk market size, yaitu PDRB, signifikan dan positif di semua koridor. Hasil tersebut mengindikasikan bahwa tipe FDI yang masuk ke masing-

masing koridor di Indonesia bersifat market seeking. Hal ini didukung dengan tidak signifikannya variabel proksi untuk natural resource availability, yaitu proporsi ekspor migas dan mineral di koridor Jawa dan koridor Bali dan Nusa Tenggara.



**Gambar 10. Nilai Slope dan Pengaruh Proporsi Ekspor Migas dan Mineral Terhadap FDI di Seluruh Koridor Ekonomi Indonesia**

Variabel proporsi migas dan mineral berpengaruh negatif terhadap realisasi FDI di koridor Sumatera. Hal tersebut berarti bahwa FDI yang masuk ke koridor Sumatera bukanlah tipe resource seeking. Hasil serupa juga didapatkan di koridor Kalimantan, yang mana natural resource availability juga menunjukkan hasil signifikan namun negatif. Hasil ini berbeda dengan fungsi koridor Kalimantan yang merupakan sentra produksi dan pengolahan hasil tambang yang seharusnya merupakan faktor penarik utama bagi investor untuk menanamkan modalnya disana.

Di sisi lain, natural resource availability menunjukkan hasil yang signifikan dan positif di koridor Sulawesi dan koridor Maluku dan Papua. Hal ini menunjukkan bahwa, selain FDI yang bertipe market seeking, ada juga FDI yang masuk ke koridor Sulawesi dan koridor Maluku-Papua dengan tipe resource seeking. Hal ini sesuai dengan hipotesis bahwa FDI yang masuk ke koridor Maluku-Papua akan bertipe resource seeking karena sumber daya alamnya yang masih melimpah.

## KESIMPULAN DAN SARAN

Perkembangan realisasi FDI masih didominasi oleh koridor Jawa, namun di koridor lain sudah mulai tumbuh. Pengeluaran belanja modal, jumlah angkatan kerja berpendidikan tinggi, dan PDRB juga masih didominasi oleh koridor Jawa. Tingkat keterbukaan perdagangan yang paling tertinggi berada di koridor Kalimantan. Sedangkan proporsi ekspor migas dan mineral terbesar ada di koridor Maluku dan Papua.

Faktor-faktor yang memengaruhi FDI berbeda-beda untuk masing-masing koridor. Di koridor Sumatera, koridor Jawa, dan koridor Sulawesi, belanja modal berpengaruh negatif terhadap realisasi FDI, sedangkan jumlah angkatan kerja memiliki pengaruh yang positif. Faktor tingkat keterbukaan perdagangan berpengaruh negatif di koridor Sumatera, koridor Bali dan Nusa Tenggara, koridor Sulawesi, dan koridor Maluku dan Papua. Di koridor Sumatera dan koridor Kalimantan, ekspor migas dan mineral berpengaruh negatif terhadap realisasi FDI, sebaliknya ekspor migas berpengaruh positif terhadap realisasi FDI di koridor Sulawesi dan koridor Maluku

dan Papua. PDRB berpengaruh positif terhadap realisasi FDI di semua koridor ekonomi.

FDI yang bermotivasi market seeking ditemukan di semua koridor ekonomi Indonesia. Sementara itu, FDI dengan motivasi resource seeking hanya ditemukan di koridor Sulawesi dan koridor Maluku dan Papua.

Pemerintah perlu mengevaluasi penggunaan pengeluaran belanja modal daerah agar tepat sasaran. Pemerintah juga perlu meningkatkan kebijakan tentang pendidikan, terutama untuk pendidikan tinggi, dan kebijakan yang dapat meningkatkan PDRB. Selain itu, pemerintah perlu mencari alternatif pengganti sumber daya alam di Maluku-Papua yang masih menjadi incaran para investor asing agar bisa dimanfaatkan dengan baik dan menghasilkan nilai tambah yang akan lebih menguntungkan wilayah setempat. Selain itu, motivasi *resource seeking* yang masih ditemukan di koridor Maluku-Papua juga belum sesuai dengan tujuan pemerintah yang ingin mengembangkan kegiatan ekonomi di masing-masing wilayah Indonesia.

## DAFTAR PUSTAKA

- Asiedu, E. 2002. On the Determinants of Foreign Direct Investment to Developing Countries: Is Africa Different?. *World Development*. 30(1), 107-119. (diakses 2 Agustus 2016). [people.ku.edu/~asiedu/FDI-in-Africa-WD.pdf](http://people.ku.edu/~asiedu/FDI-in-Africa-WD.pdf)
- Asiedu, E. 2006. Foreign Direct Investment in Africa: The Role of Natural Resources, Market Size, Government Policy, Institutions and Political Instability. *working paper*. United Nation University. (diakses 25 Juni 2016). [people.ku.edu/~asiedu/world-economy.pdf](http://people.ku.edu/~asiedu/world-economy.pdf)
- Asiedu, E. dan Lien, D.D. 2010. Democracy, Foreign Direct Investment and Natural Resources. *Working paper*. (diakses 1 Agustus 2016). [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1726587](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1726587)
- Badan Pusat Statistik. (2006-2014). *Keadaan Angkatan Kerja di Indonesia*. Jakarta: BPS
- Badan Pusat Statistik. (2007-2015). *Statistik Indonesia*. Jakarta: BPS
- Briguglio, L. 2016. *Small States And The European Union: Economic Perspectives* New York: Routledge
- Dunning, J. H. 1993. *The Globalization of Business*. (diakses 25 Juni 2016). [http://unctad.org/en/PublicationChapters/iteiitv3n1a3\\_en.pdf](http://unctad.org/en/PublicationChapters/iteiitv3n1a3_en.pdf)
- Jadhav, P. 2012. Determinants of Foreign Direct Investment in BRICS economies: Analysis of Economics, Institutional, and Political Factor. *Procedia – Social and Behavioral Science*. 37, 5-14. (diakses 26 Januari 2016). <http://www.sciencedirect.com/science/article/pii/S1877042812007495>
- Kemenkeu. 2011. *Klasifikasi Jenis Belanja*. (diakses 29 Juni 2016). <http://www.jdih.kemenkeu.go.id/fullText/2011/101~PMK.02~2011PerLamp%20III>
- Rohmana, Y. 2011. Analisis Faktor-Faktor yang Mempengaruhi Investasi Asing Langsung di Indonesia Periode 1980-2008. *Jurnal Sains dan Terapan*. 6(2). Universitas Pendidikan Indonesia (UPI) Bandung. (diakses 8 Februari 2015). <http://jurnal.upi.edu/2022/view/1119/analisis-faktor-faktor-yang-mempengaruhi-investasi-asing-langsung-di-indonesia-periode-1980-2008.html>
- Sarwedi. 2002. Investasi Asing Langsung di Indonesia dan Faktor yang Mempengaruhinya. *Jurnal Akuntansi & Keuangan*. 4(1), 17–35 Jurusan Ekonomi Akuntansi, Fakultas Ekonomi - Universitas Kristen Petra.
- Setiawan, G. 2002. The Impact of Foreign Direct Investment on Indonesian Economic Growth. *Tesis*. Seoul: KDI (Korea Development Institute) School of Public Policy and Management.
- Todoaro, M.P. dan Smith, S.C. 2003. *Pembangunan Ekonomi di Dunia Ketiga: Jilid 1*. Jakarta: Erlangga.

# DETERMINAN PERILAKU MEROKOK PADA REMAJA SEKOLAH DI INDONESIA

Titik Harsanti<sup>1</sup> dan Febri Wicaksono<sup>2</sup>

Sekolah Tinggi Ilmu Statistik  
e-mail :<sup>1</sup>titik@stis.ac.id, <sup>2</sup>febri@stis.ac.id

## Abstrak

Saat ini merokok telah menjadi masalah kesehatan secara global dan menjadi beban ekonomi yang berat. Di Indonesia, tren merokok cenderung semakin meningkat dan kecenderungan ini tidak hanya terjadi pada orang dewasa tetapi juga pada remaja. Penelitian ini bertujuan mengidentifikasi faktor-faktor yang mempengaruhi perilaku merokok bagi kalangan remaja yang bersekolah di Indonesia dengan menggunakan model logistik multivariat biner. Analisis dilakukan dengan menggunakan 5.986 sampel siswa dari *Global Youth Tobacco Survey 2014 (GYTS) 2014*. Hasilnya menunjukkan bahwa 25% siswa pernah merokok dan 15% siswa saat ini merokok. Peluang siswa untuk merokok lebih tinggi di kalangan anak laki-laki dibandingkan dengan anak perempuan. Risiko merokok yang lebih tinggi teramati di antara siswa yang memiliki teman dekat yang merokok dibandingkan dengan siswa yang tidak memiliki teman dekat yang merokok. Siswa yang salah satu atau kedua orang tuanya merokok lebih cenderung merokok dibandingkan dengan siswa yang orang tuanya tidak merokok. Siswa yang pernah melihat gurunya merokok atau pernah melihat orang-orang merokok di rumah mereka dan tempat-tempat umum lebih cenderung merokok dibandingkan dengan mereka yang tidak pernah melihat gurunya merokok atau tidak pernah melihat orang merokok di rumah mereka dan tempat umum. Temuan ini menunjukkan bahwa penegakan peraturan untuk mengurangi aksesibilitas rokok diperlukan untuk mengekang penggunaan rokok di kalangan siswa. Selain itu, intervensi dan kampanye pendidikan yang menargetkan siswa sekolah menengah juga diperlukan.

**Kata kunci:** Tembakau, Merokok, Sekolah, Remaja, Indonesia

## Abstract

*Smoking is a global public health concern and it imposes a heavy economic burden. However, the trend of smoking in Indonesia seems to be increasing and the magnitude of the problem affects not only adults but also adolescents. This paper identifies cigarette smoking determinants among school adolescents in Indonesia, using a multivariate binary logistic model. The analysis uses 5,986 samples of students from the 2014 Indonesia Global Youth Tobacco Survey (GYTS). The results show that 25% of the students have ever smoked and 15% of students are currently smoking. The students' odds of smoking are higher for boys compared to girls. Higher risk of smoking is observed among the students who have closed-peer smoking compared to students who don't have closed-peer smoking. Students whose one or both parents are smoking are more likely to smoke compared to whose parents are not smoking. Students who have seen their teacher smoking or have seen people smoking in their house and public places are more likely to smoke compared to who haven't ever seen their teacher smoking or haven't ever seen people smoking in their house and public places. These findings suggest that enforcement of legislations to decrease accessibility of cigarettes are necessary to curb the cigarette use among students. Beside that the interventions and education campaigns that target secondary school students are also needed.*

**Keywords:** Tobacco, Smoking, School, Adolescent, Indonesia

## PENDAHULUAN

Rokok merupakan salah satu kekhawatiran terbesar dunia kesehatan karena menyebabkan hampir 6 juta orang meninggal dalam setahun. Lebih dari 5 juta orang yang meninggal tersebut adalah akibat menghisap rokok secara langsung dan 600 ribu orang meninggal karena terpapar asap rokok (WHO, 2014).

Selain itu, rokok juga menyebabkan beban yang sangat berat terhadap perekonomian di seluruh dunia. Diperkirakan bahwa pengeluaran kesehatan yang disebabkan oleh rokok mencapai 5,7% dari total pengeluaran kesehatan dunia di tahun 2012. Kemudian total biaya ekonomi yang disebabkan oleh rokok diperkirakan setara dengan 1,8% dari total Produk Domestik Bruto (PDB) dunia di tahun 2012 (dihitung dari pengeluaran kesehatan dan berkurangnya produktivitas pekerja) (Goodchild, Nargis, & Tursan d'Espaignet, 2017).

WHO (2011) menyebutkan bahwa dampak negatif merokok pada usia remaja jauh lebih besar dibandingkan ketika merokok pada usia dewasa. Permulaan merokok biasanya terjadi pada saat remaja, namun efek merokok seperti kematian dan kecacatan yang terkait dengan merokok baru bisa terlihat secara nyata di masa tua (Elders, Perry, Eriksen, & Giovino, 1994). Lebih lanjut, merokok juga telah digambarkan sebagai “pintu gerbang” terhadap penggunaan narkoba dan obat-obat terlarang lainnya oleh remaja (Gilliland et al., 2006). Oleh karena itu, selain dapat menyebabkan masalah kesehatan jangka pendek seperti masalah pernafasan, kecanduan terhadap nikotin, dan resiko penggunaan obat-obatan terlarang, merokok pada remaja juga dapat menimbulkan masalah kesehatan jangka panjang karena faktanya adalah kebanyakan orang yang merokok sejak remaja cenderung untuk merokok sampai dewasa.

Tren kenaikan penggunaan rokok di Indonesia bukan hanya didominasi oleh kelompok dewasa, namun juga sudah merambah ke kelompok remaja. Kementerian Kesehatan Republik Indonesia

menyebutkan bahwa prevalensi merokok penduduk berumur 15-19 tahun telah meningkat dari 7,1% di tahun 1995 menjadi 20,5% di tahun 2013 (TCSC – IAKMI, 2014). Oleh karena itu penanganan masalah rokok perlu menjadi prioritas. Sehingga, penelitian ini bertujuan untuk menginvestigasi faktor yang mempengaruhi penggunaan rokok pada remaja di Indonesia.

## DATA DAN METODOLOGI

Penelitian ini menggunakan data *Global Youth Tobacco Survey (GYTS)* Indonesia tahun 2014. GYTS Indonesia tahun 2014 merupakan survei representatif nasional berbasis sekolah yang mendata siswa kelas 7, 8, dan 9 (WHO, 2015).

GYTS Indonesia tahun 2014 menggunakan desain sampel dua tahap (*two-stage sample design*), dengan memilih sekolah proporsional terhadap jumlah siswanya. Pada tahap pertama, dipilih 72 sekolah. Kelas-kelas yang berada pada sekolah tersebut selanjutnya dipilih secara acak dan semua siswa yang berada di dalamnya memenuhi syarat (*eligible*) untuk berpartisipasi di dalam survei. 72 sekolah, 208 kelas, dan 5.986 siswa berpartisipasi di dalam survei ini (WHO, 2015).

Unit analisis yang digunakan dalam penelitian ini adalah siswa kelas 7, 8, dan 9. Variabel terikat yang digunakan dalam penelitian ini adalah status merokok selama kurun waktu 30 hari sebelum pendataan sampai dengan pendataan berlangsung. Status merokok terdiri dari 2 kemungkinan, yaitu merokok (kode: 1) dan tidak merokok (kode: 0).

Variabel bebas yang digunakan dalam penelitian ini adalah jenis kelamin, tingkat pendidikan (kelas), keterpaparan akan rokok di rumah, keterpaparan akan rokok di dalam ruang publik tertutup, keterpaparan akan rokok di ruang publik terbuka, keterpaparan akan iklan rokok, keterpaparan akan kampanye anti rokok, guru merokok, orang tua merokok, teman dekat merokok, diskusi mengenai bahaya rokok di dalam keluarga, serta pengetahuan dan sikap terhadap bahaya merokok. Dari beberapa penelitian terkait, variabel-variabel tersebut diduga

berpengaruh secara signifikan terhadap perilaku merokok pada remaja (Hou, Xu, & Anderson, 2015; Rachmat, Thaha, & Syafar, 2013; Reda, Moges, Yazew, & Biadgilign, 2012; Ribeiro Sarmiento & Yehadji, 2015).

Dalam penelitian ini, status merokok sebagai variabel terikat merupakan variabel dikotomi, sehingga analisis yang digunakan untuk melihat determinan dari perilaku merokok menggunakan model regresi logistik biner (Kleinbaum & Klein, 2010). Model regresi logistik biner yang digunakan dalam penelitian ini dapat diformulasikan sebagai berikut

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_i x_i$$

dimana  $\pi(x)$  merupakan peluang untuk merokok,  $\beta_0$  merupakan konstanta,  $\beta_i$  merupakan parameter-parameter yang akan diestimasi,  $x_i$  merupakan variabel-variabel bebas, dan  $i$  merupakan banyaknya variabel bebas.

## HASIL DAN PEMBAHASAN

Secara umum, karakteristik sampel dalam penelitian ini tersaji dalam Tabel 1. Rata-rata umur responden adalah 13,26 tahun dengan standar deviasi sebesar 0,014. Dengan tidak memperhatikan periode waktu merokoknya, 1.500 remaja atau 25% dari total responden pernah merokok, dimana 570 remaja diantaranya (38%) mulai merokok pada usia 12 atau 13 tahun.

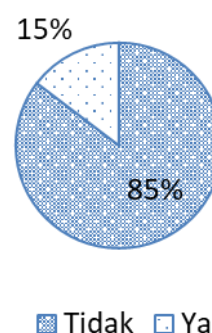
Dari keseluruhan responden, terdapat 876 (15%) responden yang merokok pada saat pendataan (kurun waktu 30 hari sebelum pendataan sampai dengan pendataan berlangsung) (Grafik 1). Sebagian besar perokok (74%) merokok tidak lebih dari satu batang rokok setiap hari, namun juga terdapat 6% perokok yang merokok lebih dari 5 batang rokok setiap harinya.

Tingginya prevalensi merokok pada siswa remaja ini salah satunya mungkin dikarenakan mudahnya akses dalam mendapatkan rokok. Studi ini menemukan bahwa 49% perokok mendapatkan rokok dengan membeli di toko, dimana 65% dari seluruh perokok membeli rokok secara eceran per batang. Meskipun di dalam

Peraturan Pemerintah Nomor 109 Tahun 2012 telah disebutkan adanya larangan penjualan rokok terhadap anak di bawah usia 18 tahun ke bawah, namun di dalam data GYTS Indonesia tahun 2014 masih dijumpai adanya siswa sekolah yang berusia kurang dari 18 tahun yang dapat membeli rokok di toko, bahkan secara eceran per batang. Hal ini menunjukkan masih lemahnya pengawasan pemerintah terhadap penjualan rokok di masyarakat.

Kemudian studi ini juga menemukan bahwa perokok remaja sekolah didominasi oleh laki-laki. Dua puluh sembilan persen diantara responden laki-laki merokok pada saat pendataan, sedangkan diantara responden perempuan, hanya dijumpai 2% diantaranya yang merokok pada saat pendataan.

Selain itu, ditemui juga bahwa 93% perokok mempunyai teman dekat yang merokok. Hal ini mengindikasikan adanya pengaruh teman sebaya yang dapat mempengaruhi seorang remaja untuk merokok.



**Grafik 1. Persentase Siswa Berdasarkan Status Merokok**

Hasil analisis regresi logistik biner menunjukkan bahwa jenis kelamin, tingkat kelas, jumlah hari responden melihat orang merokok di dalam rumah (7 hari terakhir), jumlah hari responden melihat orang merokok di dalam ruang publik tertutup (7 hari terakhir), jumlah hari responden melihat orang merokok di ruang publik terbuka (7 hari terakhir), melihat guru merokok di area sekolah, teman dekat merokok, dan orang tua merokok, serta pengetahuan dan sikap terhadap bahaya rokok berpengaruh secara signifikan terhadap perilaku merokok pada remaja yang bersekolah (lihat Tabel 2).

**Tabel 1. Karakteristik Sampel**

	Status Merokok				Total	
	Tidak		Ya		n	%
	n	%	n	%		
<b>Jenis Kelamin</b>						
Laki-laki	1,995	71	813	29	2,808	47
Perempuan	3,115	98	63	2	3,178	53
<b>Kelas</b>						
7	2,149	87	315	13	2,464	41
8	1,834	86	290	14	2,124	35
9	1,127	81	271	19	1,398	23
<b>Jumlah hari responden melihat orang merokok di dalam rumah (7 hari terakhir)</b>						
0	2,528	94	156	6	2,684	45
> 0	2,582	78	720	22	3,302	55
<b>Jumlah hari responden melihat orang merokok di dalam ruang publik tertutup (7 hari terakhir)</b>						
0	2,400	93	173	7	2,573	43
> 0	2,710	79	703	21	3,413	57
<b>Jumlah hari responden melihat orang merokok di ruang publik terbuka (7 hari terakhir)</b>						
0	2,249	93	157	7	2,406	40
> 0	2,861	80	719	20	3,580	60
<b>Pernah mendengar kampanye anti tembakau (30 hari terakhir)</b>						
Tidak	1,520	87	218	13	1,738	29
Ya	3,590	85	658	15	4,248	71
<b>Pernah melihat iklan rokok atau orang merokok di media</b>						
Tidak	1,734	90	191	10	1,925	32
Ya	3,376	83	685	17	4,061	68
<b>Salah satu atau kedua orang tua merokok</b>						
Tidak atau tidak tahu	1,741	90	189	10	1,930	32
Ya	3,369	83	687	17	4,056	68
<b>Teman dekat merokok</b>						
Tidak	2,634	98	65	2	2,699	45
Ya	2,476	75	811	25	3,287	55
<b>Diskusi keluarga mengenai bahaya merokok</b>						
Tidak	2,290	87	335	13	2,625	44
Ya	2,820	84	541	16	3,361	56
<b>Melihat guru merokok di area sekolah</b>						
Tidak atau tidak tahu	2,111	91	202	9	2,313	39
Ya	2,999	82	674	18	3,673	61



**Tabel 2. Estimasi Regresi Logistik Biner: Determinan Merokok**

	Koefisien	Stand ard Error	Rasio Kecenderu ngan (RK)	95% C.I. untuk RK	
				Lower	Upper
Jenis kelamin					
Laki-laki	2.66	0.14*	14.26	10.86	18.73
Perempuan <sup>r)</sup>					
Kelas					
9	0.22	0.11**	1.25	1.00	1.55
8	-0.04	0.10	0.96	0.79	1.18
7 <sup>r)</sup>					
Jumlah hari responden melihat orang merokok di dalam rumah (7 hari terakhir)					
> 0	0.77	0.12*	2.17	1.73	2.72
0 <sup>r)</sup>					
Jumlah hari responden melihat orang merokok di dalam ruang publik tertutup (7 hari terakhir)					
> 0	0.55	0.12*	1.73	1.37	2.19
0 <sup>r)</sup>					
Jumlah hari responden melihat orang merokok di ruang publik terbuka (7 hari terakhir)					
> 0	0.42	0.12*	1.52	1.19	1.93
0 <sup>r)</sup>					
Melihat guru merokok di area sekolah					
Ya	0.32	0.10*	1.37	1.13	1.67
Tidak atau tidak tahu <sup>r)</sup>					
Pernah melihat iklan rokok atau orang merokok di media					
Ya	0.04	0.11	1.04	0.84	1.28
Tidak <sup>r)</sup>					
Pernah mendengar kampanye anti tembakau (30 hari terakhir)					
Ya	0.02	0.10	1.02	0.84	1.23
Tidak <sup>r)</sup>					
Teman dekat merokok					
Ya	1.66	0.14*	5.28	4.00	6.97
Tidak <sup>r)</sup>					
Salah satu atau kedua orang tua merokok					
Ya	0.21	0.11**	1.24	1.00	1.52
Tidak <sup>r)</sup>					
Diskusi keluarga mengenai bahaya merokok					
Ya	0.09	0.09	1.09	0.92	1.31
Tidak <sup>r)</sup>					
Pengetahuan dan sikap terhadap bahaya rokok					
	-0.26	0.04*	0.77	0.72	0.83

Catatan: r) kategori referensi  
 \* signifikan pada  $\alpha=1\%$   
 \*\* signifikan pada  $\alpha=5\%$

Resiko remaja sekolah laki-laki untuk merokok 14,26 kali lebih besar dibandingkan dengan remaja sekolah perempuan (95% CI: 10,86 – 18,73). Studi di Jakarta, Guangdong (Cina), Nepal, Ethiopia Timur, dan Timor Leste (Global Youth

Tobacco Survey Collaborating Group, 2003; Reda et al., 2012; Ribeiro Sarmiento & Yehadji, 2015) juga menemukan adanya resiko yang lebih besar pada remaja laki-laki untuk merokok. Dalam konteks Indonesia, rendahnya resiko perempuan untuk merokok

mungkin disebabkan karena adanya nilai budaya yang mengstigmakan wanita yang merokok sebagai perilaku moral yang buruk (Barraclough, 1999).

Remaja sekolah yang memiliki teman dekat yang merokok mempunyai resiko 5,28 kali lebih besar untuk merokok dibandingkan dengan yang tidak memiliki teman dekat yang merokok (95%CI: 4,00 – 6,97). Beberapa penelitian diberbagai daerah juga mendapati adanya hubungan positif antara memiliki teman yang merokok dengan perilaku merokok remaja (Reda et al., 2012; Ribeiro Sarmiento & Yehadji, 2015; Rudatsikira, Dondog, Siziya, & Muula, 2008). Hasil ini mengindikasikan adanya tekanan sosial dari teman sebaya yang dapat mempengaruhi resiko remaja untuk merokok. Bricker dan kawan-kawan (2006) menyebutkan bahwa merokok diantara teman-teman dapat mempengaruhi inisiasi dan keberlanjutan penggunaan rokok di kalangan remaja.

Penelitian ini juga menemukan bahwa remaja sekolah yang salah satu atau kedua orang tuanya merokok mempunyai resiko 1,24 kali lebih besar untuk merokok dibandingkan dengan yang kedua orang tuanya tidak merokok (95%CI: 1,00 – 1,52). Kemudian remaja sekolah yang pernah melihat guru merokok di area sekolah mempunyai resiko 1,37 kali lebih besar untuk merokok dibandingkan dengan yang tidak pernah melihat guru merokok di area sekolah (95%CI: 1,13 – 1,67). Selain itu, remaja sekolah yang pernah melihat orang merokok di dalam ruang publik tertutup mempunyai resiko 1,73 kali lebih besar untuk merokok dibandingkan dengan yang tidak pernah melihat orang merokok di dalam ruang publik tertutup (95%CI: 1,37 – 2,19). Serta remaja sekolah yang pernah melihat orang merokok di ruang publik terbuka mempunyai resiko 1,52 kali lebih besar untuk merokok dibandingkan dengan yang tidak pernah melihat orang merokok di ruang publik terbuka (95%CI: 1,19 – 1,93). Pengaruh lingkungan yang perlu menjadi perhatian adalah bahwa remaja sekolah yang pernah melihat orang merokok di dalam rumah mempunyai resiko 2,17 kali lebih besar untuk merokok dibandingkan dengan

yang tidak pernah melihat orang merokok di dalam rumah (95%CI: 1,73 – 2,72).

Temuan ini sejalan dengan beberapa penelitian sebelumnya (Foraker, Patten, Lopez, Croghan, & Thomas, 2005; Reda et al., 2012; Ribeiro Sarmiento & Yehadji, 2015; Rudatsikira et al., 2008; Sreeramareddy, Kishore, Paudel, & Menezes, 2008). Tekanan sosial dari orang tua, guru, dan lingkungan sekitar dapat mempengaruhi psikologis remaja sehingga dapat mempengaruhi resiko mereka untuk mengkonsumsi rokok.

## KESIMPULAN DAN SARAN

Penelitian ini menemukan adanya prevalensi merokok yang cukup tinggi pada siswa remaja di Indonesia. Selain itu, penelitian ini menemukan bahwa resiko siswa untuk merokok lebih tinggi untuk laki-laki, siswa yang pernah melihat orang merokok di rumah, siswa yang pernah melihat orang merokok di dalam ruang publik tertutup, siswa yang pernah melihat orang merokok di ruang publik terbuka, siswa yang pernah melihat guru merokok di sekolah, siswa yang teman dekatnya merokok, dan siswa yang salah satu atau kedua orang tuanya merokok. Selain itu, semakin baik pengetahuan dan perilaku siswa terhadap bahaya merokok, semakin rendah resiko siswa untuk merokok.

Penelitian ini menemukan adanya kemudahan akses remaja sekolah dalam mendapatkan rokok. Hal ini mengindikasikan adanya pengawasan yang kurang baik oleh pemerintah dalam mengawasi peredaran rokok di masyarakat, walaupun sudah ada peraturan pemerintah yang dibuat untuk mengawasi hal tersebut. Sehingga dapat disarankan kepada pemerintah agar pengawasan peredaran rokok, khususnya penjualan rokok terhadap anak di bawah usia 18 tahun harus semakin diperketat.

Kemudian penelitian ini juga mendapatkan bahwa siswa yang memiliki teman dekat yang merokok cenderung lebih beresiko untuk merokok. Hasil ini mengindikasikan perlu adanya intervensi dan edukasi mengenai bahaya merokok pada

siswa sekolah menengah pertama. Kemudian perhatian juga perlu diperluas ke lingkungan sekolah dan tempat tinggal para siswa, untuk mengatasi masalah tekanan sosial di sekolah, rumah, dan di lingkungan sekitar yang dapat memberikan dampak psikologis bagi para remaja untuk mengkonsumsi rokok.

## DAFTAR PUSTAKA

- Barraclough, S. 1999. Women and tobacco in Indonesia. *Tobacco Control*, 8, 327–332. <https://doi.org/10.1136/tc.8.3.327>
- Bricker, J. B., Peterson, A. V., Andersen, M. R., Rajan, K. B., Leroux, B. G. dan Sarason, I. G. 2006. Childhood friends who smoke: Do they influence adolescents to make smoking transitions?. *Addictive Behaviors*, 31(5), 889–900. <https://doi.org/10.1016/j.addbeh.2005.07.011>
- Elders, M. J., Perry, C. L., Eriksen, M. P. dan Giovino, G. A. 1994. The report of the surgeon general: Preventing tobacco use among young people. *American Journal of Public Health*, 84(4), 543–547. <https://doi.org/10.2105/AJPH.84.4.543>
- Foraker, R. E., Patten, C. A., Lopez, K. N., Croghan, I. T., & Thomas, J. L. 2005. Beliefs and attitudes regarding smoking among young adult Latinos: a pilot study. *Preventive Medicine*, 41(1), 126–133. <https://doi.org/10.1016/j.ypmed.2004.10.018>
- Gilliland, F. D., Islam, T., Berhane, K., Gauderman, W. J., McConnell, R., Avol, E., & Peters, J. M. 2006. Regular Smoking and Asthma Incidence in Adolescents. *American Journal of Respiratory and Critical Care Medicine*, 174(10), 1094–1100. <https://doi.org/10.1164/rccm.200605-722OC>
- Global Youth Tobacco Survey Collaborating Group. 2003. Differences in Worldwide Tobacco Use by Gender: Findings from the Global Youth Tobacco Survey. *Journal of School Health*, 73(6), 207–215. <https://doi.org/10.1111/j.1746-1561.2003.tb06562.x>
- Goodchild, M., Nargis, N., & Tursan d’Espaignet, E. 2017. Global economic cost of smoking-attributable diseases. *Tobacco Control*, tobaccocontrol-2016-053305. <https://doi.org/10.1136/tobaccocontrol-2016-053305>
- Hou, X., Xu, X. dan Anderson, I. 2015. Determinants of tobacco consumption in Papua New Guinea: challenges in changing behaviors, 2(2), 1–23. <https://doi.org/10.1002/app5.85>
- Kleinbaum, D. G. dan Klein, M. 2010. *Logistic regression: a self-learning text*. Springer.
- Rachmat, Muhammad., Thaha, Ridwan Mochtar., Syafar, M. 2013. Perilaku Merokok Remaja Sekolah Menengah Pertama. *Jurnal Kesehatan Masyarakat Nasional*, 7(11), 502–508. <https://doi.org/10.21109/kesmas.v7i11.363>
- Reda, A. A., Moges, A., Yazew, B. dan Biadgilign, S. 2012. Determinants of cigarette smoking among school adolescents in eastern Ethiopia: a cross-sectional study. *Harm Reduction Journal*, 9(1), 39. <https://doi.org/10.1186/1477-7517-9-39>
- Ribeiro Sarmiento, D. dan Yehadji, D. 2015. An analysis of global youth tobacco survey for developing a comprehensive national smoking policy in Timor-Leste. *BMC Public Health*, 16(1), 65. <https://doi.org/10.1186/s12889-016-2742-5>
- Rudatsikira, E., Dondog, J., Siziya, S., & Muula, A. S. 2008. Prevalence and determinants of adolescent cigarette smoking in Mongolia. *Singapore Medical Journal*, 49(1), 57–62. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18204771>
- Sreeramareddy, C. T., Kishore, P., Paudel, J., & Menezes, R. G. 2008. Prevalence and correlates of tobacco use amongst junior collegiates in twin cities of

western Nepal: A cross-sectional, questionnaire-based survey. *BMC Public Health*, 8(1), 97. <https://doi.org/10.1186/1471-2458-8-97>

TCSC-IAKMI. 2014. *Bunga Rampai : Fakta Tembakau dan Permasalahannya*. Kemenkes RI. Jakarta

WHO. 2014. *WHO | Research for universal health coverage: World health report 2013*. WHO. World Health Organization.

WHO. 2015. *Global Youth Tobacco Survey*

(GYTS): Indonesia report 2014. Who-Searo.

[https://doi.org/http://www.searo.who.int/tobacco/documents/ino\\_gyts\\_report\\_2014.pdf](https://doi.org/http://www.searo.who.int/tobacco/documents/ino_gyts_report_2014.pdf)

WHO | Health effects of smoking among young people. 2011. WHO. Retrieved from

[http://www.who.int/tobacco/research/youth/health\\_effects/en/](http://www.who.int/tobacco/research/youth/health_effects/en/)

# PERSEPSI MASYARAKAT KELURAHAN BUKIT DURI TERHADAP PROGRAM NORMALISASI KALI CILIWUNG DI JAKARTA TAHUN 2017 SERTA VARIABEL-VARIABEL YANG MEMENGARUHINYA

Loveria Candra Puspita<sup>1</sup> dan Achmad Prasetyo<sup>2\*</sup>

<sup>1</sup>Badan Pusat Statistik, Jakarta

<sup>2</sup>Dosen Statistika, Sekolah Tinggi Ilmu Statistik, Jakarta

\*e-mail: praze@stis.ac.id

## Abstrak

Salah satu cara menangani masalah banjir adalah melakukan program normalisasi sungai. Namun, tidak semua masyarakat menerima program ini. Untuk itu, ingin diketahui persepsi masyarakat terhadap normalisasi Kali Ciliwung dan menganalisa variabel-variabel yang memengaruhinya. Data persepsi diperoleh melalui survei dengan pendekatan rumahtangga di Kelurahan Bukit Duri yang kemudian dianalisis dengan regresi logistik. Hasil penelitian menunjukkan bahwa 28 persen rumahtangga sekitar sungai dan 22 persen rumahtangga bukan sekitar sungai menolak normalisasi. Persepsi rumahtangga sekitar sungai secara signifikan dipengaruhi oleh jenis kelamin, keikutsertaan organisasi, mendapatkan sosialisasi, dan pengeluaran perkapita. Sedangkan persepsi rumah tangga yang tinggal bukan di sekitar sungai dipengaruhi status pekerjaan, keikutsertaan organisasi, dan mendapatkan sosialisasi.

**Kata kunci** : persepsi, normalisasi sungai, regresi logistik

## Abstract

*River normalization program is one of the ways to handle flood problems. However, not all communities accept this program. For that, we want to know the public perception towards normalization of Ciliwung River and analyze the variables that influence it. Perception data was obtained through survey with household approach in Bukit Duri Village which then analyzed by logistic regression. The results show that 28 percent of households around the river and 22 percent of households not around the river reject normalization. Household perceptions around the river are significantly influenced by sex, organizational participation, socialization, and per capita expenditure. The non-rivers are influenced by employment status, organizational participation, and socialization.*

**Keywords** : *perception, river normalization, logistic regression*

## PENDAHULUAN

Salah satu sungai yang melewati provinsi DKI Jakarta adalah Kali Ciliwung. Selain melewati Jakarta, Kali Ciliwung juga melewati wilayah Bogor dan Depok. Kali Ciliwung terbentang dari hulu yang terletak di Bogor yang meliputi kawasan Gunung Gede, Gunung Pangrango dan Cisarua hingga kawasan hilir pantai utara Jakarta memiliki panjang 120 km dengan luas Daerah Aliran Sungai (DAS) 387 km<sup>2</sup>. Menurut Guru Besar Fakultas Teknik UI, Ilyas (2013), saat ini ada berbagai bangunan yang dibangun di tepi tebing Kali Ciliwung. Padahal, sebaiknya pada jarak 10 meter dari tepi lereng tidak diperbolehkan dibangun bangunan karena sangat beresiko menimbulkan longsor. Banyaknya bangunan liar di sepanjang bantaran kali juga menimbulkan tingginya sampah serta limbah yang akan mengotori Kali Ciliwung.

Sampah serta limbah merupakan salah satu penyebab terjadinya banjir. Sampah serta limbah tersebut akan menyumbat aliran air di Kali Ciliwung sehingga menyebabkan volume air tidak dapat ditampung dan terjadi banjir. Deputi Bidang Sarana dan Prasarana, Direktorat Pengairan dan Irigasi mengatakan bahwa dari berbagai kajian yang telah dilakukan, banjir yang melanda daerah-daerah rawan, pada dasarnya disebabkan tiga hal. Pertama, kegiatan manusia yang menyebabkan terjadinya perubahan tata ruang dan berdampak pada perubahan alam. Kedua, peristiwa alam seperti curah hujan sangat tinggi, kenaikan permukaan air laut, badai, dan sebagainya. Ketiga, degradasi lingkungan seperti hilangnya tumbuhan penutup tanah pada *catchment area*, pendangkalan sungai akibat sedimentasi, penyempitan alur sungai dan sebagainya.

Berdasarkan data kejadian banjir yang dikumpulkan Dinas Tata Kota DKI Jakarta dan diolah oleh Badan Informasi Geografi, banjir besar menimpa Jakarta dengan *return period* 5 tahun, yaitu pada tahun 2002 dan 2007. Dari kejadian banjir tersebut, daerah yang tergenang banjir pada tahun 2002 dan 2007 dikategorikan sebagai lokasi rawan banjir sedangkan daerah yang hanya tergenang banjir pada tahun 2007

dikategorikan sebagai lokasi yang cukup rawan banjir. Menurut data Badan Penanggulangan Bencana Daerah (BPBD), bahwa pada tahun 2013, 2014 dan 2015, Kelurahan Bukit Duri selalu terdampak banjir selama tiga tahun terakhir sehingga dikategorikan sebagai daerah rawan banjir. Oleh sebab itu, warga Kelurahan Bukit Duri yang berada di bantaran Kali Ciliwung menjadi target penggusuran dalam rangka program normalisasi Kali Ciliwung. Sebanyak empat RW (RW 09, RW 10, RW 11, RW 12) di kelurahan tersebut akan digusur dan direlokasi ke Rusun Rawa Bebek. Namun pada pelaksanaannya, program normalisasi Kali Ciliwung menuai banyak respon, salah satunya adalah penolakan dari warga korban penggusuran.

Berdasarkan hal tersebut di atas maka tujuan dari penelitian ini adalah mengetahui gambaran umum rumah tangga Kelurahan Bukit Duri berdasarkan karakteristik persepsi terhadap program normalisasi Kali Ciliwung? dan mengidentifikasi faktor-faktor apa saja yang signifikan berpengaruh terhadap persepsi masyarakat menerima program normalisasi Kali Ciliwung dan kecenderungannya? Berkaitan dengan tujuan tersebut, dalam penelitian ini memiliki keterbatasan yaitu sulitnya menemukan keberadaan korban penggusuran di Kelurahan Bukit Duri sehingga persepsi korban penggusuran diperoleh dari seluruh warga Kelurahan Bukit Duri.

Menurut Walgito (2003), faktor fisiologis akan menentukan bagaimana sikap seseorang. Umur merupakan salah satu faktor fisiologis. Dimana seseorang yang umurnya lebih muda akan cenderung untuk melakukan perbuatan radikal dibandingkan seseorang yang berumur tua. Sehingga, kecenderungan untuk tidak setuju terhadap normalisasi Kali Ciliwung lebih besar pada orang-orang yang berumur muda. Peran jenis kelamin juga sangat berpengaruh terhadap pilihan seseorang dalam menentukan persepsi, dimana Parsons (1955) dalam Sarwono (2002) menyatakan bahwa kepribadian yang diharapkan ada pada laki-laki berdasarkan norma baku diantaranya adalah dominan, mandiri,

kompetitif, dan asertif, karena laki-laki diharapkan menjadi pencari nafkah dan palindung untuk keluarganya. Sebaliknya, perempuan diharapkan baik hati, senang mengasuh, suka bekerja sama, dan peka terhadap perasaan orang lain, karena perempuan diharapkan menjadi istri dan ibu yang mengurus rumah tangga dan anak-anak. Sehingga, laki-laki cenderung untuk berperilaku agresif terhadap segala sesuatu yang merugikan dirinya atau bahkan keluarga. Sebaliknya, perempuan cenderung menerima apapun yang terjadi pada dirinya.

Dalam penelitiannya, Kidamu (2015) menyimpulkan bahwa terdapat hubungan yang baik antara kecerdasan emosional dan pengambilan keputusan. Begitu juga dengan Purmaningsih (2016) berpendapat bahwa tingkat pendidikan berpengaruh positif terhadap persepsi. Sementara itu, Robbins (2003) menyatakan bahwa perbedaan pekerjaan yang dimiliki seseorang memengaruhi mereka dalam membuat suatu penilaian. Penilaian tersebut akan membentuk persepsi dari masing-masing individu.

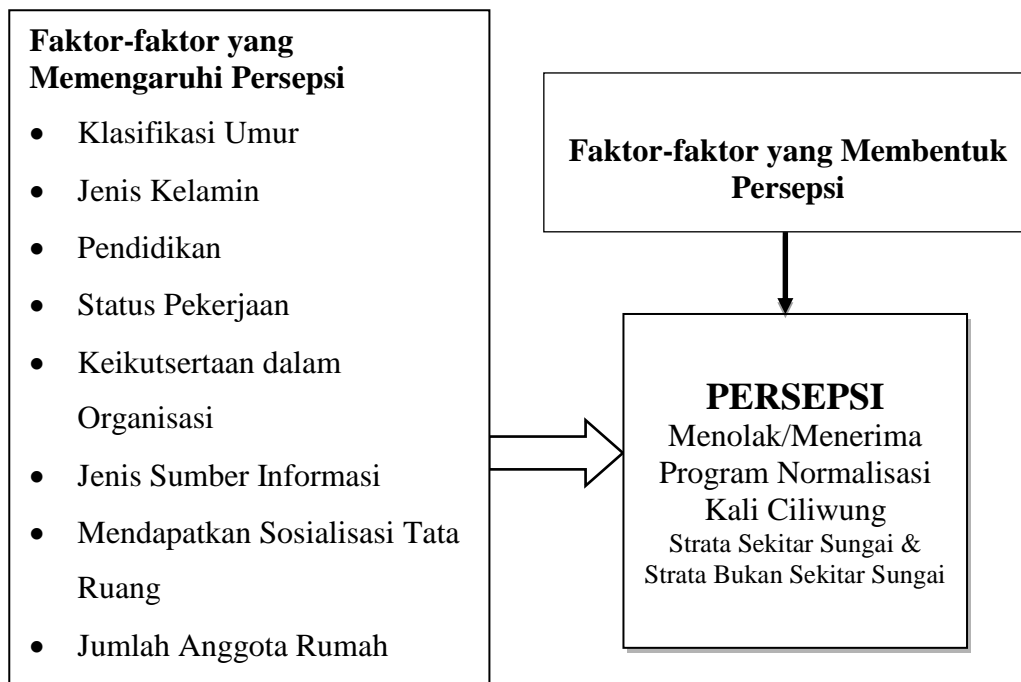
Penelitian yang dilakukan oleh Rahayu (2011) menyatakan bahwa persepsi tidak hanya dibentuk melalui lingkungan keluarga, namun juga dari lingkungan sosial masyarakat. Di dalam lingkungan masyarakat, seseorang akan memperoleh pengaruh dari budaya yang ada. Sehingga, pengaruh tersebut akan membentuk persepsi yang ada dalam masing-masing individu. Oleh karena itu, organisasi masyarakat sebagai wadah bersosialisasi antar masyarakat akan menambah peranan untuk memengaruhi sikap atau persepsi individu lainnya. Menurut Hidayat (2012), sumber informasi berperan penting bagi seseorang dalam menentukan sikap atau keputusan bertindak. Sumber informasi itu ada di mana-mana, di pasar-pasar, sekolah, rumah, lembaga-lembaga suatu organisasi komersial, buku-buku, majalah, surat kabar,

perpustakaan dan tempat-tempat lainnya. Intinya dimana suatu benda atau peristiwa berada, disana bisa tercipta informasi yang kemudian direkam dan disimpan melalui media cetak ataupun media elektronik.

Hasil analisis yang dilakukan oleh Puspita (2016) menunjukkan bahwa variabel sosialisasi berpengaruh secara langsung terhadap variabel kepatuhan sebesar 19,4 persen, variabel sosialisasi berpengaruh secara tidak langsung terhadap variabel kepatuhan melalui variabel kesadaran sebesar 38,4 persen. Sehingga dapat disimpulkan bahwa sosialisasi akan efektif meningkatkan kepatuhan apabila sosialisasi tersebut efektif meningkatkan kesadaran. Sejalan dengan itu, Luali (2006), dalam penelitiannya mengenai pengaruh faktor sosial ekonomi terhadap persepsi partisipasi masyarakat dalam pengelolaan sampah mendapatkan hasil bahwa semakin kecil jumlah anggota keluarga, semakin besar pengaruhnya terhadap persepsi.

Pola pengeluaran penduduk merupakan informasi untuk melihat kesejahteraan penduduk. Besarnya nilai nominal yang dibelanjakan baik dalam bentuk pangan maupun non pangan, secara tidak langsung dapat mencerminkan kemampuan ekonomi rumah tangga, untuk mencukupi kebutuhan yang mencakup barang dan jasa (Aminuddin, 2006). Sehingga, keadaan ekonomi yang dilihat dari rata-rata pengeluaran perkapita akan memengaruhi persepsi orang tersebut.

Dari penjelasan diatas dapat digambarkan bahwa persepsi masyarakat terhadap program normalisasi Kali Ciliwung dipengaruhi oleh klasifikasi umur, jenis kelamin, pendidikan, status pekerjaan kepala rumah tangga, keikutsertaan dalam organisasi, sumber informasi, mendapatkan sosialisasi tata ruang, jumlah anggota rumah tangga, dan rata-rata pengeluaran, seperti terlihat pada Gambar 1 dibawah ini:



Gambar 1. Kerangka pikir

## METODOLOGI

Data yang digunakan pada penelitian ini adalah data primer dan data sekunder. Data primer diperoleh dari survei di Kelurahan Bukit Duri. Sementara itu, data sekunder diperoleh dari registrasi Kantor Kelurahan Bukit Duri berupa daftar nama RT dan registrasi RT berupa daftar nama rumah tangga. Selanjutnya dengan

menggunakan peta wilayah Kelurahan Bukit Duri maka populasi rumah tangga di Kelurahan Bukit Duri dikategorikan ke dalam dua strata yaitu rumah tangga sekitar Kali Ciliwung dan rumah tangga yang bukan sekitar Kali Ciliwung, sehingga teknik sampling yang digunakan adalah *Stratified Two Stages Sampling* seperti pada Tabel 1 di bawah ini:

Tabel 1. Metode Penarikan Sampel

Strata	Keterangan
[1]	[2]
RT sekitar Kali Ciliwung	<ol style="list-style-type: none"> <li>1. Tahap pertama adalah pemilihan RT sekitar Kali Ciliwung menggunakan <i>systematic sampling</i> diurutkan berdasarkan nomor RW-RT.</li> <li>2. Tahap kedua adalah pemilihan rumah tangga di RT terpilih menggunakan <i>systematic sampling</i>.</li> </ol>
RT bukan sekitar Kali Ciliwung	<ol style="list-style-type: none"> <li>1. Tahap pertama adalah pemilihan RT bukan sekitar Kali Ciliwung menggunakan <i>systematic sampling</i> diurutkan berdasarkan nomor RW-RT.</li> <li>2. Tahap kedua adalah pemilihan rumah tangga di RT terpilih menggunakan <i>systematic sampling</i>.</li> </ol>

Menurut Asra dan Prasetyo (2015), untuk menentukan jumlah sampel dapat digunakan rumus sebagai berikut:

$$\begin{aligned}
 n &= \frac{n_{SRS-WR} \cdot Deff}{Z_{\frac{\alpha}{2}}^2 P(1-P)} \\
 &= \frac{2}{E^2} Deff
 \end{aligned}
 \tag{1}$$



Keterangan:

n = ukuran sampel

$n_{SRS-WR}$  = ukuran sampel dengan metode SRS-WR

Deff = rasio antara varians penduga

$Z_{\frac{\alpha}{2}}$  = nilai tabel normal baku

P = proporsi populasi

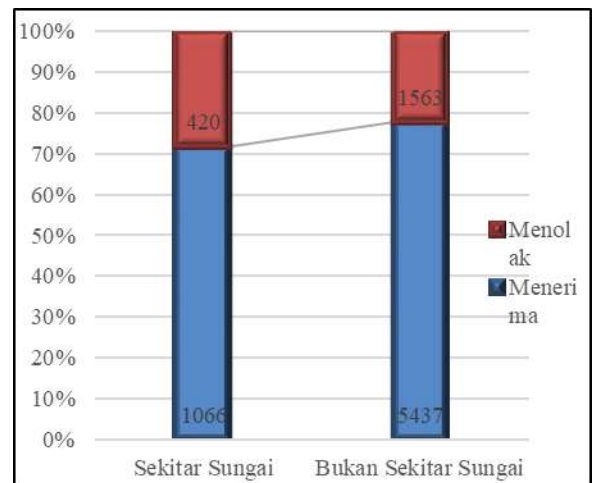
E = *margin of error*

Dalam penelitian ini, digunakan  $E=0,1$  dan  $\alpha=5\%$ . Berdasarkan rumus di atas, didapatkan minimum sampel pada penelitian ini sebesar 192,08. Oleh karena itu, sampel rumah tangga yang diambil di Kelurahan Bukit Duri sebanyak 200 rumah tangga. Sampel di Kelurahan Bukit Duri dibagi menjadi dua sampel yang terdiri dari 100 sampel rumah tangga di sekitar Kali Ciliwung dan 100 sampel rumah tangga di bukan sekitar Kali Ciliwung.

## HASIL DAN PEMBAHASAN

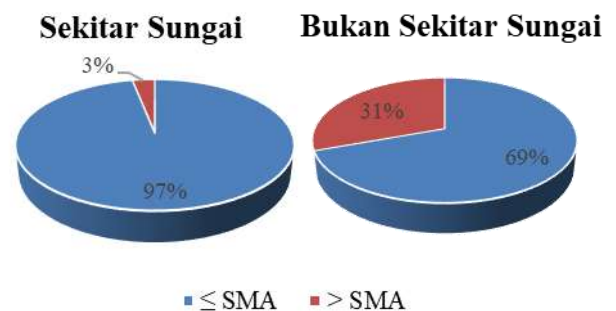
Rumah tangga sekitar sungai adalah rumah tangga yang bertempat tinggal di wilayah RT yang berbatasan langsung dengan Kali Ciliwung. Secara umum terlihat bahwa sebagian besar persepsi rumah tangga menerima adanya program normalisasi Kali Ciliwung. Program normalisasi Kali Ciliwung yang telah selesai dilaksanakan di Kelurahan Bukit Duri dianggap telah memberikan dampak positif bagi lingkungan setempat. Sebagian besar masyarakat menganggap bahwa program tersebut telah mengurangi bencana banjir yang pada tahun-tahun sebelumnya sering terjadi di Kelurahan Bukit Duri.

Terdapat sebagian persepsi warga yang menolak normalisasi Kali Ciliwung. Salah satu penyebab warga menolak program normalisasi ini adalah penggusuran yang menyebabkan sebagian warga kehilangan tempat tinggal mereka. Pada stratifikasi wilayah sekitar sungai, terdapat 28 persen rumah tangga yang menolak adanya normalisasi Kali Ciliwung. Sedangkan pada stratifikasi bukan sekitar sungai atau wilayah yang tidak berbatasan langsung dengan Kali Ciliwung, persentase yang menolak adanya normalisasi sungai sebesar 22 persen.



**Gambar 2. Jumlah rumah tangga berdasarkan persepsi dan stratifikasi wilayah di Kelurahan Bukit Duri tahun 2017**

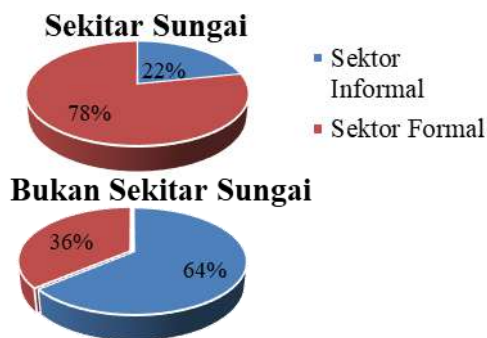
Pada Gambar 3 di bawah ini terlihat bahwa di stratifikasi wilayah sekitar sungai hanya terdapat 3 persen responden yang menolak adanya normalisasi sungai dengan pendidikan SMA ke atas. Sedangkan di stratifikasi wilayah bukan sekitar sungai ada sebesar 31 persen yang menolak adanya normalisasi sungai dengan pendidikan SMA ke atas.



**Gambar 3. Persentase rumah tangga yang menolak normalisasi Kali Ciliwung berdasarkan pendidikan yang ditamatkan dan stratifikasi wilayah tahun 2017**

Berdasarkan status pekerjaan utama, status pekerjaan dikelompokkan menjadi dua yaitu formal dan informal. Sektor formal adalah seseorang memiliki status pekerjaan sebagai buruh/karyawan/pegawai ataupun

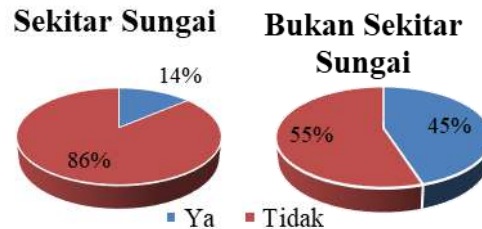
berusaha dibantu buruh tetap/buruh dibayar. Sedangkan, seseorang bekerja pada sektor informal adalah memiliki status pekerjaan berusaha sendiri, berusaha dibantu buruh tidak tetap/buruh tidak dibayar, pekerja bebas di pertanian/non pertanian, ataupun pekerja keluarga/tak dibayar. Gambar 4 menunjukkan bahwa persepsi yang menolak di wilayah sekitar sungai didominasi oleh rumah tangga yang bekerja di sektor formal yaitu sebesar 78 persen. Sedangkan, sejumlah 22 persen lainnya bekerja di sektor informal. Berbeda dengan wilayah sekitar sungai, wilayah bukan sekitar sungai memiliki persepsi menolak yang didominasi oleh rumah tangga dengan sektor pekerjaan kepala rumah tangga adalah sektor informal yaitu sebesar 64 persen. Sedangkan, sejumlah 36 persen rumah tangga yang menolak normalisasi Kali Ciliwung merupakan rumah tangga yang bekerja di sektor formal.



**Gambar 4.** Persentase rumah tangga yang menolak normalisasi Kali Ciliwung berdasarkan status pekerjaan kepala rumah tangga dan stratifikasi wilayah tahun 2017

Pada Gambar 5 dapat terlihat bahwa rumah tangga dengan persepsi menolak normalisasi Kali Ciliwung yang tidak mengikuti organisasi sosial jauh lebih banyak dibandingkan yang mengikuti organisasi sosial. Dari keseluruhan rumah tangga yang menolak normalisasi Kali Ciliwung, terdapat 86 persen diantaranya tidak mengikuti organisasi sosial dan 14 persen sisanya mengikuti organisasi sosial. Namun, wilayah bukan sekitar sungai menunjukkan perbedaan yang cukup sedikit antara komposisi mengikuti organisasi sosial dengan tidak mengikuti organisasi sosial pada rumah tangga yang menolak

normalisasi Kali Ciliwung. Hanya terdapat 55 persen rumah tangga dengan persepsi menolak normalisasi yang tidak mengikuti organisasi sosial. Sedangkan, sejumlah 45 persen sisanya mengikuti organisasi sosial di lingkungannya.



**Gambar 5.** Persentase rumah tangga yang menolak normalisasi Kali Ciliwung berdasarkan keikutsertaan organisasi sosial dan stratifikasi wilayah tahun 2017

### Strata Sekitar Sungai

Dengan menggunakan analisis regresi logistik metode *backward*, didapatkan empat dari sembilan variabel penjelas dalam penelitian ini masuk dalam model regresi logistik dan mempengaruhi secara signifikan persepsi rumah tangga di sekitar Kali Ciliwung terhadap normalisasi Kali Ciliwung, yaitu variabel jenis kelamin, keikutsertaan organisasi sosial, mendapatkan sosialisasi, dan pengeluaran perkapita. Untuk lebih jelasnya dapat dilihat pada Tabel 2 berikut:

**Tabel 2.** Pendugaan parameter, statistik uji Wald dan *odds ratio* sekitar sungai

Variabel	Dummy	$\beta$	Stat. Uji Wald	P value	Odds ratio [Exp( $\beta$ )]
(1)	(2)	(3)	(4)	(5)	(6)
Jenis Kelamin	X <sub>2</sub>	1,430	5,738	0,017	4,181
Keikutsertaan Organisasi Sosial	X <sub>5</sub>	2,456	8,892	0,003	11,662
Sosialisasi Normalisasi Kali Ciliwung	X <sub>7</sub>	-1,926	9,804	0,002	0,146
Pengeluaran Perkapita	X <sub>9</sub>	-2,212	5,847	0,016	0,109
Constant		-1,204	1,769	0,184	0,300

Adapun persamaan peluang regresi logistik yang terbentuk sebagai berikut

$$\hat{g}(x) = -1,204 + 1,43 X_2 + 2,456 X_5 - 1,926 X_7 - 2,212 X_9$$

Keterangan:

- $X_2$  : *dummy* untuk variabel jenis kelamin
- $X_5$  : *dummy* untuk variabel keikutsertaan organisasi sosial
- $X_7$  : *dummy* untuk variabel mendapatkan sosialisasi
- $X_9$  : variabel pengeluaran perkapita

Untuk mengetahui besarnya pengaruh dan kecenderungan variabel penjelas yang berpengaruh terhadap kecenderungan persepsi, dapat dilihat dari nilai  $\exp(\beta)$ . Nilai ini disebut juga dengan *odds ratio* atau rasio kecenderungan seperti yang terdapat pada Tabel 2. Berdasarkan nilai koefisien dari satu variabel penjelas yang signifikan memengaruhi persepsi, dengan menganggap variabel-variabel lain konstan, nilai *odds ratio* untuk keikutsertaan jenis kelamin adalah 4,181. Nilai tersebut dapat diartikan bahwa laki-laki memiliki kecenderungan untuk menolak normalisasi Kali Ciliwung sebesar 4,181 kali dibandingkan perempuan. Variabel lain yang signifikan adalah keikutsertaan organisasi sosial yang memiliki nilai *odds ratio* sebesar 11,662. Artinya, kecenderungan seseorang yang tidak mengikuti organisasi sosial adalah 11,662 kali dari seseorang yang mengikuti organisasi sosial untuk menolak normalisasi Kali Ciliwung. Pengeluaran perkapita juga merupakan variabel yang signifikan memengaruhi persepsi. Nilai *odds ratio* variabel pengeluaran perkapita adalah 0,109 dan bernilai negatif. Sehingga, setiap penurunan 1 juta pengeluaran perkapita akan memiliki kecenderungan 9,134 kali untuk menolak normalisasi Kali Ciliwung. Nilai *odds ratio* untuk sosialisasi adalah 0,146 dengan nilai  $\beta$  negatif yang berarti bahwa kecenderungan seseorang yang mendapatkan sosialisasi memilih untuk menolak normalisasi Kali Ciliwung adalah 6,862 kali dibandingkan seseorang yang tidak mendapatkan sosialisasi.

### Strata Bukan Sekitar Sungai

Variabel keikutsertaan organisasi sosial dan mendapatkan sosialisasi di strata bukan sekitar sungai signifikan memengaruhi persepsi sama halnya dengan di strata sekitar sungai. Variabel lainnya adalah status pekerjaan yang signifikan memengaruhi persepsi di strata bukan sekitar sungai. Hal tersebut dapat terlihat pada Tabel 3 berikut:

**Tabel 3. Pendugaan parameter, statistik uji Wald dan odds ratio strata bukan sekitar sungai**

Variabel	Dummy	$\beta$	Stat. uji Wald	Significance	Odds ratio [Exp ( $\beta$ )]
(1)	(2)	(3)	(4)	(5)	(6)
Status Pekerjaan	$X_4$	1,344	4,486	0,034	3,833
Keikutsertaan Organisasi Sosial	$X_5$	-1,624	5,771	0,016	0,197
Sosialisasi Normalisasi Kali Ciliwung	$X_7$	-1,317	5,698	0,017	0,268
Constant		0,115	0,041	0,840	1,122

Adapun persamaan peluang regresi logistik yang terbentuk sebagai berikut

$$\hat{g}(x) = 0,115 + 1,344 X_4 - 1,624 X_5 - 1,317 X_7$$

Keterangan:

- $X_4$  : *dummy* untuk variabel status pekerjaan
- $X_5$  : *dummy* untuk variabel keikutsertaan organisasi sosial
- $X_7$  : *dummy* untuk variabel mendapatkan sosialisasi

Nilai yang berbeda didapatkan pada strata bukan sekitar sungai. Pada tabel 2 dan tabel 3 dapat terlihat perbedaan antara kedua strata. Pada strata bukan sekitar sungai, variabel status pekerjaan signifikan berpengaruh terhadap kecenderungan menolak normalisasi Kali Ciliwung. Nilai *odds ratio* variabel status pekerjaan pada penelitian ini bernilai 3,833. Hal ini menunjukkan bahwa seseorang yang bekerja di sektor informal lebih cenderung menolak normalisasi Kali Ciliwung sebesar 3,833 kali

dibandingkan seseorang yang bekerja di sektor formal. Variabel lain yang signifikan adalah keikutsertaan organisasi sosial yang memiliki nilai *odds ratio* sebesar 0,197 dengan nilai  $\beta$  negatif. Artinya, kecenderungan seseorang yang mengikuti organisasi sosial untuk menolak normalisasi Kali Ciliwung adalah 5,073 kali dibandingkan seseorang yang tidak mengikuti organisasi sosial. Sementara itu untuk variabel sosialisasi nilai *odds ratio*-nya adalah 0,268 dengan nilai  $\beta$  negatif. Hal tersebut menunjukkan bahwa kecenderungan seseorang yang mendapatkan sosialisasi normalisasi Kali Ciliwung adalah 3,734 kali dari seseorang yang tidak mendapatkan sosialisasi untuk menolak normalisasi Kali Ciliwung. Hubungan yang negatif antara penerimaan sosialisasi dengan persepsi terhadap normalisasi ini sama dengan hubungan yang terjadi pada strata sekitar sungai yang hampir sebagian besar warga penerima sosialisasi merupakan korban yang merasa kehilangan tempat tinggalnya dan akan cenderung menolak normalisasi Kali Ciliwung.

## KESIMPULAN DAN SARAN

Sebagian besar persepsi rumah tangga menerima adanya program normalisasi Kali Ciliwung. Namun, masih terdapat 28 persen rumah tangga menolak adanya normalisasi Kali Ciliwung pada stratifikasi wilayah sekitar sungai. Sedangkan, persentase di wilayah bukan sekitar sungai yang menolak adanya normalisasi sungai sebesar 22 persen rumah tangga.

Selanjutnya terdapat delapan faktor yang membentuk persepsi masyarakat Kelurahan Bukit Duri terhadap program normalisasi Kali Ciliwung antara lain pertimbangan manfaat, kemampuan beradaptasi, sarana transportasi dan informasi, pertimbangan risiko, keyakinan sikap, pengetahuan, perilaku sosial, dan proses penerimaan.

Pada wilayah sekitar Kali Ciliwung, persepsi masyarakat Kelurahan Bukit Duri terhadap program normalisasi Kali Ciliwung secara signifikan dipengaruhi oleh faktor-

faktor antara lain jenis kelamin, keikutsertaan organisasi, mendapatkan sosialisasi, dan pengeluaran perkapita. Sedangkan, persepsi masyarakat Kelurahan Bukit Duri pada wilayah bukan sekitar sungai terhadap program normalisasi Kali Ciliwung secara signifikan dipengaruhi oleh faktor-faktor antara lain status pekerjaan, keikutsertaan organisasi, dan mendapatkan sosialisasi.

Berdasarkan kesimpulan tersebut maka pemerintah DKI Jakarta perlu memberikan sosialisasi secara menyeluruh kepada semua elemen masyarakat baik yang menjadi korban ataupun bukan agar mengetahui tentang manfaat program normalisasi Kali Ciliwung. Selain itu perlu memberikan kompensasi yang cukup untuk warga korban penggusuran di wilayah sekitar sungai agar tidak mempersulit kehidupan selanjutnya di tempat yang baru.

## DAFTAR PUSTAKA

- Aditya, Nicky. (2016, Oktober 5). Keluhan dari Rusun Rawa Bebek, Warga: Banyak Begal. *Kriminalitas.com*. (Diakses 22 Desember 2016). <http://kriminalitas.com/keluhan-dari-rusun-rawa-bebek-warga-banyak-begal/>
- Agresti, A. 2002. *Categorical Data Analysis Second Edition*. New Jersey: John Wiley & Sons, Inc
- Ahmad, Dalili Atika, et al. 2015. Analisis Persepsi dan Faktor yang Mempengaruhi Persepsi terhadap Penerapan Sistem Pembiayaan JKN pada Fasilitas Kesehatan Penunjang di D. I. Yogyakarta. *Journal of Management and Pharmacy Practice*, 5(4), 259-266
- Aliyati, Ratu. 2011. Permukiman Kumuh di Bantaran Ci-Liwung (Studi Kasus Kel. Manggarai-Srengseng Sawah dan Kel. Kampung Melayu-Kalisari) *Tesis*. Universitas Indonesia : Depok
- Aminuddin. 2006. *Pembangunan Ekonomi*. Ghalia Indonesia. Jakarta.

- Ardiyanto, Elvinaro, dkk. 2004. *Komunikasi Massa: Suatu Pengantar*. Simbiosia Rekatama Media
- Arfina, Onik. 2012. Analisis Perbedaan Persepsi Siswa Berdasarkan Usia, Gender, Jenis Pekerjaan, dan Lama Kursus terhadap Komunikasi Word Of Mouth *Skripsi*. Universitas Diponegoro: Semarang
- Asra, Abuzar dan Prasetyo, Achmad. 2015. *Pengambilan Sampel dalam Penelitian Survei*. Jakarta: Raja Grafindo Persada
- BPS. 2014. *Statistik Daerah Provinsi DKI Jakarta 2014*. BPS: Jakarta
- \_\_\_\_\_. 2016. Konsep ketenagakerjaan. BPS. Diakses pada tanggal 29 Januari 2017 melalui <http://bps.go.id/Subjek/view/id/6#subjekViewTab1|accordion-daftar-subjek1>
- Bappenas. Kebijakan Penanggulangan Banjir di Indonesia (Kajian). Bappenas. (Diakses 14 Januari 2017) melalui <https://bebasbanjir2025.wordpress.com/konsep-pemerintah/beppenas/>
- BPBD. 2013. *Data Rekapitulasi Kejadian Banjir Tahun 2013*. BPBD: Jakarta
- \_\_\_\_\_. 2014. Daerah Rawan Banjir DKI Jakarta. BPBD. (Diakses 25 November 2016) melalui <http://data.go.id/dataset/daerah-rawan-banjir-dki-jakarta>
- Dinas Tata Kota DKI Jakarta. 2007. *Data Kejadian Banjir*.
- Hidayat, Khairul. 2012. *Perilaku Pencarian Informasi Guru dalam Memanfaatkan Internet Untuk Memenuhi Kebutuhan Informasi di SMA Negeri 2 Lubuk Pakam* [Skripsi]. Sumatera Utara: Universitas Sumatera Utara
- Hosmer, D.W. dan S. Lemeshow. 2000. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc
- Ilyas, Tommy. 2013. Sungai Ciliwung Kini. <http://www.ui.ac.id/feature/sungai-ciliwung-kini.html> (diakses 11 Januari 2017)
- Johnson, R.A. dan Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Education Inc
- Khotimah, Husnul, dkk. 2016. Pengaruh Sosialisasi dan Pengetahuan Terhadap Minat Investor Pada Efek Syariah di Pasar Modal. *Account*, 423-433
- Kidamu, Nella. 2015. Hubungan Kecerdasan Emosional Kepala Sekolah dengan Pengambilan Keputusan di SD se Kecamatan Kota Selatan *Skripsi*. Gorontalo: Universitas Negeri Gorontalo
- Liputan 6. (2016, September 28). Video: 44 Keluarga Masih Bertahan dari Penggusuran Bukit Duri. *Liputan 6*. Diakses pada tanggal 22 Desember 2016 melalui <http://tv.liputan6.com/read/2613346/video-44-keluarga-masih-bertahan-dari-penggusuran-bukit-duri?source=search>
- Luaili, La Ode. 2006. Pengaruh faktor sosial ekonomi terhadap persepsi, sikap, dan partisipasi masyarakat dalam pengelolaan sampah : Kasus Kota Raha Kab. Muna Prov. Sulawesi Tenggara. *Tesis*. Yogyakarta: Universitas Gadjah Mada
- Malasari, Eka. 2015. Faktor-Faktor Penyebab Rendahnya Partisipasi Masyarakat Dalam Pembangunan Desa (Studi Desa Kembang Gading Kecamatan Abung Selatan Kabupaten Lampung Utara). *Skripsi*. Lampung: Universitas Lampung
- Maryono, Agus. 2009. Kajian Lebar Sempadan Sungai (Studi Kasus Sungai-sungai di Provinsi Daerah Istimewa Yogyakarta). *Dinamika Teknik Sipil*, 9(1), 56-66
- Mashita, Nani. (2016, September 28). Digusur, Warga Bukit Duri Gelar Demo Sambil Teriak Allahuakbar. *Lensa Indonesia*. (Diakses 22 Desember 2016) melalui

<http://www.lensaindonesia.com/2016/09/28/digusur-warga-bukit-duri-gelar-demo-sambil-teriak-allahu-akbar.html>

- Mulyani, Sri. 2015. Analisis Pengaruh Jenis Kelamin dan Status Pekerjaan terhadap Persepsi Etis Mahasiswa Akuntansi dengan *Love of Money* sebagai Variabel *Intervening*. *Majalah Ilmiah Solusi*, 14(3), 2-16
- Muslim AR. (2016, Juni 7). Minta Putusan Sela, Warga Bukit Duri Kecewa di Sidang Perdana. *Liputan 6*. (Diakses 22 Desember 2016) melalui <http://news.liputan6.com/read/2525676/minta-putusan-sela-warga-bukit-duri-kecewa-di-sidang-perdana?source=search>
- Normadewi, Berliana. 2012. *Analisis Pengaruh Jenis Kelamin dan Tingkat Pendidikan terhadap Persepsi Etis Mahasiswa Akuntansi dengan Love of Money sebagai Variabel Intervening* [Skripsi]. Semarang: Universitas Diponegoro
- Nugroho, Y.P. (2008). *Makna Sungai dan Praktek Pengelolaan Lingkungan Melalui Pendekatan Budaya (Studi Kasus Masyarakat Sempadan Sungai Code, Kotamadya Yogyakarta)* [Tesis]. Jakarta:Universitas Indonesia
- Poerbandono, dkk. (2014). Assessment of the effects of climate and land cover changes on river discharge and sediment yield, and an adaptive spatial planning in the Jakarta region. *Springer Science & Business Media B.V*, 73, 507-530
- Pontiawati, Ike, dkk. (2009). Manajemen Resiko Pada Pengendalian Banjir di Sungai Ciliwung. *Jurnal Teknologi UNPAK*, 1, 46-68
- Prastiwi, Devira. (29 September 2016). Wakil Ketua DPR: Penggusuran Bukit Duri Melanggar Hukum. *Liputan 6*. Diakses pada tanggal 22 Desember 2016 melalui <http://news.liputan6.com/read/2613571/wakil-ketua-dpr-penggusuran-bukit-duri-melanggar-hukum?source=search>
- Purnamaningsih, Ni Ketut Ayu. 2016. Pengaruh Gender, Usia, Tingkat Pendidikan, dan Status Sosial Ekonomi terhadap Persepsi Etis Mahasiswa Akuntansi. *Skripsi*. Universitas Udayana: Denpasar
- Puspita, Erna. 2016. Analisis jalur pengaruh sosialisasi terhadap kepatuhan wajib pajak bumi dan bangunan kota kediri dengan kesadaran sebagai variabel *intervening*. *Jurnal Akuntansi Dan Ekonomi*, 1, 1-8
- Putra, N.P. (2016, September 28). Komnas HAM Sebut Pemprov DKI Langgar Hukum Bongkar Bukit Duri. *Liputan 6*. (Diakses 22 Desember 2016) <http://news.liputan6.com/read/2613071/komnas-ham-sebut-pemprov-dki-langgar-hukum-bongkar-bukit-duri?source=search>
- \_\_\_\_\_. (2016, September 5). Warga Bukit Duri Keluhkan Tak Ada Musala di Rusun Rawa Bebek. *Liputan 6*. (Diakses 22 Desember 2016) melalui <http://news.liputan6.com/read/2594475/warga-bukit-duri-keluhkan-tak-ada-musala-di-rusun-rawa-bebek?source=search>
- Rahayu, Rehasti Dya, et al. 2011. Pengaruh Lingkungan Keluarga, Sekolah, dan Masyarakat terhadap Persepsi Gender Mahasiswa Laki-laki dan Perempuan. *Jurnal Transdisiplin Sosiologi, Komunikasi, dan Ekologi Manusia*, 5(3), 247-260. Diakses pada tanggal 30 Januari 2017 melalui [download.portalgaruda.org](http://download.portalgaruda.org)
- Revrison Baswir, et al. 2003. *Pembangunan tanpa perasaan: Evaluasi pemenuhan hak ekonomi, sosial, dan budaya*. Yogyakarta: Sabda Media
- Robbins, Stephen P. 2003. *Perilaku Organisasi*. Jakarta: Erlangga
- Rozaqi, Athok Moh Nur. 2009. Sosialisasi Kebijakan Pembangunan Pemerintah Kabupaten Bojonegoro Kepada

- Masyarakat: Studi Model Komunikasi Pembangunan. *Skripsi*. UIN Sunan Ampel Surabaya.
- Santoso, Alexander Budi. 2012. *Hubungan antara Tingkat Pendidikan Orang Tua dengan Minat Siswa dalam Bermusik di Smp N 5 Depok Sleman Yogyakarta [Skripsi]*. Yogyakarta: Universitas Negeri Yogyakarta
- Sari, Eka Puspita. 2015. Peran Media Massa dan Fungsinya Sebagai Agen Sosialisasi Gender. *Jurnal Ilmu Berbagi*, 3, 1-9
- Sarwono, S.W. 2006. *Pengantar Psikologi Umum*. Jakarta: Rajawali Pers
- Sudyasih, Tiwi, et al. 2015. Hubungan antara Status Sosial Ekonomi dengan Persepsi Masyarakat tentang Gantung Diri di Kecamatan Karangmojo Kabupaten Gunungkidul Yogyakarta. *Jurnal Ilmu Kebidanan dan Keperawatan*, 11(2), 177-183
- Sugiyono. 2005. *Memahami Penelitian Kualitatif*. Bandung: Alfabeta
- Syahputra, Ichsan. 2015. Kajian Hidrologi dan Analisa Kapasitas Tampang Sungai Krueng Langsa Berbasis HEC-HMS dan HEC-RAS. *Jurnal Teknik Sipil Universitas Abulytama*
- Tamara, Riana Monalisa. 2016. Peranan Lingkungan Sosial terhadap Pembentukan Sikap Peduli Lingkungan Peserta Didik di SMA Negeri Kabupaten Cianjur. *Jurnal Pendidikan Geografi*, 16(1), 44-55
- Undang-Undang No. 24 Tahun 1992
- Universitas Indonesia. 2013. Sungai Ciliwung Kini. *Universitas Indonesia*. Diakses pada tanggal 19 Desember 2016  
<http://www.ui.ac.id/feature/sungai-ciliwung-kini.html>
- Walgito, Bimo. 2003. *Pengantar Psikologi Umum*. Andi Yogyakarta: Yogyakarta
- Walpole, R.E. 1993. *Pengantar Statistika*. Jakarta: Gramedia Pustaka Utama
- Wibowo. 1987. *Psikologi Sosial*. Jakarta





# NAMED ENTITY RECOGNITION ON A COLLECTION OF RESEARCH TITLES

**Siti Mariyah**

The Center of Computational Statistics Study, Institute of Statistics, Jakarta-Indonesia 13330  
e-mail: [sitimariyah@stis.ac.id](mailto:sitimariyah@stis.ac.id)

## Abstrak

Judul dapat membantu pembaca untuk mendapatkan sudut pandang universal dari artikel tersebut sebagai pemahaman awal sebelum membaca konten secara keseluruhan. Pada penelitian teknis, judul memuat informasi penting. Dalam penelitian ini, kami mengembangkan teknik ekstraksi informasi untuk mengenali dan mengekstrak masalah, metode, dan domain penelitian yang terdapat dalam judul. Kami menerapkan pendekatan supervised learning pada 671 judul penelitian dalam bidang ilmu komputer dari beragam jurnal online dan prosiding seminar internasional. Kami melakukan beberapa percobaan dengan skema yang berbeda untuk mempelajari pengaruh fitur dan kinerja algoritma. Kami menguji fitur kontekstual, fitur sintaksis, dan fitur bag of words menggunakan Naïve Bayes dan Maximum Entropy. Classifier Naïve Bayes yang belajar dari kelompok set fitur pertama berhasil memprediksi kategori masing-masing token dalam dataset judul. Keakuratan dan nilai f1-score untuk setiap kelas lebih dari 0,80 karena kelompok pertama set fitur mempertimbangkan lokasi token dalam sebuah kalimat, memperhatikan token sekitar dan tag POS dari beberapa token sebelum dan sesudah. Sementara classifier Naïve Bayes yang dipelajari dari kelompok kedua dari rangkaian fitur lebih tepat mengklasifikasikan token frase daripada token kata.

**Kata Kunci:** *research titles, named entity recognition, information extraction, contextual features, naïve bayes classifier*

## Abstract

*The title can help the reader to get the universal point of view of the article as the initial understanding before reading the content as a whole. On technical research papers, the title states essential information. In this study, we aim to develop information extraction techniques to recognize and extract problem, method, and domain of research contained in a title. We apply supervised learning on 671 research titles in computer science from various online journals and international conference proceedings. We conducted some experiments with different schemas to discover the influence of features and the performance of the algorithm. We examined contextual, syntactic, and the bag of words feature sets using Naïve Bayes and Maximum Entropy. The Naïve Bayes classifier learned from the first group of the feature set is successful in predicting category of each token in title dataset. The accuracy and f1-score for each class are more than 0.80 since the first group of feature sets considers the location of a token within a sentence, considers the token and POS tag of some tokens before and after and deliberates the rules of a token. While the Naïve Bayes classifier learned from the second group of the feature set is more appropriate classifying a phrase token than a word token.*

**Keywords:** *research titles, named entity recognition, information extraction, contextual features, naïve bayes classifier*

## INTRODUCTION

Research title is a short sentence which can help the reader to get the main or universal point of view of the article as the initial understanding before reading the content as a whole. The title is also commonly used as a filter in a search engine when there is a retrieval query against a research paper in online journals or online archives. On technical research papers such as in computer science or engineering, the title states essential information. That information consists of the research problem, the method used or method proposed, and the specific research domain. A reader or a researcher should know the problem, method, and domain of research regarding the topic she/he is studying or focusing.

On the other side, information extraction opens the opportunity to extract words or phrases that are regarded as informative words or phrases. Informative means that the word or phrase describes the information a reader want to know. Information extraction technique involves a collection of natural language processing (NLP) tasks. Each method may include different NLP task which depends on the complexity of information, the format of the document, and the task itself, etc. There are three approaches to build information extraction technique, i.e., rule-based extraction, statistical or machine learning-based extraction, or hybrid approach.

In this study, we aimed to develop information extraction techniques to recognize and extract problem, method, and domain of research contained in a title. We apply supervised learning as a part of statistical or machine learning-based approach on 671 research titles in computer science from ACM Digital Library, IEEE, and some international conference proceedings. By using some learning algorithms, we constructed some named entity recognition (NER) models. Machine learning based extraction can handle the knowledge acquisition bottleneck since, in rule-based extraction, we need to construct extraction rules which requires the domain experts. The NER model identifies the

property of each word in the title then classify it into some defined categories. We conducted some experiments with different schemas to learn the influence of features and the performance of the algorithm. In this paper, we technically describe how we built the information extraction techniques in detail and suggest some recommendations which one is the best feature and model.

## LITERATURE REVIEW

NER was first introduced in the Sixth Message Understanding Conference (MUC-6) held in November 1995. Two of four goals are named entity recognition and scenario templates (traditional information extraction). NER task comprises the recognition of entity names of people, names of company or organization, place names, temporal expressions and a particular type of numerical expressions.

Suakkaphong et al. (2009) built disease named entity recognizer They used three feature sets. The first feature set is a morphological-pattern feature since biomedical terms commonly have unique prefixes and suffixes. The remaining features are word appearance and chunking and POS tag features. Then, They combined conditional random field (CRF) with bootstrapping and feature sampling. CRFs with bootstrapping implemented sequentially is more accurate than supervised CRFs.

Biomedical named entity recognition was also done by Saha et al. (2009) and Bodenreider et al. (2000). They hypothesized that the appropriate feature templates affect the performance of NER models. They conducted word clustering and selection based feature reduction approaches for NER using Maximum Entropy algorithm. The feature sets are generated without involving profound biomedical knowledge such as word feature, previous NE tags, capitalization and digit information, unique character, word normalization, prefix and suffix information, Part of Speech (POS) tags, and trigger words. They proved that the use of

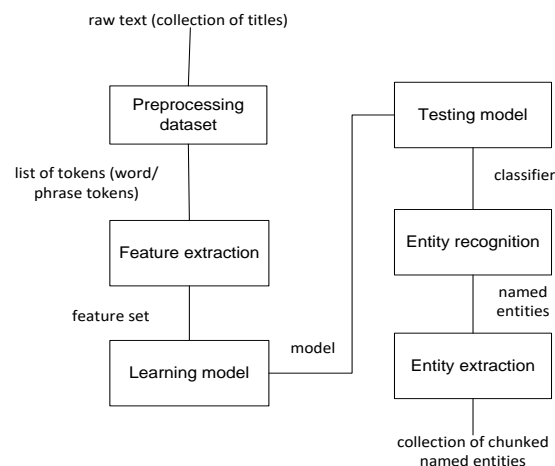
dimensionality reduction techniques could increase the performance substantially.

Bodenreider, Olivier, and Pierre Zweigenbaum (2000) developed methods to collect proper names used in biomedical terminology. The task is recognizing a word that is the appropriate name by using individual criteria owned by that word and some combination of these different criteria (capitalization, invariant words, and patterns).

Another relevant work was done by Ek et al. (2011) who conducted NER for short text messages. The characteristics of the short text message are similar to title sentence which has small windows (a few of words). They constructed regular expression and complemented with logistic regression classifier. Wu et al. (2005) used POS tag as feature set. Researches of McKenzie (2013), Mao, Xinnian et al. (2007) and Qin et al. (2008) utilized the contextual feature sets to either improve the NER results in the large-scale corpus or to reduce the noise introduced into aggregated features from disparate and generic training data. They proved that the missed entities occur when their contextual surroundings are not identified well. NER using machine learning approach are more frequent conducted than other methods. There are learning algorithms applied for NER or text classification tasks such Naïve Bayes or Multinomial Naïve Bayes performed by Fabrizio Sebastiani (2001) and Amarappa S, and Sathyanarayana S.V. (2015), Maximum Entropy applied by Ayan et al. (2006), Conditional Random Fields performed by Mao, Xinnian et al. (2007), Qin et al. (2008), and Chodey et al. (2016), Support Vector Machines applied by Fabrizio Sebastiani (2001), Thorsten Joachims (1998), and Rafi et al. (2012).

## METHODS

Extraction technique was developed by involving some tasks depicted by this following diagram:



**Figure 1. Extraction techniques**

It starts from collecting dataset. We gathered 671 research titles in computer science fields from some online journals or online archives. Then the dataset will be processed in some following tasks:

### 1. Preprocessing dataset

The dataset was validated to ensure there were no double titles. Then we conducted annotation on the dataset to tag the words or the phrases which explain problem, method, and domain of research. Annotation was done by humans who are familiar with computer science research. We tagged `<m>...</m>` for words explaining method, `<p>...</p>` for words explaining problem, and `<d>...</d>` for words explaining domain. The annotated dataset was then validated to make sure that there was no missed annotation or wrong annotation. The missed annotation means that there is a token that is not annotated. The wrong annotation means that there is a token annotated by the wrong label. By using the regular expression, we split the annotated dataset into four files. Each file contains 671 lines where each line contains the words in one category. It aimed to check whether every title contains full information (problem, method, and domain) or not.

Then, we tokenized every title sentence, made part-of-speech-tag (POS tag) for each token and mapped token with the label it owns. We labeled m for tokens flanked by `<m>...</m>` tag, p for tokens flanked by `<p>...</p>` tag, d for tokens flanked by `<d>...</d>` tag, and none for

tokens not flanked by any tag. Output in this step is collection of tokens per sentence who have its each label. We focused and used the word tokens only rather than the phrase tokens.

## 2. Feature extraction

The output from processing dataset stage is the input for this feature extraction step. The feature is information which characterizes a token. The features used significantly affect the accuracy of the classification model. We were curious which features accurately differentiate each category. We extracted some features and grouped it into two groups of the feature set. Then, these two groups would be tested with some experiments to know which group is the most relevant.

The first group of feature set:

1. Feature word: the token itself
2. Feature POS tag
3. Feature prevWord: one token before
4. Feature prevTag: POS tag of one token before
5. Feature prevBigram: two tokens before
6. Feature prevBigramTag: POS tag of two tokens before
7. Feature nextWord: one token after
8. Feature nextBigram: two tokens after
9. Feature nextTag: POS tag of one token after
10. Feature nextBigram: POS tag of two tokens after

The second group of feature set was the list resulted matching the existence of a token in a collection of the method, problem, and domain tokens. If a token exists in that collection, then the value is true. Otherwise, the value is false. The number of extracted features equals the number of tokens owned in 671 research titles. This is the example of how to extract this feature set:

The first title:

```
<m>simple algorithms</m> for <p>complex relation extraction</p> with applications to <d>biomedical ie</d>
```

The second title:

```
<m>a seed-driven bottom-up machine learning</m> framework for <p>extracting relations of various complexity</p>
```

Therefore, The method, problem, domain and none tokens are:

```
Method tokens: simple algorithms, a seed-driven bottom-up machine learning
Problem tokens: complex relation extraction, extracting relations of various complexity
Domain tokens: biomedical ie
None tokens: for with application to, framework for
```

If want to extract feature from phrase “extracting relations of various complexity”, the extracted feature is:

```
{ contain(simple): False, contain(algorithms): False, contain(a): False, contain(seed-driven): False, contain(bottom-up): False, contain(machine): False, contain(learning): False, contain(complex): False, contain(relation): False, contain(extraction): False, contain(extracting): True, contain(relations): True, contain(of): True, contain(various): True, contain(complexity): True, contain(biomedical): False, contain(ie): False, contain(for): False, contain(with): False, contain(application): False, contain(to): False, contain/framework): False, contain(for): False }
```

If want to extract feature from phrase “biomedical ie”, the extracted feature is:

```
{ contain(simple): False, contain(algorithms): False, contain(a): False, contain(seed-driven): False, contain(bottom-up): False, contain(machine): False, contain(learning): False, contain(complex): False, contain(relation): False, contain(extraction): False, contain(extracting): False, contain(relations): False, contain(of): False, contain(various): False, contain(complexity): False, contain(biomedical): True, contain(ie): True, contain(for): False, contain(with): False, contain(application): False, contain(to): False, contain/framework): False, contain(for): False }
```

## 3. Learning and testing model

In this stage, we prepared training set. The training set is a collection of extracted feature for each token in dataset then mapped with the label owned by the token. If in title dataset consists of 1000 tokens then we have 1000 feature set mapped with the label. We applied Naïve Bayes, Maximum Entropy, and Support Vector Machines using two groups of the feature set with shuffling parameter. The classification models were learned and tested by 10-fold cross-validation. We measured precision, recall, and f-measure for each category to understand the effect of shuffling parameter, the performance of feature set and algorithm.

## 4. Entity recognition and extraction

The best model is then used as a classifier which recognizes and classify every token in title sentences into problem, method, domain or none category. If any token in sentence classified as a problem,

method, or domain category, our program then chunked the sentence into tokens and extracted those tokens.

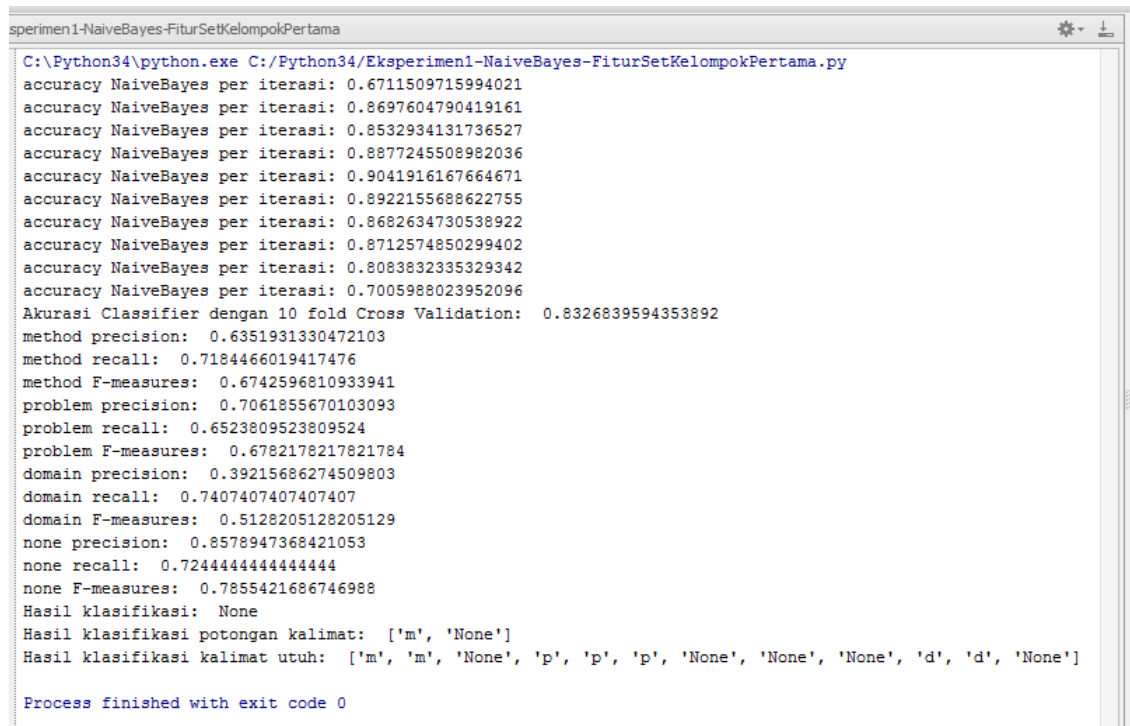
## EXPERIMENTAL STUDY

We conducted some experiments with some different conditions. The difference is defined by feature set used, shuffling

parameter and machine learning algorithm applied.

### 1. The first experiment

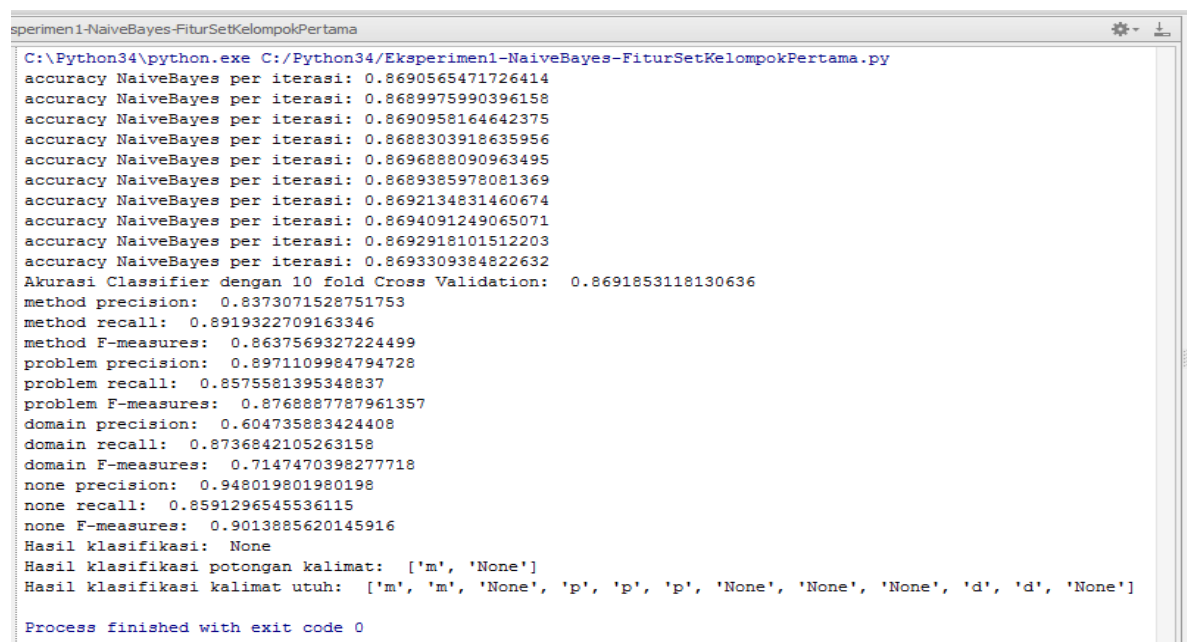
On the first experiment, we built classification model using the first group of feature set and Naïve Bayes algorithm. We applied 10-fold cross validation both on shuffled and non-shuffled training data. Shuffling the training data cause the order of the data to be random. The results are:



```
sperimen1-NaiveBayes-FiturSetKelompokPertama
C:\Python34\python.exe C:/Python34/Eksperimen1-NaiveBayes-FiturSetKelompokPertama.py
accuracy NaiveBayes per iterasi: 0.6711509715994021
accuracy NaiveBayes per iterasi: 0.8697604790419161
accuracy NaiveBayes per iterasi: 0.8532934131736527
accuracy NaiveBayes per iterasi: 0.8877245508982036
accuracy NaiveBayes per iterasi: 0.9041916167664671
accuracy NaiveBayes per iterasi: 0.8922155688622755
accuracy NaiveBayes per iterasi: 0.8682634730538922
accuracy NaiveBayes per iterasi: 0.8712574850299402
accuracy NaiveBayes per iterasi: 0.8083832335329342
accuracy NaiveBayes per iterasi: 0.7005988023952096
Akurasi Classifier dengan 10 fold Cross Validation: 0.8326839594353892
method precision: 0.6351931330472103
method recall: 0.7184466019417476
method F-measures: 0.6742596810933941
problem precision: 0.7061855670103093
problem recall: 0.6523809523809524
problem F-measures: 0.6782178217821784
domain precision: 0.39215686274509803
domain recall: 0.7407407407407407
domain F-measures: 0.5128205128205129
none precision: 0.8578947368421053
none recall: 0.7244444444444444
none F-measures: 0.7855421686746988
Hasil klasifikasi: None
Hasil klasifikasi potongan kalimat: ['m', 'None']
Hasil klasifikasi kalimat utuh: ['m', 'm', 'None', 'p', 'p', 'p', 'None', 'None', 'None', 'd', 'd', 'None']

Process finished with exit code 0
```

Figure 2. Learning performance on first group of feature set using Naïve Bayes with shuffle



```
sperimen1-NaiveBayes-FiturSetKelompokPertama
C:\Python34\python.exe C:/Python34/Eksperimen1-NaiveBayes-FiturSetKelompokPertama.py
accuracy NaiveBayes per iterasi: 0.8690565471726414
accuracy NaiveBayes per iterasi: 0.8689975990396158
accuracy NaiveBayes per iterasi: 0.8690958164642375
accuracy NaiveBayes per iterasi: 0.8688303918635956
accuracy NaiveBayes per iterasi: 0.8696888090963495
accuracy NaiveBayes per iterasi: 0.8689385978081369
accuracy NaiveBayes per iterasi: 0.8692134831460674
accuracy NaiveBayes per iterasi: 0.8694091249065071
accuracy NaiveBayes per iterasi: 0.8692918101512203
accuracy NaiveBayes per iterasi: 0.8693309384822632
Akurasi Classifier dengan 10 fold Cross Validation: 0.8691853118130636
method precision: 0.8373071528751753
method recall: 0.8919322709163346
method F-measures: 0.8637569327224499
problem precision: 0.8971109984794728
problem recall: 0.8575581395348837
problem F-measures: 0.8768887787961357
domain precision: 0.604735883424408
domain recall: 0.8736842105263158
domain F-measures: 0.7147470398277718
none precision: 0.948019801980198
none recall: 0.8591296545536115
none F-measures: 0.9013885620145916
Hasil klasifikasi: None
Hasil klasifikasi potongan kalimat: ['m', 'None']
Hasil klasifikasi kalimat utuh: ['m', 'm', 'None', 'p', 'p', 'p', 'None', 'None', 'None', 'd', 'd', 'None']

Process finished with exit code 0
```

Figure 3. Learning performance on first group of feature set using Naïve Bayes with no shuffle

**Table 1. The comparison of shuffle and no shuffle condition on first group of feature set using Naïve Bayes algorithm**

The measurements	Without shuffle	With shuffle
Classifier accuracy	0.83268	0.86919
Method precision	0.63519	0.83730
Method recall	0.71845	0.89193
Method F-Measures	0.67426	0.86376
Problem precision	0.70618	0.89711
Problem recall	0.65238	0.85755
Problem F-Measures	0.67821	0.87688
Domain precision	0.39216	0.64047
Domain recall	0.74074	0.87368
Domain F-Measures	0.51282	0.71475
None precision	0.85789	0.94802
None recall	0.72444	0.85913
None F-Measures	0.78554	0.90139

The table shows that the shuffle parameter causes the difference of classifier accuracy 0.03. It is aligned with the concept of fold cross validation which at every iteration, it divides the data into ten parts with nine parts as training and one as a testing set. The repetition is done until all elements have been a test set. The shuffle can affect the sampling of those parts. Our hypothesis is shuffle will minimize the probability a label does not appear in training set. It means that with shuffle, the distribution of the existence of each label is equal. Without shuffle, the process building up the members of 10 parts is done sequentially. Therefore, the probability of skewed distribution of category is higher.

Overall, recall values for all categories are above 0.85, and the difference of recall for each class is not significant. The precision values for the method, problem, and domain are 0.83730, 0.89711, and 0.64047. The precision for domain category is lower than others because the true positive is higher and false positive. After we evaluated the training set, the number of domain examples is more inferior than method and problem examples.

## 2. The second experiment

On the second experiment, we built classification model using the second group of feature set and Naïve Bayes algorithm. We applied 10-fold cross validation both on shuffled and non-shuffled training data.

**Table 2. The comparison of shuffle and no shuffle condition on the second group of feature set using Naïve Bayes algorithm**

The measurements	Without shuffle	With shuffle
Classifier accuracy	0.81323	0.86500
Method precision	0.0	0.92843
Method recall	None	0.73472
Method F-Measures	None	0.82029
Problem precision	0.0	0.90636
Problem recall	None	0.77963
Problem F-Measures	None	0.83823
Domain precision	0.0	0.71909
Domain recall	None	0.98807
Domain F-Measures	None	0.83234
None precision	1	0.98609
None recall	0.83146	0.95795
None F-Measures	0.90798	0.97182

Table 2 tells the performance of classifier from the second group of the feature set without and with the shuffle. The result of this experiment is much different with the last experiment. Without shuffle, the classifier failed to detect a problem, method, and domain tokens. It is explained by the values of precision, recall, and f-measures for all categories. If compared with the same treatment (with shuffle), this classifier learned from the first group of feature set performs almost equal with the classifier acquired from the second group of the feature set.

**Table 3. The comparison of the group of feature set using Naïve Bayes algorithm with shuffle**

The classifier	The First Group of Feature Set	The Second Group of Feature Set
Classifier accuracy	0.86919	0.86500
Method F-Measures	0.86376	0.82029
Problem F-Measures	0.87688	0.83823
Domain F-Measures	0.71475	0.83234
None F-Measures	0.90139	0.97182

Table 3 shows that the classifiers from two groups are almost similar. The first classifier is accurate for classifying method and problem tokens, while the second classifier is accurate for recognizing domain and none tokens. Our hypothesis is method and problem tokens are good explained with contextual and syntactic features. It means

that method and problem tokens may have regular tokens previous and after with regular POS tag.

### 3. The third experiment

On this experiment, we examined Maximum Entropy (MaxEnt) algorithm to validate the effect of different feature set on classifier. We trained the model with 10-fold cross validation and shuffle.

**Table 4. The comparison of the group of feature set using maximum entropy algorithm with shuffle**

The classifier	The First Group of Feature Set	The Second Group of Feature Set
Classifier accuracy	0.83975	0.25216
Method F-Measures	0.86918	None
Problem F-Measures	0.84124	None
Domain F-Measures	0.01047	0.40059
None F-Measures	0.88192	None

Table 4 tells us that accuracy classifier on the first group around 83.975% is better than on the second group of the feature set. It is aligned with the f-measures for the method, problem, and none categories. The interesting one is MaxEnt fails to classify domain category using the first group off; we feature set. It is caused by precision value for domain is 1.0, but the recall is 0.00526. It means that coverage ability of MaxEnt classifier for domain category is low. MaxEnt also miscarries the second group of the feature set.

From three experiments conducted, we concluded that Naïve Bayes classifier is robust on both the first and the second group of feature sets. Naïve Bayes classifier with the first group of feature set outperforms than others. It also delivers informative features. The informative feature means that the feature is the most significant feature in determining a token belongs to a category. The shuffle improves the performance a classifier than it is not shuffled.

The first group of feature set consists of a word, tag, prevWord, prevTag, prevBigram, prevBigramTag, nextWord, nextTag, nextBigram, nextBigramTag. Using Naïve Bayes with shuffle and 10-fold

cross validation, the accuracy acquired is 0.86919. It means that 86,919% of test set will classified correctly. The following descriptions are the explanation for every informative feature.

```

Most Informative Features
word = 'for'           None : p = 243.7 : 1.0
prevWord = 'for'      p : m = 211.5 : 1.0
prevBigram = '-'      m : d = 113.1 : 1.0
word = 'using'        None : m = 83.2 : 1.0
prevWord = '-'        m : d = 79.5 : 1.0
word = '.'            None : d = 72.2 : 1.0
nextWord = '-'        None : d = 70.0 : 1.0
nextTag = '-'         None : p = 57.1 : 1.0
word = 'a'            None : p = 40.9 : 1.0
prevWord = 'using'    m : p = 40.2 : 1.0
prevWord = 'from'     d : m = 38.8 : 1.0
prevBigramTag = '-'   m : d = 37.3 : 1.0
prevTag = '-'         m : d = 35.9 : 1.0
prevWord = 'in'       d : m = 34.4 : 1.0
nextBigramTag = 'IN JJ' m : d = 30.5 : 1.0
word = 'in'           None : d = 27.7 : 1.0
nextTag = 'VBG'       d : p = 26.9 : 1.0
prevBigramTag = 'NN NNS' m : d = 26.5 : 1.0
nextWord = 'for'      m : d = 25.8 : 1.0
nextWord = 'using'    p : m = 24.8 : 1.0

```

**Figure 4. The Most Informative Features from The First Group of Feature Set**

- The word 'for' appears 243 times on none class than problem class. It explains the word 'for' has high probability to be classified as none category and not belongs to problem, domain, and domain classes.
- PrevWord = 'for' occurs 211 times on problem class than on method class. It means that a word or a phrase preceded by the word 'for' has high chance to be classified as problem class.
- The third (prevBigram = '-'), the fifth (prevWord = '-'), the twelfth (prevBigramTag = '-'), and the thirteenth information (prevTag = '-') explain that a token which does have any previous token is more frequent classified as method class than domain class. It indicates that a word or a phrase at the beginning of the title sentence has high chance to be classified as method class. It is aligned with the fact. We observed directly some title sentences which prove this information.

The first title: <m>simple algorithms</m> for <p>complex relation extraction</p> with applications to <d>biomedical ie</d>

The second title: <m>a seed-driven bottom-up machine learning</m> framework for <p>extracting relations of various complexity</p>

- d. The tenth information (prevWord = ‘using’) appears 40 times on the method class than on the problem class. It shows that a word or a phrase preceded by the word ‘using’ has more chance to be classified as method class.
- e. The eleventh (prevWord = ‘from’) and the fourteenth information (prevWord = ‘in’) appear more than 30 times on domain class than on the method class. It describes that a word or a phrase preceded by the word ‘from’ or ‘in’ has a higher probability to be classified as domain class.
- f. The fifteenth (nextBigramTag = ‘IN JJ’) occurs 30 times and the eighteenth (prevBigramTag = ‘NN NNS’) appears 26 times on class method than on class domain. It indicates that a word or a phrase preceded by noun words will be classified as method class.
- g. The seventeenth (nextTag = ‘VBG’) occurs 26 times more on domain class than problem class. It means that a word or a phrase followed by gerund (verb + ‘ing’) has a higher probability to be classified as domain class.
- h. The nineteenth (nextWord = ‘for’) appears 25 times more on the method class and the twentieth information (nextWord = ‘using’) occurs 20 times on problem class. It indicates that a word or a phrase followed by the word ‘for’ will be classified as method class and followed by the word ‘using’ has higher chance to be classified as problem class.

Most Informative Features			
contain(for) = True	None : m	=	291.8 : 1.0
contain(using) = True	None : m	=	107.3 : 1.0
contain(in) = True	None : d	=	69.5 : 1.0
contain(a) = True	None : p	=	61.7 : 1.0
contain(extraction) = True	p : d	=	47.0 : 1.0
contain(and) = True	m : None	=	30.2 : 1.0
contain(classification) = True	p : d	=	26.3 : 1.0
contain(of) = True	p : d	=	24.6 : 1.0
contain(information) = True	p : d	=	22.2 : 1.0
contain(an) = True	None : p	=	21.7 : 1.0
contain(summarization) = True	p : d	=	20.3 : 1.0
contain(method) = True	m : p	=	17.7 : 1.0
contain(approach) = True	m : None	=	14.2 : 1.0
contain(knowledge) = True	m : None	=	13.0 : 1.0
contain(models) = True	m : d	=	12.3 : 1.0
contain(on) = True	None : p	=	12.1 : 1.0
contain(traffic) = True	p : d	=	11.0 : 1.0
contain(algorithm) = True	m : p	=	11.0 : 1.0
contain(detection) = True	p : None	=	11.0 : 1.0
contain(fuzzy) = True	m : p	=	9.7 : 1.0

**Figure 5. The Most Informative Features from The Second Group of Feature Set**

The picture tells about:

- a. If a word or a phrase is/contains a word ‘for’, ‘using’, ‘in’, ‘a’, or ‘an’, then the word or phrase has more chance to be classified as none class.
- b. If a word or a phrase is/contains a word ‘extraction’, ‘classification’, ‘information’, ‘summarization’, ‘traffic’, or ‘detection’, then then the word or phrase has higher chance to be classified as problem class.
- c. If a word or a phrase is/contains a word ‘method’, ‘approach’, ‘knowledge’, ‘models’, ‘algorithm’, or ‘fuzzy’, then the word or phrase has more chance to be classified as method class.

We conducted significance test to examine two hypotheses. The first hypothesis is the performance of Naïve Bayes and MaxEnt classifier learned from the first group of feature set is same. The second hypothesis is the performance of two classifiers are different, one classifier is better than another. This is the significance test algorithm:

1. The data is partitioned into  $k$  disjoint test sets  $T_1, T_2, \dots, T_k$  with same size. The minimum size is 30.
2. For  $i$  from 1 to  $k$ , do #  $k = 10$   
 Use  $T_i$  for the test set and the remaining data for training set  $S_i$   
 $S_i \leftarrow \{D_0 - T_i\}$  #  $S_i$  : training set  
 $h_A \leftarrow L_A(S_i)$  #  $L_A$ : Naïve Bayes classifier  
 $h_B \leftarrow L_B(S_i)$  #  $L_B$ : MaxEnt classifier  
 $\delta_i \leftarrow error_{T_i}(h_A) - error_{T_i}(h_B)$
3. Return:

$$\bar{\delta} = \frac{1}{k} \sum_{i=1}^k \delta_i$$

The result of  $\bar{\delta} = -0.029512$

Next step is measuring confidence interval. We took confidence interval 90% so that the confidence interval estimation for

$$\delta: \bar{\delta} \pm t_{N,k-1} s_{\bar{\delta}}$$

Where:

$$s_{\bar{\delta}} = \sqrt{\frac{1}{k(k-1)} \sum_{i=1}^k (\delta_i - \bar{\delta})^2}$$

$$s_{\bar{\delta}} = \sqrt{\frac{1}{10(10-1)} \sum_{i=1}^{10} (\delta_i - (-0.029512))^2}$$



$$= 7,86554E - 05$$

The value  $t_{N,k-1} = t_{90,9} = 1.833$  is acquired from t-table. The confidence interval is:

$$-0.029512 \pm 1.833*(7,86554E-05)$$

$$= -0.029512 \pm 0.000144$$

The upper limit of the interval:

$$-0.029512+0.000144 = -0.02936582$$

The lower limit of the interval is

$$-0.029512-0.000144 = -0.0296542$$

The error difference is -0.029512. It means that the error of Naïve Bayes classifier is less than MaxEnt classifier. The upper and lower limit of the interval has small range, approximately 0.000004. It shows that with 90% of confidence, we can conclude that Naïve Bayes classifier is better than MaxEnt classifier but the accuracy of both classifiers is not significant different.

After we got the best classifier, we conduct the post processing to extract the word or phrase belongs to method, problem, and domain categories on research title dataset. The post processing includes classification each token in every title sentence and token chunking. This is the example of post processing result:

Title sentence: large scale learning of relation extraction rules with distant supervision from the web

After classification: large **p** scale **p** learning **p** of **p** relation **p** extraction **p** rules **p** with **none** distant **m** supervision **m** from **none** the **none** web **d**

Chunking result:

Method class: distant supervision

Problem class: large scale learning of relation extraction rules

Domain class: web

To enrich analysis and answer the research problem, we examined the Naïve Bayes classifiers constructed from two groups of the feature set. We deliver the chunking results from four titles:

**Table 5. The post processing results of naïve bayes classifier constructed from the first group of feature set**

Title Sentence	Predicted class	Actual class
simple algorithms for complex relation extraction with applications to biomedical ie	['m', 'm', 'None', 'p', 'p', 'p', 'None', 'None', 'None', 'd', 'd']	['m', 'm', 'None', 'p', 'p', 'p', 'None', 'None', 'None', 'd', 'd']
a seed-driven bottom-up machine learning framework for extracting relations of various complexity	['m', 'm', 'd', 'm', 'm', 'm', 'None', 'p', 'p', 'p', 'p', 'p']	['m', 'm', 'm', 'm', 'm', 'None', 'None', 'p', 'p', 'p', 'p', 'p']
a machine learning approach for efficient traffic classification	['None', 'm', 'm', 'm', 'None', 'p', 'p', 'p']	['None', 'm', 'm', 'm', 'None', 'p', 'p', 'p']
ddos attack detection at local area networks using information theoretical metrics	['p', 'p', 'p', 'p', 'p', 'p', 'd', 'None', 'm', 'm', 'd']	['p', 'p', 'p', 'p', 'p', 'p', 'p', 'p', 'None', 'm', 'm', 'm']

Tables 5 shows that there is no wrong prediction on the 1st and the 3rd sentences. But on the 2nd and the 4th sentences, the Naïve Bayes classifier tends to misclassify the domain class.

**Table 6. The post processing results of naïve bayes classifier constructed from the second group of feature set**

Title Sentence	Predicted class	Actual class
simple algorithms for complex relation extraction with applications to biomedical ie	['d', 'd', 'None', 'd', 'p', 'p', 'None', 'd', 'None', 'd', 'd']	['m', 'm', 'None', 'p', 'p', 'p', 'None', 'None', 'None', 'd', 'd']
a seed-driven bottom-up machine learning framework for extracting relations of various complexity	['None', 'd', 'd', 'd', 'm', 'm', 'None', 'd', 'p', 'p', 'd', 'd']	['m', 'm', 'm', 'm', 'm', 'None', 'None', 'p', 'p', 'p', 'p', 'None']
a machine learning approach for efficient traffic classification	['None', 'd', 'm', 'm', 'None', 'd', 'd', 'p']	['None', 'm', 'm', 'm', 'None', 'p', 'p', 'p']
ddos attack detection at local area networks using information theoretical metrics	['d', 'd', 'p', 'd', 'd', 'd', 'd', 'None', 'p', 'd', 'd']	['p', 'p', 'p', 'p', 'p', 'p', 'p', 'None', 'm', 'm', 'm']

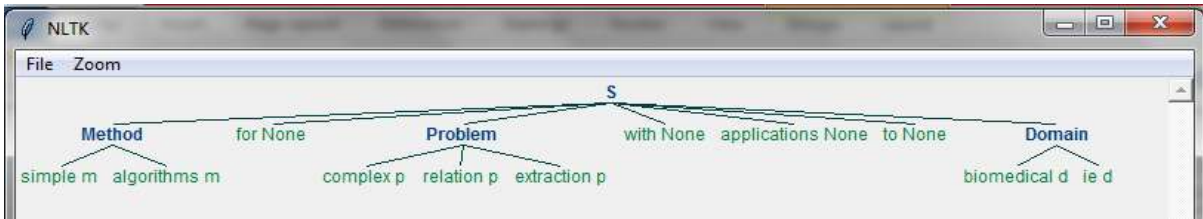


Figure 6. The chunking result of the first title using Naïve Bayes classifier learned from the first group of feature set



Figure 7. The chunking result of the second title using Naïve Bayes classifier learned from the first group of feature set



Figure 8. The chunking result of the third title using Naïve Bayes classifier learned from the first group of feature set

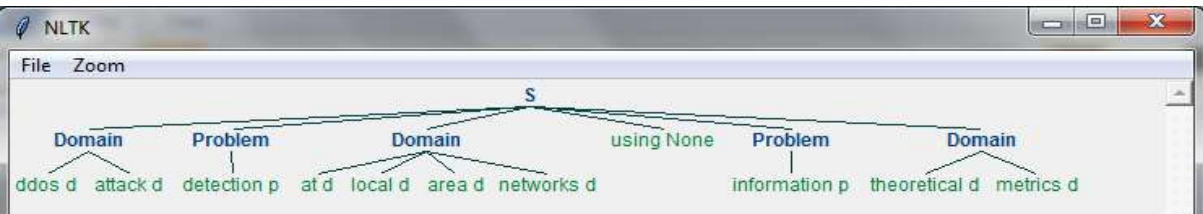


Figure 9. The chunking result of the fourth title using Naïve Bayes classifier learned from the first group of feature set

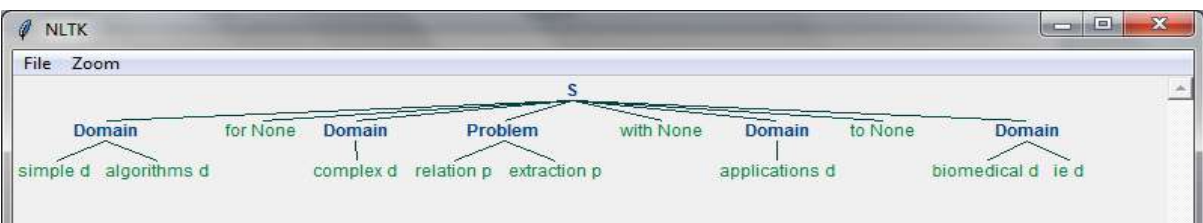
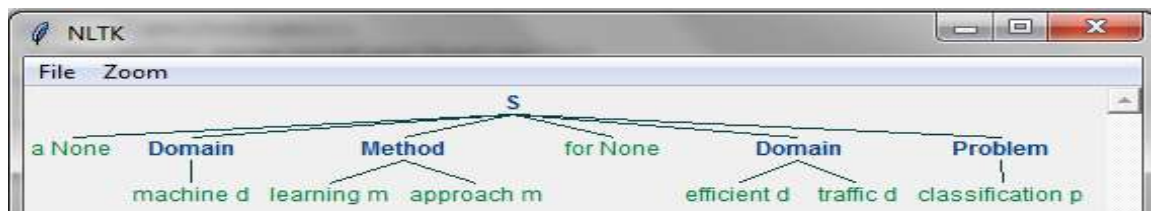


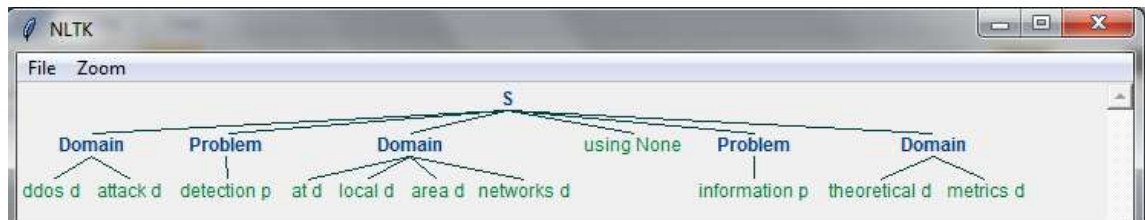
Figure 10. The chunking result of the first title using Naïve Bayes classifier learned from the second group of feature set



Figure 11. The chunking result of the second title using Naïve Bayes classifier learned from the second group of feature set



**Figure 12. The chunking result of the third title using Naïve Bayes classifier learned from the second group of feature set**



**Figure 13. The chunking result of the fourth title using Naïve Bayes classifier learned from the second group of feature set**

Table 6 tells us that Naïve Bayes classifier learned from the second group of feature set also tends to misclassify the domain class. The domain class is mostly classified as the method class. This classifier is not appropriate to predict class of a word instead of a phrase. If we examined to classified a phrase such as ‘biomedical ie’ or ‘complex relation extraction’ then this classifier will predict ‘biomedical ie’ as domain class and ‘complex relation extraction’ as problem class.

## RESULTS AND CONCLUSIONS

There are some aspects we learn from the experimental study. The first, the labeling process should be consistent since the inconsistent label for tokens can influence the modeling process and might worse the model itself. The annotated dataset has to be validated before it is used for modeling to check the consistency of labels and the completeness of labeled tokens. Shuffle on training set produces more accurate classifier than without shuffle because shuffle lets each category/class has equal data distribution on the dataset. Therefore, each class has its representatives on both the training and testing set.

On the small size dataset, the 10-fold cross validation is an appropriate method to construct and validate/test the models instead of holdout method. The Naïve Bayes

classifier learned from the first group of the feature set is successful in predicting category of each token in title dataset. The accuracy and f1-score for each class are more than 0.80 since the first group of feature set considers the contextual and syntactic feature of a token. This classifier determines the location of a token within a sentence, considers the token and POS tag of some tokens before and after and deliberates the rules of a token. While the Naïve Bayes classifier learned from the second group of the feature set is more appropriate classifying a phrase token than a word token. This classifier just considering the tokens owned by a phrase instead determines the characteristics of word token. The definition of the token in our experimental study is a word.

We believe that it is a good idea to try the same information extraction techniques we have built on the large title dataset from various research fields. We also encourage to conduct semi-supervised learning in classifier modeling because the cost for annotation is expensive. The idea is utilizing the limited annotated titles to construct a classifier then applying the ensemble methods to improve the performance of the classifier.

## REFERENCES

- Ayan, Necip Fazil, and Bonnie J. Dorr. 2006. A Maximum Entropy Approach to Combining Word Alignments.

- Proceedings of the Human Language Technology Conference of the NAACL, Main Conference* (June): 96–103.
- Bodenreider, Olivier, and Pierre Zweigenbaum. 2000. Identifying Proper Names in Parallel Medical Terminologies. *Studies in Health Technology and Informatics* 77: 443–47.
- Chodey, Krishna Prasad, and Gongzhu Hu. 2016. Clinical Text Analysis Using Machine Learning Methods. *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on*.
- Dimililer, Nazife, Ekrem Varoğlu, and Hakan Altınçay. 2009. Classifier Subset Selection for Biomedical Named Entity Recognition. *Applied Intelligence* 31(3): 267–82.
- Ek, Tobias, Camilla Kirkegaard, Håkan Jonsson, and Pierre Nugues. 2011. Named Entity Recognition for Short Text Messages. *Procedia - Social and Behavioral Sciences* 27(Pacling): 178–87.
- Joachims, Thorsten. 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In *The 10th European Conference on Machine Learning*, , 137–42.
- Mao, Xinnian et al. 2007. Using Non-Local Features to Improve Named Entity Recognition Recall. In *Proceedings of the 21st Pasific Asia Conference on Language, Information, and Computation*, 303–10. [http://dSPACE.wul.waseda.ac.jp/dSPACE/bits/tream/2065/29132/1/PACLIC\\_21\\_00\\_031\\_Mao.pdf](http://dSPACE.wul.waseda.ac.jp/dSPACE/bits/tream/2065/29132/1/PACLIC_21_00_031_Mao.pdf).
- McKenzie, Amber. 2013. Focused Training Sets to Reduce Noise in NER Feature Models. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, , 411–15. <http://www.aclweb.org/anthology/N13-1042>.
- Nadeau, D. 2007. A Survey of Named Entity Recognition and Classification. *Linguisticae Investigationes* (30): 3–26. <http://nlp.cs.nyu.edu/sekine/papers/li07.pdf>.
- Qin, Ying, Taozheng Zhang, and Xiaojie Wang. 2008. Chinese Named Entity Recognition with New Contextual Features. *2008 International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 2008*: 1–6.
- Rafi, Muhammad, Sundus Hassan, and Mohammad Shahid Shaikh. 2012. Content-Based Text Categorization Using Wikitology. *International Journal of Computer Science Issues* 9(4): 9. <http://arxiv.org/abs/1208.3623>.
- S, Amarappa, and Sathyanarayana S.V. 2015. Kannada Named Entity Recognition and Classification (NERC) Based on Multinomial Naïve Bayes (MNB) Classifier. *International Journal on Natural Language Computing* 4(4): 39–52. <http://www.airccse.org/journal/ijnlc/papers/4415ijnlc04.pdf>.
- Saha, Sujan Kumar, Sudeshna Sarkar, and Pabitra Mitra. 2009. Feature Selection Techniques for Maximum Entropy Based Biomedical Named Entity Recognition. *Journal of Biomedical Informatics* 42(5): 905–11. <http://dx.doi.org/10.1016/j.jbi.2008.12.012>.
- Sebastiani, Fabrizio. 2001. Machine Learning in Automated Text Categorization. *Journal ACM Computing Surveys (CSUR)* 34(1): 1–47. <http://arxiv.org/abs/cs/0110053>.
- Suakkaphong, Nichalin, Zhu Zhang, and Hsinchun Chen. 2009. Disease Named Entity Recognition Using Semisupervised Learning and Conditional Random Fields. *Journal of The American Society for Information Science and Technology* 3(2): 80–90.
- Wu, Tianhao, William M Pottenger, and Computer Science. 2005. A Semi-Supervised Active Learning Algorithm for Information Extraction from Textual Data. *Journal of the American Society for Information Science and Technology* 56(3): 258–71. <http://doi.wiley.com/10.1002/asi.20119>.

# Petunjuk Penulisan

## JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

Naskah dikirim dalam bentuk *softcopy* ke alamat email [uppm@stis.ac.id](mailto:uppm@stis.ac.id) disertai dengan daftar riwayat hidup ringkas penulis. Format naskah mengacu pada Petunjuk Penulisan Naskah berikut:

Naskah dibuat menggunakan *Microsoft Office Word* 2010. Seluruh bagian dalam naskah diketik dengan huruf *Times New Roman*, ukuran 12, spasi 1, ukuran kertas A4 dan margin 2 cm untuk semua sisi kecuali margin kiri 3 cm, jumlah halaman 15-20. Untuk kepentingan penyuntingan naskah, seluruh bagian naskah (termasuk tabel, gambar dan persamaan matematika) dibuat dalam format yang dapat disunting oleh editor.

Gaya penulisan naskah untuk Jurnal Aplikasi Statistika dan Komputasi Statistik ditulis dalam *Bahasa Indonesia* dengan gaya naratif. Pembabakan dibuat sederhana dan sedapat mungkin menghindari pembabakan bertingkat. Tabel dan gambar harus mencantumkan sumber jika dari data sekunder. Tabel, gambar dan persamaan matematika diberi nomor secara berurut sesuai dengan kemunculannya. Semua kutipan dan referensi dalam naskah harus tercantum dalam daftar pustaka, dan sebaliknya sumber bacaan yang tercantum dalam daftar pustaka harus ada dalam naskah. Format sumber: Nama Penulis dan Tahun. Nomor dan judul tabel diletakkan di bagian atas tabel dan dicetak tebal, sedangkan nomor dan judul gambar diletakkan di bagian bawah gambar dan dicetak tebal.

### Bagian naskah berisi:

**Judul.** Judul tidak melebihi 12 kata dalam Bahasa Indonesia.

**Data Penulis.** Berisi nama lengkap semua penulis tanpa gelar, asal institusi, dan alamat email.

**Abstrak.** Ditulis dalam Bahasa Inggris dan Bahasa Indonesia, maksimum 100 kata untuk masing-masing abstrak dan berisikan tiga hal yaitu topik yang dibahas, metodologi yang dipergunakan dan hasil yang didapatkan.

**Kata Kunci.** Berisi kata atau frasa (maksimum 5 subjek) yang sering dipergunakan dalam naskah dan dianggap mewakili dan atau terkait dengan topik yang dibahas.

**Pendahuluan.** Memuat latar belakang, studi sebelumnya yang relevan, permasalahan ataupun hipotesis yang akan diuji dalam penelitian, ruang lingkup penelitian, serta tujuan dari penelitian.

**Metodologi** terdiri atas:

**a. Tinjauan Referensi.** Bagian ini menguraikan landasan konseptual dari tulisan dan berisi alasan teoritis mengapa pertanyaan penelitian dalam artikel diajukan. Di samping itu penulis dapat mengutip studi yang relevan sebelumnya untuk melengkapi justifikasi mengenai kerangka pikir penelitian.

**b. Metode Analisis.** Bagian ini berisi informasi teoritis dan teknis yang cukup memadai untuk pembaca dapat mereproduksi penelitian dengan baik termasuk di dalamnya uraian mengenai jenis dan sumber data serta variabel yang digunakan. Dalam hal keperluan verifikasi hasil, editor dan mitra bestari (*reviewer*) berhak meminta data mentah (*raw data*) yang digunakan penulis.

**Hasil dan Pembahasan.** Tuliskan hasil yang didapat berdasarkan metode yang digunakan disertai analisis terhadap variabel-variabelnya. Dapat disajikan berupa tabel, gambar, hasil pengujian hipotesis dengan disertai uraian analitis yang mengangkat poin-poin penting berdasarkan konsepsi teoritisnya.

**Kesimpulan dan Saran.** Bagian ini memuat kesimpulan dari hasil dan implikasinya secara akademis, dan saran yang dapat diberikan berdasarkan temuan dari pembahasan. Bagian ini juga memuat keterbatasan penelitian dan kemungkinan penelitian lanjutan yang dapat dilakukan dengan penggunaan/pengembangan variabel, metode analisis ataupun cakupan wilayah penelitian lainnya.

**Daftar Pustaka.** Daftar pustaka disusun berdasarkan urutan abjad dengan ketentuan sebagai berikut:

Publikasi Buku

1. Penulis satu orang  
Enders, Walter. 2010. *Applied Econometric Time Series, Third Edition*. New Jersey: Wiley.
2. Penulis dua orang  
Pyndick, Robert. S. dan Rubinfeld, Daniel L. 2009. *Microeconomics, Seventh Edition*. New Jersey: Pearson Education.
3. Penulis tiga orang  
Fotheringham, A. S., Brunson, C, dan Charlton, M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex: John Wiley & Sons.

Artikel dalam jurnal

Romer, P. 1993. Idea Gaps and Object Gaps in Economic Development. *Journal of Monetary Economics*, Vol. 32 (3), 543–573.

Artikel online

Woodward, Douglas P. 1992. Locational Determinants of Japanese Manufacturing Start-Ups in the United States. *Southern Economic Journal*, Vol. 58 (3), 690-708. <http://www.jstor.org/discover/10.2307/1059836> (Diakses 1 September, 2014).

Buku yang ditulis oleh lembaga atau organisasi

BPS. 2009. *Analisis dan Penghitungan Tingkat Kemiskinan 2008*. Jakarta: BPS.

Kertas kerja (working papers)

Edwards, S. 1990. Capital Flows, Foreign Direct Investment, and Debt-Equity Swaps in Developing Countries. *NBER Working Paper*, 3497.

Makalah yang direpresentasikan

Zhang, Kevin H. 2006. Foreign Direct Investment and Economic Growth in China: A Panel Data Study for 1992-2004. *Conference of WTO, China, and Asian Economies*. Beijing.

Karya yang tidak dipublikasikan

Hartono, Djoni. 2002. Analisis Dampak Kebijakan Harga Energi terhadap Perekonomian dan Distribusi Pendapatan di DKI Jakarta: Aplikasi Model Komputasi Keseimbangan Umum (Computable General Equilibrium Model). *Tesis*. Jakarta.

Artikel di koran, majalah, dan periodik sejenis

Reuters. (2014, September 17). Where is Inflation?. *Newsweek*.