

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 11, NOMOR 2, DESEMBER 2019 ISSN 2086 – 4132
AKREDITASI NOMOR: 747/Akred/P2MI-LIPI/04/2016

Pengaruh Perkembangan Finansial terhadap Konvergensi Ekonomi di ASEAN: Pendekatan Data Panel Dinamis Tahun 2001-2016

AISYAH FITRI YUNIASIH, KRISMANTI TRI WAHYUNI, dan IRMA KORIMAH

Determinasi Partisipasi Sekolah Anak Penyandang Disabilitas di Indonesia Tahun 2015

DILLA CITRA DEWI dan EKARIA SOEBIJARTO

Hierarchical Bayes Spasial untuk Estimasi Pengeluaran Perkapita Level Kecamatan

DWI ASIH SEPTI WAHYUNI

Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode *Naive Bayes*

ALIFTA AINURROCHMAH, MEMI NOR HAYATI, dan ANDI M. ADE SATRIYA

Peramalahan *Time Series* Menggunakan Gaussian Kernel PCA dan Autoregressive

KASIFUL APRIYANTO

NoSQL Model *Data Warehouse* Metadata Survei Dinamis (Studi Kasus: Survei Rumah Tangga)

LUTFI RAHMATUTI MAGHFIROH dan IBNU SANTOSO



PUSAT PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
POLITEKNIK STATISTIKA 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: Dr. Erni Tri Astuti

Ketua Dewan Redaksi: Dr. Nasrudin

Koordinator Jurnal Ilmiah: Dr. Ernawati Pasaribu

Mitra Bestari: Dr. Budiasih
Dr. Ahmadriswan Nasution
Dr. Timbang Sirait
Dr. Azka Ubaidillah
Novia Budi Parwanto, Ph.D.
Yunarso Anang, Ph.D.

Pelaksana Redaksi: Dr. Ernawati Pasaribu
Geri Yesa Ermawan, S.Tr.Stat.

Alamat Redaksi:
Politeknik Statistika STIS
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 kehadiran Allah, Tuhan Yang Maha Esa, “Jurnal Aplikasi Statistika dan Komputasi Statistik” Volume 11, Nomor 2, Desember 2019 dapat diterbitkan. Jurnal ilmiah ini dapat terwujud atas partisipasi semua pihak, penulis internal dilingkungan Politeknik Statistika STIS maupun penulis eksternal, 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 agar publikasi yang dihasilkan menjadi salah satu sarana untuk memberikan sosialisasi statistika bagi masyarakat.

Jakarta, Desember 2019
Ketua Dewan Redaksi,

Dr. Nasrudin

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

VOLUME 11, NOMOR 2, DESEMBER 2019

AKREDITASI NOMOR: 747/Akred/P2MI-LIPI/04/2016

DAFTAR ISI

<u>Pengantar Redaksi</u>	iii
<u>Daftar Isi</u>	iv
<u>Abstrak</u>	v-xii
Pengaruh Perkembangan Finansial terhadap Konvergensi Ekonomi di ASEAN: Pendekatan Data Panel Dinamis Tahun 2001-2016 <u>Aisyah Fitri Yuniasih, Krismanti Tri Wahyuni, dan Irma Korimah</u>	1-14
Determinasi Partisipasi Sekolah Anak Penyandang Disabilitas di Indonesia Tahun 2015 <u>Dilla Citra Dewi dan Ekaria Soebijarto</u>	15-28
<i>Hierarchical Bayes</i> Spasial untuk Estimasi Pengeluaran Perkapita Level Kecamatan <u>Dwi Asih Septi Wahyuni</u>	29-36
Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode <i>Naive Bayes</i> <u>Alifta Ainurrochmah, Memi Nor Hayati, dan Andi M. Ade Satriya</u>	37-48
Peramalan <i>Time Series</i> Menggunakan Gaussian Kernel PCA dan Autoregressive <u>Kasiful Apriyanto</u>	49-54
NoSQL Model <i>Data Warehouse</i> Metadata Survei Dinamis (Studi Kasus: Survei Rumah Tangga) <u>Lutfi Rahmatuti Maghfiroh dan Ibnu Santoso</u>	55-74

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

DDC: 315.98

Aisyah Fitri Yuniasih, Krismanti Tri Wahyuni, dan Irma Korimah

Pengaruh Perkembangan Finansial Terhadap Konvergensi Ekonomi di ASEAN: Pendekatan Data Panel Dinamis Tahun 2001-2016

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 1 – 8

Abstrak

ASEAN merupakan salah satu perhimpunan negara-negara berpotensi ekonomi tinggi di dunia yang masih mengalami permasalahan disparitas perekonomian. Penelitian ini bertujuan untuk menggambarkan perkembangan finansial negara-negara ASEAN, mengidentifikasi adanya konvergensi regional pendapatan per kapita negara-negara ASEAN, dan pengaruh perkembangan finansial serta faktor-faktor yang memengaruhi pendapatan per kapita negara-negara ASEAN. Oleh karena itu, beberapa variabel dari sektor finansial ditambahkan sebagai variabel kontrol ke dalam model pertumbuhan ekonomi augmented Solow yang menjadi dasar penelitian ini. Hasil dari model teoritis yang diperoleh dari penelitian ini menunjukkan bahwa perkembangan finansial meningkatkan proses konvergensi. Model panel dinamis FD-GMM (*First Different Generalized Method of Moment*) yang digunakan dalam penelitian ini menunjukkan bahwa perkembangan finansial dapat mempercepat proses konvergensi perekonomian di antara negara-negara ASEAN karena memberikan pengaruh positif terhadap pertumbuhan ekonomi.

Kata kunci: konvergensi, perkembangan finansial, panel dinamis, FD-GMM

DDC: 315.98

Dilla Citra Dewi dan Ekaria Soebijarto

Determinan Partisipasi Sekolah Anak Penyandang Disabilitas di Indonesia Tahun 2015

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 9 – 26

Abstrak

Partisipasi bersekolah anak penyandang disabilitas di Indonesia masih rendah serta masih banyak anak disabilitas yang belum mendapat pendidikan yang memadai. Penelitian ini bertujuan untuk mengetahui variabel-variabel pada level individu dan pada level provinsi yang memengaruhi partisipasi sekolah anak penyandang disabilitas di Indonesia berdasarkan data Survei Penduduk Antar Sensus (SUPAS 2015). Metode analisis yang digunakan dalam penelitian ini adalah multilevel regresi logistik biner dua level. Hasil yang didapatkan antara lain variabel-variabel pada level individu yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas yaitu tingkat kesulitan disabilitas anak, jenis disabilitas anak, tingkat pendidikan Kepala Rumah Tangga (KRT), serta daerah tempat tinggal. Kemudian variabel-variabel pada level provinsi yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas adalah rasio ketersediaan Sekolah Luar Biasa (SLB) dan jarak ketersediaan SLB.

Kata kunci: disabilitas, anak penyandang disabilitas, pendidikan, partisipasi sekolah, multilevel regresi logistik

DDC: 315.98

Dwi Asih Septi Wahyuni

Hierarchical Bayes Spasial untuk Estimasi Pengeluaran Perkapita Level Kecamatan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 27 – 38

Abstrak

Badan Pusat Statistik merilis angka pengeluaran perkapita setiap tahun untuk 514 kabupaten/kota di Indonesia. Angka pengeluaran perkapita diperoleh melalui Survei Sosial Ekonomi Nasional (Susenas). Susenas hanya dapat memberikan data yang akurat hingga level kabupaten/kota. Disisi lain, kebutuhan data hingga level kecamatan bahkan desa/kelurahan semakin meningkat guna memotret keberhasilan tujuan pembangunan berkelanjutan (*sustainable development goals*) setiap kabupaten/kota. Penelitian ini mengusulkan metode *small area estimation* dengan menggabungkan data survei dan sensus menggunakan model *hierarchical bayes* dengan mempertimbangkan keterkaitan antar wilayah (bobot spasial). Estimasi pengeluaran perkapita level kecamatan mampu memberikan hasil yang akurat dengan menggabungkan data pengeluaran perkapita sebagai variabel dependen dari Susenas dan data sensus dari Potensi Desa (Podes) sebagai variabel independen menggunakan model *hierarchical bayes* spasial.

Kata kunci: *small area estimation*, *hierarchical bayes*, spasial, pengeluaran perkapita

DDC: 315.98

Aliftha Ainurrochmah, Memi Nor Hayati,
dan Andi M. Ade Satriya

Perbandingan Klasifikasi Analisis
Diskriminan Fisher dan Metode *Naive*
Bayes

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 11, Nomor 2, Desember
2019, hal 39 – 52

Abstrak

Klasifikasi adalah suatu teknik pembentukan model dari data yang telah diketahui kelompok klasifikasinya. Model tersebut digunakan untuk mengklasifikasikan obyek baru. Analisis diskriminan Fisher merupakan teknik multivariat untuk memisahkan obyek-obyek dalam kelompok yang berbeda. *Naive* Bayes merupakan pengklasifikasian berdasarkan probabilitas dan teorema Bayes dengan asumsi independensi yang kuat. Penelitian ini bertujuan melakukan perbandingan tingkat akurasi klasifikasi dengan analisis diskriminan Fisher dan metode *naive* Bayes pada status pembayaran premi nasabah asuransi. Data yang digunakan memiliki 4 variabel bebas yaitu pendapatan, usia, masa pembayaran premi dan besar pembayaran premi. Hasil kesalahan akurasi dengan menggunakan nilai APER (*Apparent Rate Error*) menunjukkan bahwa metode *naive* Bayes memiliki tingkat akurasi yang lebih tinggi sebesar 15,38% daripada analisis diskriminan Fisher sebesar 46,15% dalam menganalisis status pembayaran premi nasabah asuransi.

Kata kunci: klasifikasi, analisis diskriminan Fisher, *naive* Bayes, asuransi

DDC: 315.98

Kasiful Apriyanto

Peramalan *Time Series* Menggunakan
Gaussian Kernel PCA dan Autoregressive

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 11, Nomor 2, Desember
2019, hal 53 – 66

Abstrak

Peramalan *time series* menjadi bagian penting dalam pengambilan keputusan karena mampu memberikan gambaran atau kejadian mendatang berupa data prediksi. Paper ini menawarkan gaussian kernel PCA dan Autoregressive (KPCA-AR) sebagai metode peramalan untuk melakukan prediksi data. Metode KPCA-AR bekerja dengan melakukan pemetaan data ke dimensi yang lebih tinggi menggunakan kernel dengan distribusi gaussian. Setelah itu dilakukan transformasi data dengan PCA agar dimensi yang dihasilkan dapat direduksi dengan varian maksimum sehingga tidak mengurangi karakteristik data secara signifikan. Data inilah yang kemudian digunakan untuk melakukan peramalan menggunakan autoregressive. Paper ini juga membandingkan beberapa metode peramalan lainnya seperti ARIMA, ANN, SVM, dan Eksponensial *Smoothing*. Hasil menunjukkan bahwa KPCA-AR secara umum mampu memberikan prediksi yang baik dan bisa digunakan sebagai alternative dari metode perhitungan yang ada dilihat dari kelebihan ataupun kekurangannya.

Kata kunci: Gaussian, Kernel, PCA, Autoregressive, *Forecasting*

DDC: 315.98

Lutfi Rahmatuti Maghfiroh dan
Ibnu Santoso

NoSQL Model *Data Warehouse* Metadata
Survei Dinamis (Studi Kasus: Survei
Rumah Tangga)

Jurnal Aplikasi Statistika & Komputasi
Statistik, Volume 11, Nomor 2, Desember
2019, hal 67 – 78

Abstrak

Suatu survei yang dilakukan antar periode cenderung mengalami perubahan metadata. Meski demikian, seluruh data dan metadata tersebut haruslah disimpan secara terintegrasi untuk keperluan data retrieval yang efisien. Tantangan yang muncul dengan perubahan metadata ini adalah ketika dilakukan suatu *query* untuk data lebih dari satu periode. Satu *query* untuk data satu periode tentu tidak efisien. Di sisi lain, satu *query* untuk data beberapa periode dapat menyebabkan terjadinya masalah konsistensi data karena adanya perubahan metadata. Pada penelitian ini, kami melakukan studi lebih lanjut terhadap beberapa model metadata data warehouse yang dapat mengakomodasi perubahan metadata secara dinamis seperti *data warehouse* tradisional, *Multi Version Data Warehouse* (MVDW) dan Manajemen metadata menggunakan model non-relasional seperti NoSQL. Dari hasil penelitian sebelumnya didapatkan bahwa model non-relasional pada *database* NoSQL menghasilkan waktu *query* yang lebih cepat dan penyimpanan yang lebih efisien dibandingkan model relasional, namun model tersebut belum dapat menangani model survei rumah tangga yang berbeda-beda. Penelitian ini bertujuan untuk menghasilkan model *data warehouse* yang dapat mengelola perubahan metadata pada survei rumah tangga yang berbeda-beda secara dinamis. Dari hasil pengujian didapatkan hasil bahwa model yang diusulkan telah dapat menangani perubahan metadata dari model survei rumah tangga yang berbeda-beda dengan waktu dan

tempat penyimpanan yang tidak jauh berbeda dari hasil penelitian sebelumnya.

Kata kunci: Metadata, *Data warehouse*, MVDW, NoSQL

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

DDC: 315.98

Aisyah Fitri Yuniasih, Krismanti Tri Wahyuni, dan Irma Korimah

Pengaruh Perkembangan Finansial Terhadap Konvergensi Ekonomi di ASEAN: Pendekatan Data Panel Dinamis Tahun 2001-2016

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 1 – 8

Abstract

ASEAN is one of the associations of high-economic potentialized countries in the world which still has problems in economic disparity. This research is aimed to describe the financial development of ASEAN countries, identify the regional convergence of income per capita of ASEAN countries, and the influence of financial development and the factors that influence the income per capita of ASEAN countries. As a result, several variables of financial sector are added as control variables to a model of economic growth, augmented Solow, which is the basis of this research. The result of the theoretical model acquired from this research shows that the financial development increases the convergence process. The dynamic panel model FD-GMM (First Different Generalized Method of Moment) used in this research shows that the financial development could accelerate the process of economic convergence among ASEAN countries as it gives positive impacts towards the economic growth.

Keywords: convergence, financial development, dynamic panel, FD-GMM

DDC: 315.98

Dilla Citra Dewi dan Ekaria Soebijarto

Determinan Partisipasi Sekolah Anak Penyandang Disabilitas di Indonesia Tahun 2015

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 9 – 26

Abstract

Percentage of children with disabilities in Indonesia who have not received education yet (no or never attended school) based on Susenas 2012 is 35,25 % and those who do not attend school anymore are 16,03 %, while those who still study in school are only 48,73 % (BPS and Kemensos, 2013). This shows that Children with disabilities in Indonesia are still lacking in school participation and many children have not received adequate education. This research aims to determine the individual and contextual variables that affect the school participation of children with disabilities in Indonesia based on SUPAS 2015. Analytical method used in this research is multilevel binary logistic regression with two levels. The obtained result is variables at individual level that significantly affected school participation of children with disabilities are severity of disability, the type of disability, the education level of the household head, and the area of residence. Variables at provincial level that significantly affected the school participation of children with disabilities are ratio of school's availability and distance of school's availability.

Keywords: disability, children with disabilities, education, school participation, multilevel logistic regression

DDC: 315.98

Dwi Asih Septi Wahyuni

Hierarchical Bayes Spasial untuk Estimasi Pengeluaran Perkapita Level Kecamatan

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 27 – 38

Abstract

The Central Bureau of Statistics releases annual per capita expenditure figures for 514 districts / cities in Indonesia. The per capita expenditure figures are obtained through the National Socio-Economic Survey (Susenas). Susenas can only provide accurate data down to the district / city level. On the other hand, the need for data up to the sub-district level and even villages / wards is increasing in order to capture the success of the sustainable development goals of each district / city. This study proposes a small area estimation method by combining survey and census data using a hierarchical Bayes model by considering the relationship between regions (spatial weight). The estimation of expenditure per capita at the sub-district level is able to provide accurate results by combining the per capita expenditure data as the dependent variable from Susenas and census data from the Village Potential Survey (Podes) as an independent variable using a bayesspasial hierarchical model.

Keywords: small area estimation, hierarchical bayes, spatial, expenditure per capita

DDC: 315.98

Alifta Ainurrochmah, Memi Nor Hayati, dan Andi M. Ade Satriya

Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode *Naive Bayes*

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 39 – 52

Abstract

Classification is a technique to form a model of data that is already known to its classification group. The model was formed will be used to classify new objects. Fisher discriminant analysis is multivariate technique to separate objects in different groups. Naive Bayes is a classification technique based on probability and Bayes theorem with assumption of independence. This research has a goal to compare the level of classification accuracy between Fisher's discriminant analysis and Naive Bayes method on the insurance premium payment status customer. The data used four independent variables that is income, age, premium payment period and premium payment amount. The results of misclassification using the APER (Apparent Rate Error) indicate that the naive Bayes method has a higher level of accuracy is 15,38% than Fisher's discriminant analysis is 46,15% on the insurance premium payment status customer.

Keywords: classification, Fisher discriminant analysis, naive Bayes, insurance

DDC: 315.98

Kasiful Apriyanto

Peramalan *Time Series* Menggunakan Gaussian Kernel PCA dan Autoregressive

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 53 - 66

Abstract

Time series forecasting becomes an important part in decision making because it is able to provide a picture or future event in the form of predictive data. This paper offers the Gaussian kernel PCA and Autoregressive (KPCA-AR) as a forecasting method to predict data. The KPCA-AR method works by mapping data to a higher dimension using a kernel with a gaussian distribution. After that the transformation of data with PCA is done so that the resulting dimensions can be reduced with maximum variants so as not to significantly reduce the characteristics of the data. This data is then used to forecast using autoregressive. This paper also compares several other forecasting methods such as ARIMA, ANN, SVM, and Exponential Smoothing as a comparison. The results show that KPCA-AR in general is able to give good predictions and can be used as an alternative to existing calculation methods seen from the advantages or disadvantages.

Keywords: Gaussian, Kernel, PCA, Autoregressive, Forecasting

DDC: 315.98

Lutfi Rahmatuti Maghfiroh dan Ibnu Santoso

NoSQL Model *Data Warehouse* Metadata Survei Dinamis (Studi Kasus: Survei Rumah Tangga)

Jurnal Aplikasi Statistika & Komputasi Statistik, Volume 11, Nomor 2, Desember 2019, hal 67 - 78

Abstract

A survey conducted between periods tends to experience changes in metadata. However, all data and metadata must be stored in an integrated manner for data retrieval efficiency. There are difficulty in query performed for data in between periods because of metadata changes. One query for one period data is certainly not efficient. On the other hand, one query for data for several periods can cause data consistency problems due to changes in metadata. In the latest study, we conduct further studies on several data warehouse metadata models that can accommodate dynamic metadata changes such as traditional data warehouse, Multi Version Data Warehouse (MVDW) and metadata management using non-relational models such as NoSQL. From the results of previous studies it was found that the non-relational model in the NoSQL database resulted in faster query time and more efficient storage than the relational model, but the model had not been able to handle different household survey models. This study aims to produce a data warehouse model that can manage changes in metadata in household surveys that vary dynamically. The results of the test show that the proposed model has been able to handle changes in metadata from household survey models that vary with time and place of storage that are not much different from the results of previous studies.

Keywords: Metadata, Data warehouse, MVDW, NoSQL

PENGARUH PERKEMBANGAN FINANSIAL TERHADAP KONVERGENSI EKONOMI DI ASEAN: PENDEKATAN DATA PANEL DINAMIS TAHUN 2001-2016

Aisyah Fitri Yuniasih¹, Krismanti Tri Wahyuni², Irma Korimah³

Politeknik Statistika STIS
e-mail: ¹aisyah.fy@bps.go.id

Abstrak

ASEAN merupakan salah satu perhimpunan negara-negara berpotensi ekonomi tinggi di dunia yang masih mengalami permasalahan disparitas perekonomian. Penelitian ini bertujuan untuk menggambarkan perkembangan finansial negara-negara ASEAN, mengidentifikasi adanya konvergensi regional pendapatan per kapita negara-negara ASEAN, dan pengaruh perkembangan finansial serta faktor-faktor yang memengaruhi pendapatan per kapita negara-negara ASEAN. Oleh karena itu, beberapa variabel dari sektor finansial ditambahkan sebagai variabel kontrol ke dalam model pertumbuhan ekonomi *augmented Solow* yang menjadi dasar penelitian ini. Hasil dari model teoritis yang diperoleh dari penelitian ini menunjukkan bahwa perkembangan finansial meningkatkan proses konvergensi. Model panel dinamis FD-GMM (*First Different Generalized Method of Moment*) yang digunakan dalam penelitian ini menunjukkan bahwa perkembangan finansial dapat mempercepat proses konvergensi perekonomian di antara negara-negara ASEAN karena memberikan pengaruh positif terhadap pertumbuhan ekonomi.

Kata kunci: konvergensi, perkembangan finansial, panel dinamis, FD-GMM

Abstract

ASEAN is one of the associations of high-economic potentialized countries in the world which still has problems in economic disparity. This research is aimed to describe the financial development of ASEAN countries, identify the regional convergence of income per capita of ASEAN countries, and the influence of financial development and the factors that influence the income per capita of ASEAN countries. As a result, several variables of financial sector are added as control variables to a model of economic growth, augmented Solow, which is the basis of this research. The result of the theoretical model acquired from this research shows that the financial development increases the convergence process. The dynamic panel model FD-GMM (First Different Generalized Method of Moment) used in this research shows that the financial development could accelerate the process of economic convergence among ASEAN countries as it gives positive impacts towards the economic growth.

Keywords: convergence, financial development, dynamic panel, FD-GMM

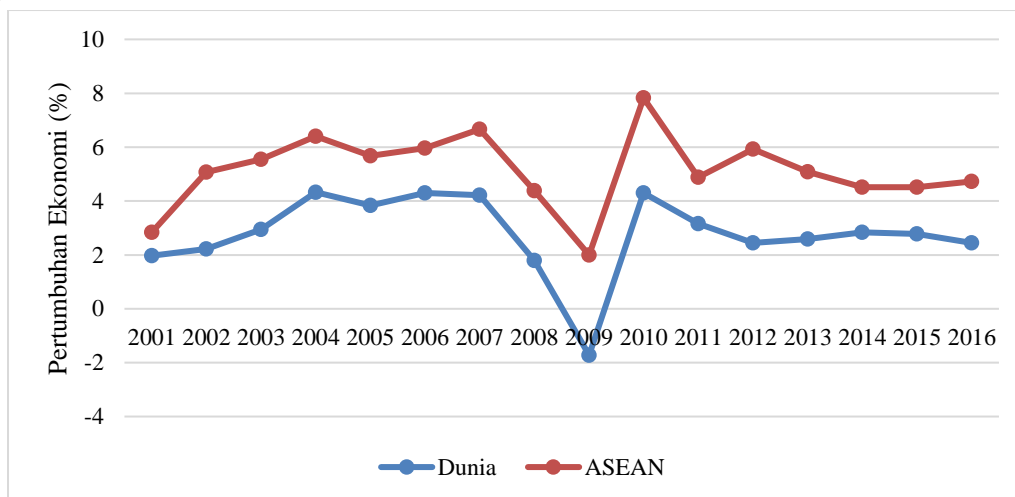
PENDAHULUAN

1. Latar Belakang

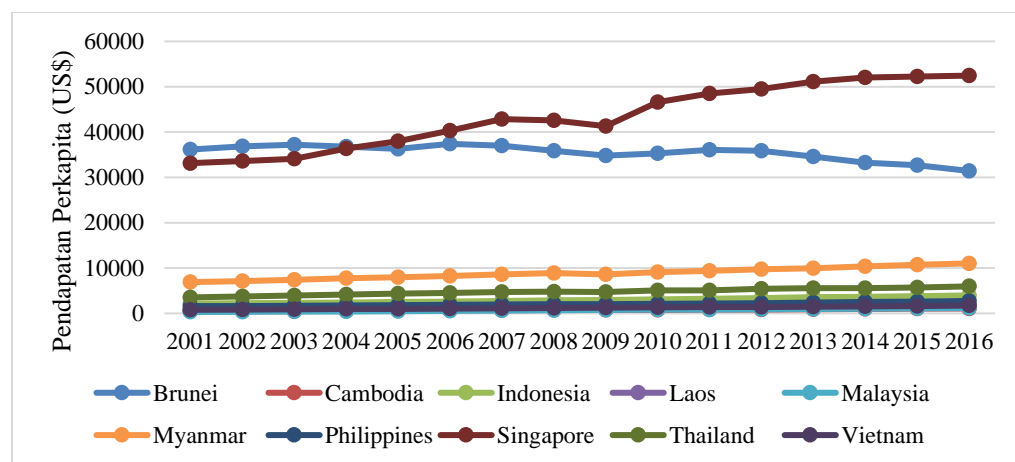
Association of Southeast Asian Nations atau biasa disingkat ASEAN merupakan perhimpunan bangsa-bangsa yang secara teritorial berada di kawasan Asia Tenggara. Sampai akhir tahun 2018, terdapat 10 negara di kawasan Asia Tenggara yang menjadi anggota ASEAN (ASEAN, 2018). Dengan jumlah penduduk 640.507.126 jiwa atau hanya 8,58 % dari total penduduk dunia pada tahun 2016 berdasarkan data UNCTAD, pertumbuhan ekonomi negara-negara ASEAN selalu berada di atas pertumbuhan ekonomi dunia (**Gambar 1**). Hal ini menjadikan ASEAN sebagai salah satu kawasan berpotensi ekonomi tinggi di dunia.

Masyarakat Ekonomi ASEAN (MEA) mulai diberlakukan secara resmi

pada tahun 2015 dengan tujuan menjadikan kawasan Asia Tenggara sebagai kawasan yang stabil, makmur, dan berdaya saing tinggi dengan aliran bebas barang, jasa, investasi, modal, dan tenaga kerja terampil serta pembangunan ekonomi yang merata disertai pengurangan kemiskinan dan kesenjangan sosial ekonomi di tiap-tiap negara yang menjadi anggotanya (Das, 2017). Namun, pemerataan ekonomi bagi seluruh penduduk kawasan ASEAN sepertinya belum bisa terwujud. Pada tahun 2016, data UNCTAD menunjukkan bahwa pendapatan perkapita masyarakat negara Singapura sebesar US\$ 52.458 merupakan yang tertinggi di antara negara ASEAN, setara dengan 48 kali lipat pendapatan perkapita masyarakat negara Kamboja yang terendah di antara negara ASEAN yang hanya mencapai US\$ 1.078 (**Gambar 2**).



Gambar 2. Pertumbuhan Ekonomi Dunia dan ASEAN Tahun 2001-2016



Gambar 1. Pendapatan Perkapita Negara-Negara ASEAN Tahun 2001-2016

Faktor yang memengaruhi perbedaan pendapatan per kapita antar negara dari sisi *supply* yang berkaitan dengan pertumbuhan nilai tambah diantaranya adalah peningkatan volume produksi, peningkatan akumulasi modal, dan peningkatan produktivitas faktor produksi (Mhadhbi, 2014). Perkembangan finansial diduga dapat berkontribusi terhadap peningkatan pertumbuhan ekonomi melalui peningkatan volume produksi, akumulasi modal, dan produktivitas faktor produksi (Lu, *et al*, 2007). Integrasi finansial merupakan salah satu elemen utama terkait implementasi MEA dalam pergerakan modal (Almekinders, *et al*, 2015).

Berdasarkan latar belakang dan permasalahan yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk menggambarkan perkembangan finansial negara-negara ASEAN, mengidentifikasi adanya konvergensi regional pendapatan per kapita negara-negara ASEAN, dan pengaruh perkembangan finansial serta faktor-faktor yang memengaruhi pendapatan per kapita negara-negara ASEAN. Penurunan ketimpangan antar negara ASEAN diperlukan dalam mewujudkan pembangunan ekonomi yang merata sesuai tujuan utama implementasi MEA. Akselerasi proses konvergensi regional pendapatan per kapita negara ASEAN dapat dilakukan dengan memberikan porsi intervensi terhadap determinan pendapatan per kapita yang lebih besar bagi negara dengan pendapatan per kapita yang lebih rendah.

2. Tinjauan Pustaka

Beberapa penelitian yang sudah membahas tentang konvergensi perekonomian di negara ASEAN menunjukkan hasil yang berbeda-beda. Kebanyakan penelitian menganalisis tentang konvergensi perekonomian di negara ASEAN 5 yang terdiri dari negara pendiri ASEAN yaitu Indonesia, Malaysia, Filipina, Thailand, dan Singapura. Konvergensi sigma dan konvergensi beta dengan metode analisis data panel tidak terbukti terjadi di antara negara ASEAN 5 pada periode 1960-1999 (Michelis and

Neaime, 2004). Namun, Ismail (2008) menemukan bahwa konvergensi tidak bersyarat dan konvergensi bersyarat terjadi diantara negara ASEAN 5 pada periode 1960-2004 dengan kecepatan konvergensi 1.6%-16.6% dengan pendekatan data panel dinamis yang heterogen Pooled Mean Group Estimator (PMGE). Jayanthakumaran dan Lee (2013) juga menemukan bukti terjadinya konvergensi stokastik di antara negara ASEAN 5 dengan metode analisis deret waktu yang menerapkan uji akar unit dengan dua buah *structural break* endogen.

MEA tentunya mengharapkan konvergensi perekonomian terjadi tidak hanya di antara negara ASEAN 5 saja. Chowdhary, *et al* (2011) melakukan uji konvergensi di antara negara ASEAN-9 (ASEAN tanpa Myanmar) dan menemukan bahwa konvergensi sigma terjadi dengan stabil walaupun perlahan pada periode tahun 1990-2008. Guggler dan Vonoli (2016) meneliti konvergensi perekonomian 9 negara ASEAN (tanpa Myanmar) periode 2000-2014 dengan metode Pooled OLS. Hasilnya kecepatan konvergensi antar negara ASEAN relatif rendah jika dibandingkan dengan kawasan lain seperti Uni Eropa. Devasmita (2018) juga menemukan bahwa terjadi konvergensi sigma dan konvergensi beta di antara 10 negara ASEAN selama periode tahun 2000-2014. Sayangnya, belum ada penelitian yang memasukkan perkembangan finansial dalam menganalisis konvergensi perekonomian di negara ASEAN.

Teori pertumbuhan ekonomi tradisional Neoklasik mengabaikan peran dari perkembangan finansial (Mhadhbi, 2014). Lu, *et al* (2007) mensarikan tiga teori yang mendasari hubungan antara perkembangan finansial dan pertumbuhan ekonomi. Hipotesis yang pertama, hipotesis "*supply-leading*", yang pada awalnya diperkenalkan oleh Schumpeter (1912), menyatakan bahwa sektor finansial yang baik dapat menstimulasi inovasi teknologi sehingga berperan sebagai faktor eksogen yang memengaruhi pertumbuhan ekonomi seperti yang diasumsikan oleh teori pertumbuhan ekonomi Solow. Dalam hal

ini, perkembangan finansial berperan secara proaktif meningkatkan pertumbuhan ekonomi (Fung, 2009). Kedua, hipotesis “*demand-following*” yang pada awalnya diperkenalkan oleh Robinson (1952) yang menyatakan bahwa pertumbuhan ekonomi menciptakan permintaan terhadap berbagai jenis produk-produk finansial yang harus diakomodasi oleh sistem finansial yang ada sehingga evolusi sistem keuangan merupakan proses berkelanjutan konsekuensi dari proses pembangunan ekonomi. Dalam hal ini, perkembangan finansial berperan sebagai faktor endogen yang memengaruhi pertumbuhan ekonomi seperti yang diasumsikan oleh teori pertumbuhan ekonomi Endogen. Hipotesis yang ketiga, hipotesis “*irrelevance of finance*” yang pertama kali diperkenalkan oleh Lucas (1988) memberikan sudut pandang yang berbeda. Hipotesis ini menyangkal eksistensi hubungan jangka panjang antara perkembangan finansial dan pertumbuhan ekonomi karena faktor finansial hanya bisa signifikan dalam perekonomian yang kuat sektor finansialnya.

Divergensi ekonomi antara negara miskin dan negara kaya pada dasarnya dipengaruhi oleh perbedaan peningkatan volume produksi, peningkatan akumulasi modal, dan peningkatan produktivitas faktor produksi yang secara umum dipengaruhi oleh perbedaan kecepatan transfer teknologi. Keterbatasan finansial menghalangi negara miskin untuk mengambil keuntungan penuh dari proses transfer teknologi yang ada (Aghion, *et al*, 2004). Proses transfer teknologi memerlukan biaya karena negara penerima transfer teknologi biasanya tidak dapat secara langsung menerapkan teknologi dari luar tanpa melakukan penyesuaian. Padahal, seiring kemajuan teknologi global, peningkatan porsi investasi juga diperlukan untuk mengiringinya. Sayangnya, semakin rendah tingkat perkembangan finansial di suatu negara menyebabkan semakin rendah pula multiplier dari kredit yang bisa dinikmati oleh negara tersebut yang diperlukan untuk mengakselerasi transfer teknologi. Peluang sebuah negara akan

konvergen dengan negara lain merupakan fungsi naik terhadap tingkat perkembangan finansialnya.

Secara umum, perkembangan finansial dapat memengaruhi performa perekonomian secara keseluruhan melalui 2 jalur (Mhadhbi, 2014). Pertama, *volume effect* dimana kegiatan finansial meningkatkan tabungan yang bisa menjadi sumber untuk melakukan investasi. Dalam hal ini, perkembangan finansial dapat meningkatkan akumulasi modal yang dapat meningkatkan pertumbuhan ekonomi (perkembangan finansial → akumulasi modal → pertumbuhan ekonomi). Yang kedua, *allocation effect* dimana perkembangan finansial memberikan banyak pilihan alokasi sumber daya untuk investasi. Disini, dengan akumulasi modal yang tersedia, kita bisa memilih alternatif investasi finansial yang paling menguntungkan sehingga dapat meningkatkan pertumbuhan ekonomi (akumulasi modal → perkembangan finansial → pertumbuhan ekonomi).

Mhadhbi (2014) menguji secara empiris hubungan perkembangan finansial dan pertumbuhan ekonomi 110 negara maju dan negara berkembang periode 1973-2012 dengan menggunakan metode analisis data panel dinamis *Generalized Method of Moments*. Hasilnya perkembangan finansial berhubungan positif dengan pertumbuhan ekonomi.

Karshenasan dan Mobarak (2012) meneliti mengenai hubungan antara perkembangan finansial dengan konvergensi pertumbuhan ekonomi di 64 negara dengan kategori pendapatan yang berbeda pada periode 1975-1979 yang diestimasi menggunakan estimator panel System-GMM (Generalized-Method-of-Moment). Hasilnya perkembangan finansial berpengaruh positif terhadap pertumbuhan pendapatan per kapita.

METODE PENELITIAN

Penelitian ini menggunakan data panel 10 negara Anggota ASEAN yang terdiri dari Brunei Darussalam, Kamboja, Indonesia, Laos, Malaysia, Myanmar,

Filipina, Singapura, Thailand, dan Vietnam periode tahun 2001-2016. Dengan menggabungkan data *cross section* dan *time series*, penelitian ini diharapkan dapat memberikan informasi yang lebih banyak dan lebih bervariasi dalam melihat heterogenitas (efek perilaku) negara-negara di ASEAN. Perkembangan perilaku negara-negara ASEAN dianalisis dengan mempelajari *dynamic of adjustment* dari variabel endogen, yaitu pendapatan per kapita yang diproksi dari PDB per kapita.

Banyak variabel ekonomi pada dasarnya bersifat dinamis atau kondisi di tahun berjalan dipengaruhi oleh kondisi di waktu sebelumnya. Hubungan ini ditandai dengan adanya *lag* variabel endogen yang berkorelasi dengan variabel endogen (selanjutnya akan menjadi variabel eksogen). Keberadaan *lag* yang bersifat random menimbulkan masalah endogenitas karena memunculkan korelasi dengan *error*, sehingga asumsi klasik regresi tidak dapat dipenuhi (Gujarati, 2003). Penggunaan *least square* akan menghasilkan parameter yang tidak konsisten dan bias. Model data panel statis tidak dapat mengakomodir adanya *lag* variabel endogen sehingga tidak dapat digunakan dalam model yang dibangun dalam penelitian ini.

Model umum panel dinamis tanpa variabel eksogen berdasarkan Baltagi (2005) adalah:

$$y_{it} = \delta y_{i,t-1} + u_{it} \dots\dots\dots (1)$$

dimana:

y_{it} = nilai variabel endogen negara ke-i pada periode ke-t, $i = 1, 2, \dots, N$ dan $t = 1, 2, \dots, T$;

$y_{i,t-1}$ = *lag* variabel endogen;

u_{it} = $\mu_i + v_{it}$ yang memenuhi asumsi $\mu_i \sim IID(0, \sigma_\mu^2)$ dan $v_{it} \sim IID(0, \sigma_v^2)$;

δ = konstanta dari *lag* variabel endogen.

Estimasi yang digunakan pada model panel dinamis adalah *Generalized Method of Moments* (GMM). Sedangkan prosedur yang digunakan dalam penelitian ini *First*

Difference GMM (FD GMM) atau *Arellano Bond GMM* (AB GMM).

Transformasi *first difference* dilakukan untuk menghilangkan efek individu μ_i dalam rangka menyelesaikan masalah ketidakkonsistenan parameter δ dengan $N \rightarrow \infty$ dan T tertentu sebagai berikut:

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + [(\mu_i + v_{it}) - (\mu_i + v_{i,t-1})]$$

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + [v_{it} - v_{i,t-1}]$$

$$\Delta y_{it} = \Delta y_{i,t-1} + \Delta v_{it} \dots\dots\dots (2)$$

Jika diestimasi dengan *least square*, maka estimator menjadi tidak konsisten, sehingga diperlukan variabel instrumen yang merupakan variabel yang berkorelasi kuat dengan variabel endogen dan *lag* variabel endogen tetapi tidak berkorelasi dengan *error*.

Penelitian ini menginvestigasi hubungan antara perkembangan finansial dan pertumbuhan ekonomi dengan menggunakan analisis data panel dinamis GMM yang mengatasi masalah ekonometrik yang disebabkan oleh tidak diobservasinya efek spesifik *cross section* dan endogenitas dari variabel bebas dalam model yang menggunakan *lag* dari variabel dependen dalam persamaan regresi pertumbuhan ekonomi. Penggunaan data panel yang menggabungkan data deret waktu dan data *cross section* memberikan informasi tambahan tentang perubahan antar waktu dan determinannya untuk mendapatkan hasil yang lebih tepat.

Variabel endogen yang akan dimodelkan pada penelitian ini adalah pendapatan perkapita. Produk Domestik Bruto (PDB) per kapita digunakan sebagai variabel proksi dari pendapatan per kapita yang dihasilkan di suatu negara. Berdasarkan Todaro dan Smith (2006), pertumbuhan output selalu bersumber dari satu atau lebih faktor-faktor produksi yaitu penambahan modal (melalui bagian pendapatan yang ditabungkan dan diinvestasikan), kenaikan kuantitas pekerja dan peningkatan kualitas tenaga kerja

tercermin dari teknologi. Dalam perekonomian terbuka, modal mengalir dari negara kaya (rasio modal per tenaga kerja tinggi) ke negara yang miskin (rasio modal per tenaga kerja rendah) dengan imbalan tingkat investasi yang lebih tinggi. Negara maju tidak akan terus mengalami pertumbuhan ekonomi yang besar. Sementara itu modal dan tenaga kerja di negara berkembang belum berperan maksimal dalam perekonomian sehingga akan mengalami pertumbuhan yang semakin meningkat hingga mencapai *diminishing return*, sebagaimana yang terjadi pada negara kaya. Kondisi ini mengakibatkan adanya konvergensi output per kapita (pendapatan per kapita).

Konvergensi dapat diukur dengan menggunakan konvergensi sigma dan konvergensi beta. Konvergensi sigma diukur dengan menggunakan koefisien variasi logaritma natural pendapatan per kapita. Kecenderungan konvergensi sigma terjadi jika dispersi dari pendapatan perkapita menurun sepanjang waktu penelitian (Barro dan Martin, 2004). Sedangkan konvergensi beta terjadi jika perekonomian negara miskin tumbuh lebih cepat daripada ekonomi negara kaya, sehingga negara miskin bisa mengejar perekonomian daerah kaya. Konvergensi beta dibagi menjadi dua jenis yaitu konvergensi beta absolut dan konvergensi beta kondisional. Konvergensi beta absolut menggunakan pendapatan per kapita tahun sebelumnya sebagai satu-satunya variabel eksogen, sedangkan konvergensi beta kondisional menambahkan variabel eksogen lainnya dalam model, yang menunjukkan bahwa karakteristik antar negara berbeda-beda.

Model yang diajukan dalam penelitian ini mengacu pada persamaan konvergensi beta kondisional Karagiannis (2007) sebagai berikut:

$$\frac{y_{it}-y_{i,t-1}}{y_{i,t-1}} = -\alpha y_{i,t-1} + \beta X_{it} + u_{it} \dots\dots (3)$$

dimana:

- y_{it} = PDB per kapita negara ke-i periode ke-t;
- $y_{i,t-1}$ = PDB per kapita negara ke-i periode ke-(t-1);

- α = kecepatan konvergensi (*speed of convergence*);
- X_{it} = matriks variabel eksogen negara ke-i periode ke-t;
- β = vektor kolom parameter untuk X_{it} ;
- u_{it} = *error term*.

Konvergensi terjadi jika nilai α negatif, artinya terdapat hubungan terbalik antara pertumbuhan PDB per kapita dengan PDB per kapita tahun sebelumnya (Young, Higgins, dan Levy, 2007). Hal ini menyebabkan negara miskin mampu mengejar ketertinggalannya dengan negara kaya (konvergensi beta kondisional).

Estimasi parameter dengan menggunakan pendekatan panel dinamis merupakan modifikasi yang dilakukan dengan asumsi:

$$\frac{y_{it}-y_{i,t-1}}{y_{i,t-1}} \approx \ln\left(\frac{y_{it}}{y_{i,t-1}}\right) \dots\dots\dots(4)$$

Sehingga persamaan (3) dapat ditulis menjadi:

$$\ln\left(\frac{y_{it}}{y_{i,t-1}}\right) = -\alpha y_{i,t-1} + \beta X_{it} + u_{it} \dots\dots(5)$$

Berdasarkan sifat operasi logaritma natural maka:

$$\ln y_{it} - \ln y_{i,t-1} = -\alpha y_{i,t-1} + \beta X_{it} + u_{it} \dots\dots\dots(6)$$

Jika dilakukan pemindahan dari ruas kiri ke ruas kanan:

$$\ln y_{it} = (1 - \alpha)y_{i,t-1} + \beta X_{it} + u_{it} \dots\dots(7)$$

Dimisalkan $(1 - \alpha) = \beta_1$ maka:

$$\ln y_{it} = \beta_1 y_{i,t-1} + \beta X_{it} + u_{it} \dots\dots\dots(8)$$

Modifikasi yang dilakukan tersebut menyebabkan konvergensi beta kondisional terjadi jika koefisien lag variabel endogen kurang dari satu ($\beta_1 < 1$) dan signifikan. Konvergensi akan terjadi semakin cepat apabila koefisien lag variabel endogen semakin mendekati nol karena pertumbuhan PDB per kapita semakin kecil seiring berjalannya waktu. Kecepatan konvergensi dapat dihitung sebagai:

$$\text{Implied } \lambda = -\ln(\beta_1) \dots\dots\dots(9)$$

Selanjutnya Jan dan Chaudhary (2011) menjelaskan *half life of convergence* sebagai berikut:

$$e^{-\lambda t} = \frac{1}{2} \dots \dots \dots (10)$$

Dengan menggunakan operasi logaritma natural maka:

$$\lambda t = \ln 2 \dots \dots \dots (11)$$

Sehingga dapat dituliskan:

$$t = \frac{\ln 2}{\lambda} \dots \dots \dots (12)$$

Half life of convergence adalah waktu yang dibutuhkan semua negara untuk menutupi setengah kesenjangan awal.

Dalam rangka mengetahui gambaran perkembangan finansial suatu negara secara lebih menyeluruh, penelitian ini menggunakan dua buah proksi perkembangan finansial yaitu *Financial Institution Index* (FI) dan *Financial Market Index* (FM), bersumber dari *International Monetary Fund* (IMF) yang perhitungannya berdasarkan penelitian Svirydzenka (2016). Perhitungan FI dan FM dilakukan berdasarkan tiga dimensi yaitu kedalaman, akses, dan efisiensi. Penghitungan FI meliputi lembaga keuangan bank, perusahaan asuransi, reksadana, dana pensiun, dan lembaga keuangan nonbank lainnya, sedangkan penghitungan FM meliputi pasar saham dan obligasi. Kedua proksi ini diharapkan dapat menggambarkan jalur perkembangan finansial yang berbeda, namun keduanya memengaruhi pertumbuhan ekonomi dan dapat merepresentasikan perkembangan finansial yang sebenarnya di suatu negara.

Berdasarkan modifikasi model panel dinamis tersebut, spesifikasi model konvergensi perekonomian negara ASEAN yang digunakan dalam penelitian ini adalah sebagai berikut:

1) Konvergensi beta absolut

$$\ln(GDPC_{it}) = \beta_1 \ln(GDPC_{i,t-1}) + u_{it} \dots \dots \dots (13)$$

2) Konvergensi beta kondisional tanpa perkembangan finansial

$$\ln(GDPC_{it}) = \beta_1 \ln(GDPC_{i,t-1}) + \beta_2 \ln(GCF_{it}) + \beta_3 \ln(GOV_{it}) +$$

$$\beta_4 \ln(TO_{it}) + \beta_5 INF_{it} + u_{it} \dots \dots \dots (14)$$

3) Konvergensi beta kondisional dengan perkembangan finansial

$$\ln(GDPC_{it}) = \beta_1 \ln(GDPC_{i,t-1}) + \beta_2 \ln(GCF_{it}) + \beta_3 \ln(GOV_{it}) + \beta_4 \ln(TO_{it}) + \beta_5 INF_{it} + \beta_6 \ln(FI_{it}) + \beta_7 \ln(FM_{it}) + u_{it} \dots \dots (15)$$

dimana:

β_1 = kecepatan konvergensi (*speed of convergence*);

GDPC = *Gross Domestic Product Per Capita* (Pendapatan Per Kapita Riil dalam satuan US\$);

GCF = *Gross Capital Formation* (Proporsi Pembentukan Modal Bruto terhadap PDB dalam satuan %);

GOV = *Government Expenditure* (Proporsi Pengeluaran Pemerintah terhadap PDB dalam satuan %);

TO = *Trade Openness* (Proporsi Keterbukaan Perdagangan (Ekspor ditambah Impor) terhadap PDB dalam satuan %);

INF = Inflasi;

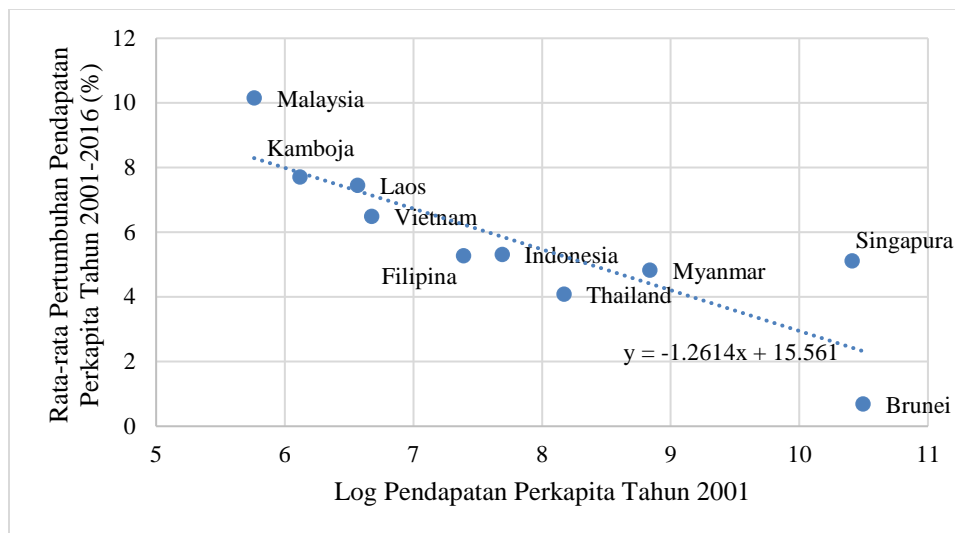
FI = *Financial Institutional Index* (Indeks Institusi Finansial);

FM = *Financial Market Index* (Indeks Pasar Finansial);

u_{it} = *error term*.

HASIL DAN PEMBAHASAN

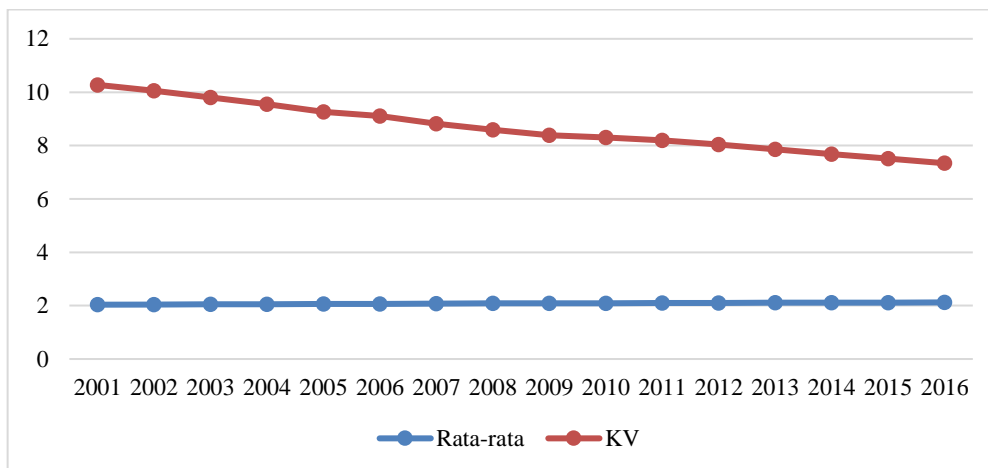
Perekonomian kawasan yang semakin maju, termasuk ASEAN, tidak dapat dilepaskan dari kajian mengenai kesenjangan ekonomi antar negara yang berada di dalam kawasan itu. Pendapatan perkapita negara-negara ASEAN secara umum mengalami peningkatan selama periode 2001 hingga 2016, sebagaimana ditunjukkan pada **Gambar 2**. Namun, berdasarkan **Gambar 2** tersebut masih belum dapat menjelaskan kecenderungan adanya penurunan kesenjangan pendapatan per kapita yang seharusnya ditunjukkan dengan selisih pendapatan per kapita yang semakin mengecil antar negara pada akhir



Gambar 3. Scatter Plot Rata-rata Pertumbuhan Pendapatan Perkapita Tahun 2001-2016 dan Log Pendapatan Perkapita Tahun 2001 untuk Negara-Negara ASEAN

periode penelitian. Meskipun masih terdapat *gap* yang cukup besar, pada **Gambar 3**

variasi dari logaritma natural pendapatan per kapita semua negara yang ada di



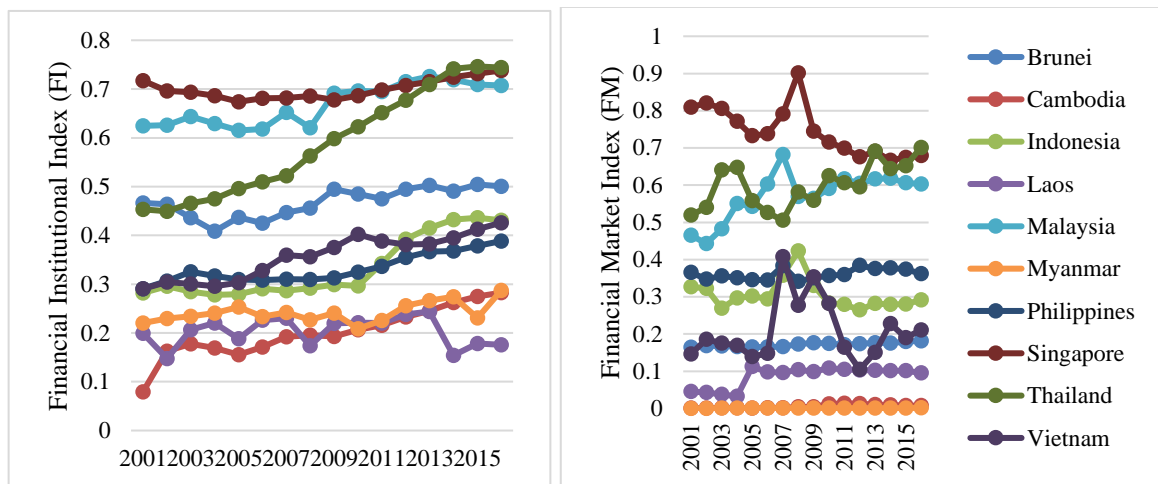
Gambar 4. Rata-rata dan Koefisien Variasi dari Logaritma Natural Pendapatan Perkapita Negara-negara ASEAN Tahun 2001-2016

secara riil menunjukkan bahwa pertumbuhan pendapatan per kapita negara-negara ASEAN tumbuh secara konvergen atau terjadi penurunan kesenjangan perekonomian antar negara-negara ASEAN (dilihat dari *slope* yang negatif). Kesenjangan yang terjadi akan menurun dikarenakan pertumbuhan pendapatan per kapita di negara berpendapatan rendah akan lebih cepat dibandingkan dengan pertumbuhan pendapatan per kapita negara berpendapatan tinggi sehingga pada akhirnya negara berpendapatan rendah akan mampu mengejar ketertinggalannya (Barro dan Sala-i-Martin, 2004).

Hal yang sama juga dapat dilakukan dengan menggunakan ukuran koefisien

ASEAN. Berdasarkan **Gambar 4**, selama tahun 2001 sampai 2016 nilai koefisien variasi logaritma natural pendapatan per kapita terus mengalami penurunan. Tren ini memperlihatkan terjadinya konvergensi sigma selama periode penelitian. Walaupun rata-ratanya meningkat sangat lambat, namun variasi dari pendapatan per kapita negara-negara ASEAN semakin menurun atau pendapatan per kapita negara-negara di ASEAN semakin homogen

Perkembangan finansial negara-negara ASEAN cukup menunjukkan prospek yang menjanjikan. Pada periode penelitian dari tahun 2001 hingga 2016, hampir seluruh negara anggota ASEAN mengalami peningkatan *Financial*



Gambar 5. *Financial Institutional Index* dan *Financial Market Index* Negara-negara ASEAN Tahun 2001-2016

Institutional Index (FI Index) dan *Financial Market Index* (FM Index) (Gambar 5). Penurunan FI Index hanya terjadi di negara Laos yaitu hanya 0.02 poin dari 0.20 di tahun 2001 menjadi 0.18 di tahun 2016 sedangkan penurunan FM Index terjadi di negara Indonesia (0.03 poin), Filipina (0.004 poin) dan Singapura (0.13 poin). Negara ASEAN dengan tingkat perkembangan finansial tertinggi adalah Singapura yang secara rata-rata pada periode 2001-2016 memiliki FI Index dan

FM Index yang tertinggi. Singapura memang dikenal sebagai salah satu pusat finansial terbesar di dunia (Almekinders, *et al.*, 2015). Kamboja merupakan negara ASEAN dengan rata-rata FI Index terendah dan Myanmar merupakan dan FM Index yang paling rendah pada periode 2001-2016.

Konvergensi sigma dapat mendeskripsikan kondisi pendapatan perkapita secara sederhana. Selanjutnya, konvergensi beta dihitung dengan

Tabel 1. Estimasi konvergensi ekonomi ASEAN Periode 2001-2016

Variabel endogen: Logaritma Natural GDP perkapita

Parameter	Konvergensi beta absolut		Konvergensi beta kondisional tanpa perkembangan finansial		Konvergensi beta kondisional dengan perkembangan finansial	
$\ln(\text{GDPC})_{t-1}$	0.9653	(0.0000)****	0.9930	(0.0000)****	0.5673	(0.0090)****
$\ln(\text{GCF})$			-0.0196	(0.4200)	0.0568	(0.2560)
$\ln(\text{GOV})$			-0.0386	(0.1640)	-0.1490	(0.0280)***
$\ln(\text{TO})$			0.0709	(0.0150)***	0.1648	(0.0220)***
INF			-0.0004	(0.0100)***	-0.0010	(0.1120)*
$\ln(\text{FI})$					0.5189	(0.0410)***
$\ln(\text{FM})$					0.1786	(0.0910)**
Wald Test	14604.770	(0.0000)****	4447.7000	(0.0000)****	6828.1800	(0.0000)****
Uji AB						
m1	-1.9117	(0.0559)**	-1.9783	(0.0479)***	1.7110	(0.0871)**
m2	-0.7704	(0.4411)	-0.9370	(0.3487)	-0.7557	(0.4498)
Uji Sargan	9.6526	(1.0000)	6.4031	(1.0000)	2.4557	(1.0000)
Speed of convergence	0.0353		0.0070		0.5669	
Half time convergence	19.6268		98.6740		1.2228	

Sumber: data diolah

Ket: *, **, *** dan **** berturut-turut menyatakan variabel signifikan pada taraf nyata 15 %, 10 %, 5 %, dan 1 %.

menggunakan model regresi panel dinamis FD-GMM (*First Different Generalized Method of Moment*). Sebelum dianalisis lebih lanjut, dilakukan uji spesifikasi model panel dinamis terhadap hasil estimasi model yang diperoleh. Uji konsistensi parameter dilakukan dengan menggunakan uji Arellano Bond untuk menguji bahwa tidak ada hubungan antara *first order* dan *second order* dengan *error* persamaan *first difference* dengan hipotesis sebagai berikut:

$$H_0: E(\Delta v_{it} \Delta v_{i,t-j}) = 0$$

$$H_1: E(\Delta v_{it} \Delta v_{i,t-j}) \neq 0$$

dimana $j = 1, 2$

Estimator yang digunakan dalam model sudah konsisten, ditunjukkan dari uji Arellano Bond (AB) dengan *p-value* m_1 sebesar 0,0559 yang signifikan pada taraf nyata 10 persen atau tolak H_0 pada m_1 dan gagal tolak H_0 pada m_2 dilihat dari *p-value* m_2 sebesar 0,4411 yang tidak signifikan di taraf nyata 1 persen, 5 persen, 10 persen, ataupun 15 persen.

Sementara itu, validitas hasil estimasi dibuktikan dengan uji Sargan dengan hipotesis sebagai berikut:

$$H_0: E(W' \Delta v_i) = 0 \text{ (Tidak ada hubungan antara variabel instrumen dengan error)}$$

$$H_1: E(W' \Delta v_i) \neq 0 \text{ (Ada hubungan antara variabel instrumen dengan error)}$$

Hasil estimasi menunjukkan bahwa nilai *p-value* sebesar 1,0000 atau belum cukup bukti untuk menolak H_0 , sehingga dapat dikatakan bahwa variabel instrumen yang digunakan sudah valid.

Tabel 1 menunjukkan bahwa konvergensi beta absolut juga terjadi selama periode penelitian, terlihat dari nilai koefisien dari lag variabel endogen yang kurang dari satu dan signifikan pada tingkat signifikansi 1 persen. Berdasarkan **Tabel 1**, persamaan yang terbentuk dinyatakan sebagai berikut:

1. Konvergensi beta absolut

$$\ln(GDPC_{it}) = 0,9653 \ln(GDPC_{i,t-1})^{****} + \dots \dots \dots (16)$$

2. Konvergensi beta kondisional tanpa perkembangan finansial

$$\ln(GDPC_{it}) = 0,9930 \ln(GDPC_{i,t-1})^{****} - 0,0196 \ln(GCF) - 0,0386 \ln(GOV) +$$

$$0,0709 \ln(TO)^{***} - 0,0004 \ln(INF)^{***} + \dots \dots \dots (17)$$

3. Konvergensi beta kondisional dengan perkembangan finansial

$$\ln(GDPC_{it}) = 0,5673 \ln(GDPC_{i,t-1})^{****} + 0,0568 \ln(GCF) - 0,1490 \ln(GOV)^{***} + 0,1648 \ln(TO)^{***} - 0,0010 \ln(INF)^* + 0,5189 \ln(FI)^{***} + 0,1786 \ln(FM)^{**} + \dots \dots \dots (18)$$

Berdasarkan model persamaan yang terbentuk, selama kurun waktu 2001 hingga 2016 di kawasan ASEAN terjadi konvergensi beta absolut karena koefisien lag variabel pendapatan per kapita sebesar 0,9653 atau kurang dari satu dan signifikan, dengan kecepatan konvergensi sebesar 3,53 persen. Artinya, sebesar 3,53 persen kesenjangan pendapatan per kapita antar negara-negara ASEAN saat ini dengan kondisi *steady state*-nya berkurang dalam waktu satu tahun sedangkan waktu untuk menutup setengah kesenjangan selama 20 tahun. Analisis menggunakan model konvergensi beta absolut hanya dapat menjelaskan kecepatan konvergensi saja karena hanya menggunakan lag variabel endogen sebagai satu-satunya variabel eksogen dan tidak melihat faktor-faktor lain. Model ini tidak dapat menjelaskan kebijakan yang dapat dilakukan karena lag variabel endogen sudah terjadi pada periode sebelumnya sehingga tidak dapat diintervensi pada tahun berjalan. Oleh karena itu perlu ditambahkan variabel endogen yang lain, yang lebih realistis terhadap kondisi perekonomian.

Pertumbuhan ekonomi merupakan ukuran yang digunakan untuk menggambarkan kondisi perekonomian suatu negara dari tahun ke tahun. Pertumbuhan ekonomi dari sisi penawaran menekankan peningkatan faktor produksi tenaga kerja, stok modal (aset) dan produktivitas dari kedua faktor produksi tersebut. Namun, penelitian-penelitian terbaru menekankan pentingnya pengembangan sektor keuangan sebagai faktor modal dalam meningkatkan pembangunan ekonomi (Mhadhbi, 2014). Dalam sejarah perkembangan ekonomi

makro, Irving Fisher dan Milton Friedman juga sudah terlebih dahulu mengubah pemikiran teori ekonomi klasik yang menekankan sektor riil menjadi pemikiran modern yang berpendapat bahwa sektor moneter sama pentingnya dengan sektor riil, bahkan sektor moneter lebih dominan karena dampak perubahan fenomena moneter bisa lebih cepat dan sangat luas (Lipsey, 2008).

Kajian konvergensi dengan memasukkan variabel eksogen yang sesuai dengan karakteristik negara masing-masing dilakukan dengan model konvergensi beta kondisional. Berdasarkan Tabel 2, dibuat model konvergensi beta kondisional tanpa perkembangan finansial dan dengan perkembangan finansial. Hasilnya, konvergensi juga terjadi di kawasan ASEAN pada periode penelitian, sebagaimana yang terjadi pada model konvergensi beta absolut. Kecepatan konvergensi model dengan variabel perkembangan finansial jauh lebih besar, yaitu mencapai 56,69 persen dibandingkan tanpa perkembangan finansial. Artinya, adanya variabel-variabel pada sektor keuangan meningkatkan kecepatan konvergensi. Hal ini menyebabkan *half life of convergence* dapat diturunkan dari 98,67 tahun menjadi hanya 1,22 tahun untuk menutupi setengah kesenjangan awal pendapatan perkapita riil antar negara di kawasan ASEAN.

Analisis model konvergensi beta kondisional dengan perkembangan finansial memberikan hasil koefisien regresi variabel investasi yang didefinisikan dari proporsi pembentukan modal bruto terhadap PDB atau *Gross Capital Formation* sebesar 0,0568, namun tidak signifikan berpengaruh terhadap PDB per kapita. Menurut penelitian Mhadhbi (2014) variabel investasi merupakan variabel kunci dalam pertumbuhan ekonomi karena investasi merupakan salah satu komponen ekonomi pembentuk PDB. Teori Solow menyatakan bahwa peningkatan investasi akan meningkatkan jumlah stok modal dan tingkat output dalam perekonomian (Mankiew, 2012). Jika tingkat output meningkat, maka pada akhirnya akan

meningkatkan pendapatan per kapita dan pertumbuhan ekonomi Negara ASEAN. Dalam hasil studi ini, koefisien regresi variabel investasi bertanda positif tetapi tidak berpengaruh signifikan pada tingkat kepercayaan yang digunakan. Penelitian ini semakin membuktikan bahwa pendapatan yang dihasilkan di negara-negara ASEAN tidak lagi didominasi oleh faktor-faktor riil, tetapi lebih dipengaruhi oleh faktor moneter, sesuai dengan teori yang diungkapkan Irving Fisher dan Milton Friedman.

Selanjutnya, variabel *Government Expenditure* menghasilkan koefisien regresi sebesar -0,1490. Dengan kata lain, pengeluaran pemerintah berpengaruh negatif terhadap pertumbuhan ekonomi. Artinya, setiap peningkatan 1 persen proporsi pengeluaran pemerintah negara-negara ASEAN terhadap PDB akan menurunkan pertumbuhan ekonomi negara-negara ASEAN sebesar 0,1490 persen, *ceteris paribus*. Hasil ini sesuai dengan penelitian Mhadhbi (2014) dan Karshenasan dan Mubarak (2012). Namun, hasil penelitian Mhadhbi (2014) menyimpulkan bahwa pengeluaran pemerintah tidak signifikan terhadap pertumbuhan ekonomi pada total sampel baik negara berkembang maupun negara maju. Hal ini disebabkan karena pengeluaran untuk publik pada beberapa sektor tidak teralokasikan dengan efisien sehingga tidak menaikkan pertumbuhan ekonomi. Sedangkan menurut argumen Keynes, jika defisit anggaran dibiayai dengan meminjam, suku bunga akan naik dan memberikan dampak negatif terhadap pertumbuhan ekonomi investasi swasta. Oleh karena itu, investasi pemerintah dapat digunakan sebagai proksi untuk keterlibatan pemerintah dalam akumulasi modal yang dapat memengaruhi pertumbuhan secara positif (Karshenasan dan Mubarak, 2012).

Variabel *Trade Openness* memiliki hasil koefisien regresi positif sebesar 0,1648 atau setiap peningkatan keterbukaan perdagangan sebesar 1 persen akan meningkatkan pertumbuhan ekonomi negara-negara ASEAN sebesar 0,1648 persen, *ceteris paribus*. Hasil ini sejalan

dengan penelitian Karshenasan dan Mubarak (2012) dan Mhadhbi (2014). Pada penelitian Mhadhbi (2014), variabel *Trade Openness* signifikan dan positif untuk negara maju dan total sampel negara-negara objek penelitian, sedangkan pada negara berkembang justru bertanda signifikan dan negatif. Karshenasan dan Mubarak (2012) menyatakan bahwa keterbukaan ekonomi yang dilakukan suatu negara memungkinkan perekonomian negara tersebut tumbuh lebih cepat karena dapat mengalokasikan sumber daya lebih efisien dan memungkinkan adanya difusi pengetahuan dan transfer teknologi sehingga investasi dapat masuk untuk mendorong pertumbuhan ekonomi.

Kemudian, variabel inflasi memiliki koefisien sebesar -0,0010 yang berarti bahwa jika terjadi peningkatan inflasi sebesar 1 persen maka pertumbuhan ekonomi negara-negara ASEAN akan menurun sebesar 0,0010 persen dengan asumsi *ceteris paribus*. Hasil ini sesuai dengan penelitian Mhadhbi (2014) yang menyimpulkan inflasi berpengaruh signifikan dan negatif pada total sampel baik negara berkembang maupun negara maju. Inflasi yang tinggi dapat mengubah keputusan investasi pada jangka panjang dan memengaruhi pertumbuhan ekonomi. Selain itu, inflasi yang berpengaruh negatif terhadap pertumbuhan ekonomi dapat disebabkan karena ketidakstabilan finansial. Sedangkan berdasarkan teori Keynesian, adanya inflasi menjadi insentif bagi para pekerja untuk menuntut kenaikan upah yang menjadi *shock supply* bagi perusahaan sehingga dilakukan rasionalisasi output oleh perusahaan dengan menurunkan jumlah output yang menyebabkan turunnya pertumbuhan ekonomi.

Variabel *Financial Institutional Index* atau indeks institusi finansial memiliki koefisien regresi sebesar 0,5189. Hal ini berarti bahwa indeks institusi finansial berpengaruh positif terhadap pertumbuhan ekonomi. Peningkatan 1 persen indeks institusi finansial akan meningkatkan pertumbuhan ekonomi negara-negara ASEAN sebesar 0,5189 persen. Tidak

berbeda dengan variabel *Financial Institution Index*, variabel *Financial Market Index* atau indeks pasar finansial yang juga berpengaruh positif terhadap pertumbuhan ekonomi negara-negara ASEAN sebesar 0,1786 yang artinya peningkatan 1 persen indeks pasar finansial akan meningkatkan pertumbuhan ekonomi negara ASEAN sebesar 0,1786 persen. Peningkatan pertumbuhan ekonomi karena adanya perkembangan finansial meningkatkan kecepatan konvergensi pendapatan per kapita menuju kondisi *steady-state*. Hal ini sesuai dengan hasil penelitian Mhadhbi (2014) dan Karshenasan dan Mobarak (2012). Pada penelitian Karshenasan dan Mobarak (2012), menyimpulkan bahwa variabel perkembangan finansial yang diukur dari empat dimensi yaitu (i) rasio *liquid liabilities* terhadap PDB (M_3/PDB), (ii) *liquid liabilities* dikurangi *narrow money* (M_3-M_1), (iii) kredit yang disalurkan kepada sektor privat (PRIV), dan (iv) kredit bank (BANK) berpengaruh signifikan dan positif meningkatkan pertumbuhan ekonomi. Penelitian yang sama oleh Mhadhbi (2014) juga menghasilkan kesimpulan yang sama dengan variabel perkembangan finansial yang diukur dari tiga dimensi yaitu (i) rasio *liquid liabilities* terhadap PDB (*Depth*), (ii) kredit privat (*Privat*), dan (iii) kredit yang dikeluarkan oleh bank ke sektor swasta terhadap PDB (*Bank*).

KESIMPULAN

Berdasarkan hasil analisis dan tujuan dari penelitian yang dilakukan maka diambil kesimpulan sebagai berikut. *Pertama*, pada periode penelitian, hampir seluruh negara anggota ASEAN mengalami peningkatan perkembangan finansial yang diukur dari *Financial Institution Index* (FI) dan *Financial Market Index* (FM). Peningkatan perkembangan finansial tertinggi yaitu Singapura baik diukur dari FI *index* maupun FM *index*, sedangkan Kamboja merupakan salah satu negara ASEAN dengan FI *index* terendah dan Myanmar dengan FM *index* terendah. *Kedua*, dilihat dari *scatter plot* antara log pendapatan per kapita dan rata-rata

pertumbuhan pendapatan per kapita dapat dikatakan terjadi konvergensi antar negara ASEAN ditunjukkan dari penurunan kesenjangan perekonomian antara negara-negara dengan pendapatan per kapita rendah dan negara-negara dengan pendapatan per kapita tinggi. Ketiga, estimasi model panel dinamis FD-GMM yang terpilih menghasilkan kesimpulan bahwa perkembangan finansial berpengaruh terhadap pertumbuhan ekonomi dan mempercepat terjadinya konvergensi. Hasil dari estimasi yang didapatkan yaitu variabel *FI Index* dan *FM Index* berpengaruh signifikan dan positif terhadap pertumbuhan ekonomi. Selain itu, variabel lag pendapatan per kapita, *government spending*, *trade openness*, dan inflasi berpengaruh signifikan terhadap pertumbuhan ekonomi sedangkan variabel *gross capital formation* berpengaruh tidak signifikan terhadap pertumbuhan ekonomi.

DAFTAR PUSTAKA

- Aghion, P., Howitt, P., Mayer-Foulkes, D. (2004). *The Effect of Financial Development on Convergence: Theory and Evidence*. NBER Working Paper, 10358.
- Almekinders, G., Fukuda, S., Mourmouras, A., Zhou, J. (2015). *ASEAN Financial Integration*. IMF Working Paper.
- ASEAN. (2018). *A resilient and Innovative ASEAN Community Annual Report 2017-2018*. ASEAN Secretariat Jakarta.
- Badunenko, O., Romero-Avila, D.. (2013). *Financial Development And The Sources Of Growth And Convergence*. *International Economic Review*. 54(2), 629-663.
- Baltagi, Badi H. (2005). *Econometric Analysis of Panel Data 3rd Edition*. England: John Wiley & Sons.
- Barro, Robert J and Xavier Sala-i Martin. (2004). *Economic Growth 2nd Edition*. London: MIT Press.
- Chowdhary, R., Jore, S., Thakur, R., Agrawal, K., & Geete, V. (2011). *Convergence of GDP per capita in ASEAN countries*. *Prestige International Journal of Management and Research*, 3(2), 1-9.
- Das, S. B., (2017). *Mind the Gap: Explaining Implementation Shortfalls in the ASEAN Economic Community*. ISEAS Economics Working Paper No. 2017-7.
- Devasmita, J, (2018). "Economic integration and income convergence in the EU and the ASEAN," *Journal of Economics Library, KSP Journals*, vol. 5(1), 1-11, March.
- Fung, M. K. (2009). *Financial development and economic growth: Convergence or divergence?*. *Journal of International Money and Finance*, 28, 56–67.
- Gugler, P., Vanoli, L. (2017). *Convergence or divergence of prosperity within the ASEAN community? A crucial issue for the success of the ASEAN economic community (AEC) process*, *International Journal of Emerging Markets*, 12(3), 610-624.
- Gujarati, Damodar N. (2003). *Basic Econometrics 4th Edition*. New York: McGraw-Hill.
- Ismail, N. W., (2008). *Growth and Convergence in ASEAN: A Dynamic Panel Approach*. *Int. Journal of Economics and Management*, 2(1), 127-140.
- Jan, Sajjad Ahmad and A.R. Chaudhary. (2011). *Testing The Conditional Convergence Hypothesis for Pakistan*. *Pak J. Commer Soc. Sci*. 2011 Vol. 5 (1), 117-128
- Jayanthakumaran, K. and Lee, S. (2013). *Evidence on the convergence of per capita income: a comparison of founder members of the Association of South East Asian Nations and the South Asian Association of Regional Cooperation*. *Pacific Economic Review*, 18 (1), 108-121.
- Karagiannis, Stellos. (2007). *The Knowledge-Based Economy, Convergence and Economic Growth: Evidence from the European Union*. Centre of Planning and Economic Research No 91.

- Karshenasan, A., Mobarak, A. (2012). *The relationship between financial development and economic growth convergence. Middle East Journal Of Business*. 7(4). 14-22.
- Lu, X., Fausten, D. K., Smyth, R. (2007). *Financial Development, Capital Accumulation and Productivity Improvement: Evidence from China. Journal of Chinese Economic and Business Studies*, 5(3), 227–242.
- Lucas, R. E. (1988). *On the mechanics of economic development, Journal of Monetary Economics*, 22, 3–42.
- Lipsey, G. Richard, Courant N. Paul, ragan T.S. Christopher. (2008). *Economics 13th Edition*. USA: Pearson Addison Wesley.
- Mankiew, N. Gregory. (2012). *Macroeconomics Eighth Edition*. New York: Worth Publishers.
- Mhadhbi, K. (2014). *Financial Development and Economic Growth: A Dynamic Panel Data Analysis. International Journal of Econometrics and Financial Management*, 2(2), 48-58.
- Michelis, L., Neaime, S. (2004), *Income convergence in the Asia-Pacific Region, Journal of Economic Integration*, Vol. 19(3), 470-498
- Robinson, J. (1952). *The generalization of the General Theory, The Rate of Interest and Other Essays*, 67–146 (London: Macmillan).
- Schumpeter, J. A. (1912). *Theorie der Wirtschaftlichen Entwicklung* (Leipzig, Germany: Dunker & Humblot). Translated by Redvers Opie (1934) (Cambridge, MA: Harvard University Press).
- Svirydzenka, K. (2016). *Introducing a New Broad-based Index of Financial Development*. IMF Working Paper.
- Todaro, M. dan S. Smith. (2006). *Pembangunan Ekonomi Edisi ke-9*. Jakarta: Erlangga.
- Young, Highins and Levy. (2007). *Sigma Convergence versus Beta Convergence: Evidence from U.S. Country-Level Data*, The Ohio State University.

DETERMINAN PARTISIPASI SEKOLAH ANAK PENYANDANG DISABILITAS DI INDONESIA TAHUN 2015

Dilla Citra Dewi¹, Ekaria Soebijarto²

¹BPS Kabupaten Pohuwato, ²Politeknik Statistika STIS
e-mail: ¹dilla.citra@bps.go.id, ²ekaria@stis.ac.id

Abstrak

Partisipasi bersekolah anak penyandang disabilitas di Indonesia masih rendah serta masih banyak anak disabilitas yang belum mendapat pendidikan yang memadai. Penelitian ini bertujuan untuk mengetahui variabel-variabel pada level individu dan pada level provinsi yang memengaruhi partisipasi sekolah anak penyandang disabilitas di Indonesia berdasarkan data Survei Penduduk Antar Sensus (SUPAS 2015). Metode analisis yang digunakan dalam penelitian ini adalah multilevel regresi logistik biner dua level. Hasil yang didapatkan antara lain variabel-variabel pada level individu yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas yaitu tingkat kesulitan disabilitas anak, jenis disabilitas anak, tingkat pendidikan Kepala Rumah Tangga (KRT), serta daerah tempat tinggal. Kemudian variabel-variabel pada level provinsi yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas adalah rasio ketersediaan Sekolah Luar Biasa (SLB) dan jarak ketersediaan SLB.

Kata kunci: disabilitas, anak penyandang disabilitas, pendidikan, partisipasi sekolah, multilevel regresi logistik

Abstract

Percentage of children with disabilities in Indonesia who have not received education yet (no or never attended school) based on Susenas 2012 is 35,25 % and those who do not attend school anymore are 16,03 %, while those who still study in school are only 48,73 % (BPS and Kemensos, 2013). This shows that children with disabilities in Indonesia are still lacking in school participation and many children have not received adequate education. This research aims to determine the individual and contextual variables that affect the school participation of children with disabilities in Indonesia based on SUPAS 2015. Analytical method used in this research is multilevel binary logistic regression with two levels. The obtained result is variables at individual level that significantly affected school participation of children with disabilities are severity of disability, the type of disability, the education level of the household head, and the area of residence. Variables at provincial level that significantly affected the school participation of children with disabilities are ratio of school's availability and distance of school's availability.

Keywords: *disability, children with disabilities, education, school participation, multilevel logistic regression*

PENDAHULUAN

Disabilitas merupakan bagian dari kondisi manusia yang mana keberadaan penyandang disabilitas tidak dapat dipisahkan dari kehidupan. Hal ini karena penyandang disabilitas juga merupakan bagian dari masyarakat. Banyak penyandang disabilitas kesulitan untuk mendapatkan akses yang setara terhadap pendidikan, kesempatan bekerja, dan pelayanan kesehatan, serta kesulitan mendapatkan pelayanan yang berkaitan dengan kebutuhan khusus yang mereka butuhkan (WHO, 2011) dan (Lamichhane, 2015). Padahal seharusnya perlakuan yang sama dalam pekerjaan, pendidikan, aksesibilitas, dan informasi teknologi juga merupakan hak bagi setiap penyandang disabilitas yang bertujuan untuk mencapai kemandirian, kesetaraan, dan kesejahteraan sosial bagi penyandang disabilitas (BPS & Kemensos, 2013). Hak-hak yang berhubungan dengan penyandang disabilitas juga telah diatur dalam Undang-Undang Republik Indonesia No. 8 Tahun 2016 tentang Penyandang Disabilitas. Disebutkan dalam pasal 1 UU ini bahwa “penyandang disabilitas adalah setiap orang yang mengalami keterbatasan fisik, intelektual, mental, dan/atau sensorik dalam jangka waktu lama yang dalam berinteraksi dengan lingkungan dapat mengalami hambatan dan kesulitan untuk berpartisipasi secara penuh dan efektif dengan warga negara lainnya berdasarkan kesamaan hak.”

Perkembangan penyandang disabilitas di Indonesia cenderung mengalami peningkatan. Berdasarkan hasil Susenas, pada tahun 2006 terdapat 1,38 persen penduduk Indonesia yang mengalami disabilitas, persentase ini meningkat pada tahun 2012 menjadi sebesar 2,45 persen. Demikian pula pada tahun 2015, persentase ini juga mengalami peningkatan, data Survei Penduduk Antar Sensus (SUPAS 2015) menunjukkan bahwa pada tahun 2015 persentase penyandang disabilitas sebesar 8,56 persen dari seluruh penduduk Indonesia.

Seperti yang telah disebutkan sebelumnya bahwa penyandang disabilitas

memiliki hak yang sama untuk mendapatkan pendidikan, hak untuk mendapatkan pendidikan ini juga telah diatur dalam UU No.8 Tahun 2016 tentang Penyandang Disabilitas pasal 5 ayat (1). Maka dari itu penelitian ini akan memfokuskan pembahasan pada salah satu permasalahan pada penyandang disabilitas yaitu mengenai pendidikan bagi anak penyandang disabilitas.

Berdasarkan hasil Susenas tahun 2012, persentase anak penyandang disabilitas yang belum memperoleh pendidikan (tidak/belum pernah sekolah) sebesar 35,25 persen serta yang tidak bersekolah lagi sebesar 16,03 persen, sedangkan yang masih bersekolah hanya 48,73 persen (BPS dan Kemensos, 2013). Hal ini menunjukkan bahwa partisipasi bersekolah anak penyandang disabilitas masih rendah serta masih banyak anak penyandang disabilitas yang belum mendapat pendidikan yang memadai.

Selain masalah rendahnya partisipasi sekolah, ketersediaan fasilitas pendidikan bagi anak penyandang disabilitas juga merupakan hal yang harus diperhatikan oleh pemerintah. Ini karena salah satu faktor penting untuk meningkatkan partisipasi sekolah anak penyandang disabilitas adalah tersedianya fasilitas pendidikan (Huisman & Smits, 2008) (Sugiarto, 2015). Berdasarkan data statistik Sekolah Luar Biasa (SLB) oleh Kemendikbud dapat diketahui bahwa jumlah SLB di Indonesia pada tahun 2015/2016 terdapat adalah sebanyak 1962 SLB, dimana 26,56 persen merupakan SLB berstatus negeri yang dikelola oleh pemerintah, sedangkan 73,44 persen lainnya merupakan SLB berstatus swasta yang dikelola oleh pihak swasta. Sedikitnya jumlah SLB yang dikelola oleh pemerintah ini mengindikasikan kurangnya peran pemerintah untuk memfasilitasi anak penyandang disabilitas agar dapat bersekolah.

Permasalahan lain yang berkaitan dengan pendidikan anak penyandang disabilitas adalah ketimpangan persentase partisipasi bersekolah anak penyandang disabilitas antarprovinsi di Indonesia

(Lampiran 1). Berdasarkan hasil Susenas tahun 2012, masih banyak provinsi di Indonesia memiliki persentase anak penyandang disabilitas yang bersekolah yang berada di bawah persentase nasional. Hal ini mengindikasikan bahwa terdapat ketidakseimbangan pembangunan fasilitas pendidikan antarprovinsi di Indonesia serta akses pelayanan pendidikan bagi anak penyandang disabilitas yang kurang merata antarprovinsi di Indonesia.

WHO (2001) menyatakan bahwa partisipasi dipengaruhi oleh kondisi kesehatan serta faktor lingkungan dan faktor personal. Sehingga dengan berpedoman kepada ICF (2001) tersebut maka partisipasi termasuk partisipasi bersekolah anak penyandang disabilitas yang dikaji pada penelitian ini dapat dipengaruhi oleh faktor lingkungan (wilayah) dan faktor personal (individu). Mengingat keterbatasan data yang ada maka dalam penelitian ini variabel-variabel pada level wilayah (provinsi) yang diteliti meliputi rasio ketersediaan SLB, jarak ketersediaan SLB, dan rasio murid-guru SLB, sementara itu variabel-variabel pada level individu yang diteliti meliputi jenis kelamin anak, jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, jenis kelamin KRT, serta tingkat pendidikan KRT (Lampiran 2).

Terdapat beberapa penelitian terdahulu yang relevan dengan penelitian ini. Lamichhane & Kawakatsu (2015) menemukan bahwa variabel-variabel yang berpengaruh terhadap partisipasi sekolah anak yaitu tingkat keparahan disabilitas anak, pendidikan tertinggi KRT, jenis kelamin KRT, bekerja di sektor pertanian, total pengeluaran rumah tangga dalam sebulan, serta jumlah anggota rumah tangga yang berusia kerja. Penelitian lain yang dilakukan oleh Sugiarto (2015) memperoleh hasil bahwa usia anak, jenis kecacatan anak, tingkat keparahan kecacatan anak, pendidikan KRT, status bekerja KRT, Anggota Rumah Tangga (ART) yang bersekolah, serta kondisi disabilitas KRT berpengaruh terhadap keikutsertaan anak dengan disabilitas untuk bersekolah.

Berdasarkan uraian di atas tujuan penelitian ini yaitu mengetahui profil penyandang disabilitas berdasarkan status partisipasi sekolah dan menganalisis faktor-faktor yang memengaruhi kecenderungan berpartisipasi sekolah anak penyandang disabilitas. Penelitian ini akan memfokuskan pembahasan pada salah satu permasalahan pada penyandang disabilitas yaitu mengenai pendidikan bagi anak penyandang disabilitas.

METODOLOGI

1. Metode Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekunder yang bersumber dari hasil Survei Penduduk Antar Sensus (SUPAS 2015) yang diselenggarakan oleh Badan Pusat Statistik (BPS). Data SUPAS 2015 merupakan data terbaru yang dapat digunakan untuk mendeteksi disabilitas. Unit analisis dalam penelitian ini adalah anak penyandang disabilitas berusia 7-18 tahun di Indonesia. Usia ini dipilih dikarenakan usia 7-18 tahun merupakan usia sekolah di Indonesia. Jenjang pendidikan di Indonesia yang dibagi menjadi pendidikan dasar (SD/MI/ sederajat) usia 7-12 tahun, pendidikan pertama (SMP/MTs/ sederajat) usia 12-15 tahun, dan pendidikan menengah (SMA/MA/ sederajat) usia 16-18 tahun (BPS, 2012).

Variabel respons yang digunakan pada penelitian ini adalah partisipasi sekolah anak penyandang disabilitas. Sedangkan variabel penjelasnya dibagi menjadi dua bagian sesuai dengan struktur data penelitian yang berbentuk hierarki, yaitu variabel-variabel pada level individu dan pada level provinsi. Mengingat keterbatasan data yang ada maka dalam penelitian ini variabel-variabel pada level wilayah (provinsi) yang diteliti meliputi rasio ketersediaan SLB, jarak ketersediaan SLB, dan rasio murid-guru SLB, sementara itu variabel-variabel pada level individu yang diteliti meliputi jenis kelamin anak, jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal,

jenis kelamin KRT, serta tingkat pendidikan KRT (Lampiran 2).

Variabel-variabel pada level individu, yaitu jenis kelamin anak, jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, jenis kelamin KRT, serta tingkat pendidikan KRT. Kemudian untuk variabel-variabel pada level provinsi dalam penelitian ini yaitu variabel rasio ketersediaan SLB dan rasio murid-guru SLB yang diperoleh dari data publikasi Statistik Sekolah Luar Biasa (SLB) 2015/2016 oleh Kemendikbud, serta variabel jarak ketersediaan SLB yang diperoleh dari data publikasi Kode dan Data Wilayah Administrasi Pemerintahan oleh Kemendagri, 2015.

2. Metode Analisis

Metode analisis yang digunakan dalam penelitian ini adalah analisis deskriptif dan analisis inferensia. Analisis deskriptif untuk mengetahui gambaran karakteristik anak penyandang disabilitas berdasarkan status partisipasi sekolahnya. Sedangkan analisis inferensia untuk menganalisis variabel-variabel yang memengaruhi dan kecenderungan berpartisipasi sekolah anak penyandang disabilitas. Analisis inferensia yang digunakan dalam penelitian ini adalah analisis multilevel regresi logistik biner dua level (bilevel regresi logistik biner). Level satu untuk tingkat individu dan level dua untuk tingkat provinsi. Menurut Hox (2010) analisis multilevel merupakan analisis yang digunakan untuk mengatasi masalah data dengan struktur hirarki. Model multilevel yang digunakan adalah model multilevel dengan *random intercept* karena diasumsikan bahwa pengaruh variabel bebas setiap kelompok adalah sama. Analisis multilevel regresi logistik biner digunakan karena variabel respons memiliki dua kategori yaitu anak penyandang disabilitas yang tidak bersekolah ($y = 0$) dan anak penyandang disabilitas yang bersekolah ($y = 1$).

Tahapan analisis diawali dengan uji kebebasan *Chi-square* dan uji *U Mann-Whitney*. Uji ini bertujuan untuk mengetahui variabel-variabel penjelas yang

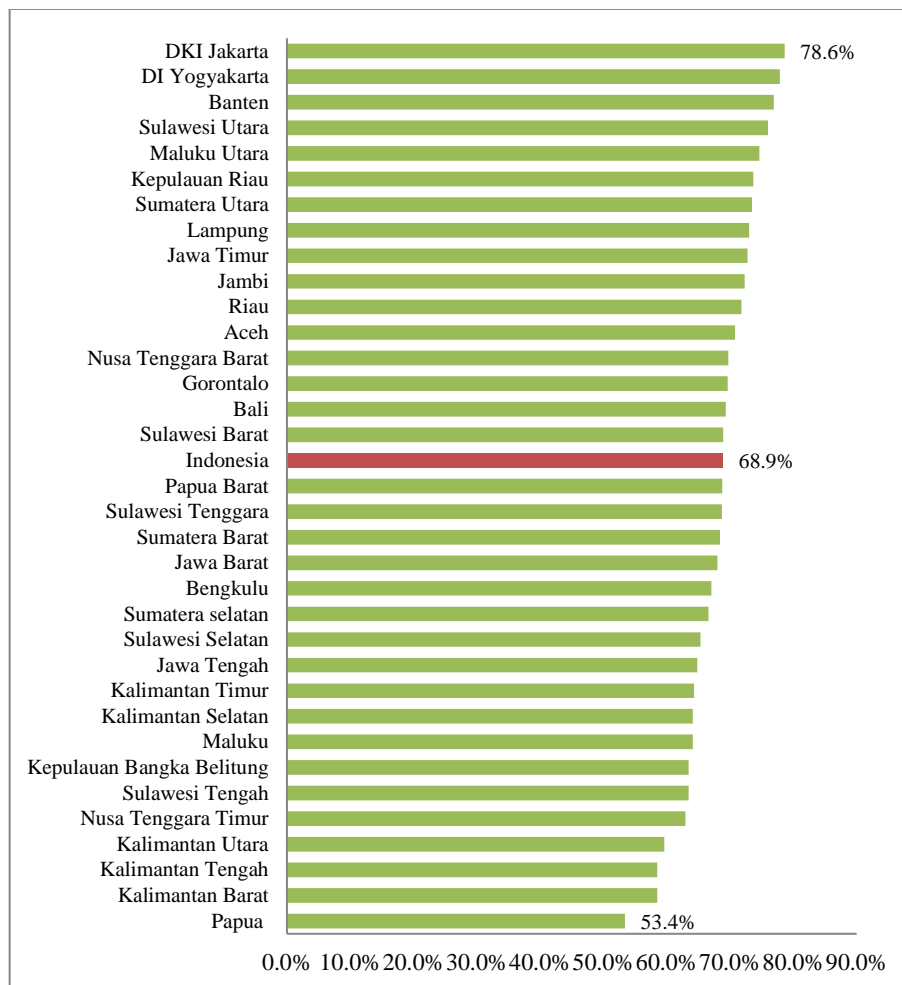
signifikan berhubungan dengan variabel partisipasi sekolah anak penyandang disabilitas untuk analisis bilevel regresi logistik biner. Kemudian melakukan pengujian signifikansi *random effect* dengan *Likelihood Ratio Test* untuk mengetahui apakah model multilevel regresi logistik biner lebih cocok digunakan daripada model regresi logistik biner satu level. Selanjutnya melakukan penghitungan variasi antar unit di level-2 menggunakan nilai *Intraclass Correlation Coefficient* (ICC).

Lalu dilakukan pengujian simultan yang bertujuan untuk mengetahui pengaruh seluruh variabel penjelas secara bersama-sama terhadap partisipasi sekolah anak penyandang disabilitas. Jika pengujian parameter secara simultan memberikan kesimpulan bahwa terdapat paling sedikit satu variabel penjelas yang memengaruhi partisipasi sekolah anak penyandang disabilitas, maka tahapan selanjutnya menguji variabel penjelas secara parsial. Pengujian parsial digunakan untuk mengetahui variabel penjelas yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas secara parsial. Setelah mengetahui variabel-variabel yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas, tahap selanjutnya adalah menginterpretasikan nilai *Odds Ratio* (OR).

HASIL DAN PEMBAHASAN

Salah satu indikator yang digunakan sebagai pemantau program pendidikan yang telah digulirkan pemerintah adalah partisipasi sekolah. Indikator ini mencerminkan keefektifan program pendidikan dalam menyerap potensi pendidikan yang terdapat dalam masyarakat. Semakin efektifnya suatu program dapat dilihat dari semakin tingginya nilai indikator tersebut (BPS, 2012). Berdasarkan hasil pengolahan, terdapat 68,9 persen anak penyandang disabilitas usia 7-18 tahun di Indonesia yang bersekolah, sedangkan 31,1 persen lainnya tidak bersekolah.

Gambar 1 di bawah ini menyajikan keragaman persentase anak penyandang



Sumber: SUPAS 2015 (diolah)

Gambar 1. Persentase anak penyandang disabilitas usia 7-18 tahun yang masih bersekolah di Indonesia tahun 2015

disabilitas usia 7-18 tahun yang masih bersekolah antarprovinsi di Indonesia. Persentase anak penyandang disabilitas yang masih bersekolah tertinggi terdapat di Provinsi DKI Jakarta yaitu sebesar 78,6 persen. Sedangkan Provinsi Papua memiliki persentase anak penyandang disabilitas yang masih bersekolah terendah yaitu hanya sebesar 53,4 persen.

Dengan merujuk anak penyandang disabilitas yang bersekolah secara nasional, maka masih banyak provinsi yang memiliki persentase anak penyandang disabilitas yang bersekolah yang berada di bawah sebesar 68,9 persen. Berdasarkan Gambar 2 dapat diketahui pula bahwa terdapat keragaman serta perbedaan persentase anak penyandang disabilitas yang masih bersekolah antarprovinsi di Indonesia. Keragaman persentase ini mengindikasikan adanya efek dari karakteristik antarprovinsi yang menyebabkan adanya

perbedaan persentase anak penyandang disabilitas. Ini berarti bahwa masih terdapat ketidakseimbangan pembangunan fasilitas pendidikan antarprovinsi di Indonesia. Artinya, akses pelayanan pendidikan bagi anak penyandang disabilitas antarprovinsi di Indonesia belum merata.

Apabila dikaji lebih lanjut berdasarkan karakteristik sosial demografi dan status partisipasi sekolah, partisipasi bersekolah anak penyandang disabilitas yang lebih rendah terjadi pada anak penyandang disabilitas laki-laki (68,5%), mengalami lebih dari satu kesulitan (disabilitas ganda) (42,5%), memiliki tingkat kesulitan disabilitas yang selalu kesulitan (30,8%), tinggal di daerah perdesaan (64,3%), memiliki kepala rumah tangga perempuan (66,5%), serta memiliki kepala rumah tangga dengan tingkat pendidikan SMP ke bawah (63,8%). Tabel 1 berikut menyajikan persentase partisipasi

Tabel 1. Persentase partisipasi sekolah anak penyandang disabilitas menurut karakteristiknya

Variabel	Kategori	Partisipasi Sekolah (%)	
		Bersekolah	Tidak Bersekolah
(1)	(2)	(3)	(4)
Jenis kelamin anak	1. Laki-laki	68,5	31,5
	2. Perempuan	69,3	30,7
Jenis disabilitas anak	1. Lebih dari satu kesulitan	42,5	57,5
	2. Kesulitan penglihatan	83,1	16,9
	3. Kesulitan pendengaran	80,7	19,3
	4. Kesulitan berjalan	79,9	20,1
	5. Kesulitan menggerakkan tangan	77,1	22,9
	6. Kesulitan konsentrasi	76,1	23,9
	7. Gangguan perilaku/emosional	82,9	17,1
	8. Kesulitan berbicara/komunikasi	60,2	39,8
	9. Kesulitan mengurus diri sendiri	87,2	12,8
Tingkat kesulitan disabilitas anak	1. Selalu kesulitan	30,8	69,2
	2. Seringkali kesulitan	54,9	45,1
	3. Sedikit kesulitan	80,6	19,4
Daerah tempat tinggal	1. Perkotaan	74,8	25,2
	2. Perdesaan	64,3	35,7
Jenis kelamin KRT	1. Laki-laki	69,4	30,6
	2. Perempuan	66,5	33,5
Tingkat pendidikan KRT	1. \leq SMP	63,8	36,2
	2. $>$ SMP	80,4	19,6

Sumber: SUPAS 2015 (diolah)

sekolah anak penyandang disabilitas menurut karakteristiknya.

Sebelum melakukan analisis bilevel regresi logistik untuk partisipasi sekolah anak penyandang disabilitas, dilakukan uji kebebasan *Chi-square* dan uji *U Mann-Whitney*. Uji ini untuk mengetahui bagaimana hubungan antara variabel partisipasi sekolah anak penyandang disabilitas dan variabel jenis kelamin anak, jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, jenis kelamin KRT, tingkat pendidikan KRT, rasio ketersediaan SLB, jarak ketersediaan SLB, serta rasio murid-guru SLB. Semua merupakan variabel berskala kategorik, maka pengujian bagaimana hubungannya dengan partisipasi sekolah anak penyandang disabilitas dilakukan dengan menggunakan uji kebebasan *Chi-square*. Berdasarkan hasil pengujian *Chi-square* diketahui bahwa variabel jenis disabilitas anak, tingkat kesulitan disabilitas

anak, daerah tempat tinggal, jenis kelamin KRT, dan tingkat pendidikan KRT signifikan berhubungan dengan variabel partisipasi sekolah anak penyandang disabilitas karena nilai *p-value* yang didapatkan kurang dari *alpha* 0,05. Sementara itu, variabel jenis kelamin anak tidak signifikan berhubungan dengan variabel partisipasi sekolah anak penyandang disabilitas. Pengujian hubungan antara variabel partisipasi sekolah anak penyandang disabilitas dan rasio ketersediaan SLB, jarak ketersediaan SLB, serta rasio murid-guru SLB yang merupakan variabel numerik dilakukan dengan menggunakan uji *U Mann-Whitney*. Berdasarkan hasil uji *U Mann-Whitney* dapat diketahui bahwa partisipasi sekolah menurut variabel rasio ketersediaan SLB dan jarak ketersediaan SLB signifikan berbeda karena nilai *p-value* yang didapatkan kurang dari *alpha* 0,05. Hal ini mengindikasikan bahwa terdapat hubungan

antara variabel partisipasi sekolah anak penyandang disabilitas dengan variabel rasio ketersediaan SLB dan jarak ketersediaan SLB. Sementara itu, variabel rasio murid-guru SLB tidak signifikan berhubungan dengan variabel partisipasi sekolah anak penyandang disabilitas. Setelah dilakukan pengujian hubungan dan didapatkan variabel-variabel yang signifikan berhubungan dengan partisipasi sekolah anak penyandang disabilitas yaitu variabel daerah tempat tinggal, jenis kelamin KRT, tingkat pendidikan KRT, jenis disabilitas anak, tingkat kesulitan disabilitas anak, rasio ketersediaan SLB dan jarak ketersediaan SLB maka selanjutnya akan dilakukan analisis bilevel regresi logistik untuk partisipasi sekolah anak

penyandang disabilitas dengan menggunakan ketujuh variabel tersebut.

Pengujian kecocokan model bilevel regresi logistik untuk partisipasi sekolah anak penyandang disabilitas memberikan hasil bahwa terdapat *random effect* yang signifikan. Sehingga model multilevel regresi logistik biner lebih cocok digunakan daripada model regresi logistik biner satu level ($LR=96,31 > \chi^2_{(0,05;1)} = 3,84$).

Selanjutnya nilai *Intraclass Corelation Coefficient* (ICC) yang diperoleh sebesar 0,0174 berarti bahwa variasi partisipasi sekolah anak penyandang disabilitas di Indonesia yang disebabkan oleh perbedaan karakteristik antarprovinsi sebesar 1,74 persen. Menurut Nezlek (2008) jika penelitian memiliki data yang berstruktur hierarki, maka analisis multilevel perlu

Tabel 2. Hasil uji setiap variabel dalam model bilevel regresi logistik untuk partisipasi sekolah anak penyandang disabilitas

Variabel (1)	Koefisien (2)	p-value (3)	Odds Ratio (4)
Konstanta	-1,615	0,000*	0,1992
Variabel pada Level Individu			
Jenis Disabilitas Anak			
Lebih dari satu kesulitan**			
Kesulitan penglihatan	1,1842	0,000*	3,2703
Kesulitan pendengaran	1,238	0,000*	3,449
Kesulitan berjalan	0,821	0,000*	2,2743
Kesulitan menggerakkan tangan	1,249	0,000*	3,4885
Kesulitan konsentrasi	0,917	0,000*	2,502
Gangguan perilaku/emosional	1,1522	0,000*	3,1662
Kesulitan berbicara/berkomunikasi	0,2933	0,034*	1,340
Kesulitan mengurus diri sendiri	1,543	0,000*	4,6777
Tingkat Kesulitan Disabilitas Anak			
Selalu kesulitan**			
Seringkali kesulitan	0,945	0,000*	2,5746
Sedikit kesulitan	1,916	0,000*	6,7968
Daerah Tempat Tinggal			
Perdesaan **			
Perkotaan	0,132	0,019*	1,141
Tingkat Pendidikan KRT			
≤ SMP**			
> SMP	0,744	0,000*	2,105
Variabel pada Level Provinsi			
Rasio Ketersediaan SLB	8×10^{-5}	0,015*	1,001
Jarak Ketersediaan SLB	-2×10^{-5}	0,001*	0,9991,0

Sumber: SUPAS 2015 (diolah)

Keterangan:

*) Signifikan pada taraf uji 5 persen

***) Menyatakan kategori referensi

untuk digunakan meskipun nilai ICC mendekati 0.

Pengujian seluruh variabel secara bersamaan dalam model bilevel regresi logistik untuk partisipasi sekolah anak penyandang disabilitas memberikan hasil bahwa terdapat paling sedikit satu variabel penjelas dalam model yang memengaruhi partisipasi sekolah anak penyandang disabilitas ($G=2333,6654 > \chi^2_{(0,05;15)} = 25$). Pengujian setiap variabel dalam model bilevel regresi logistik untuk partisipasi sekolah anak penyandang disabilitas memberikan hasil bahwa variabel jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, tingkat pendidikan KRT, rasio ketersediaan SLB, dan jarak ketersediaan SLB signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas. Sementara itu, variabel jenis kelamin KRT tidak signifikan memengaruhi variabel partisipasi sekolah anak penyandang disabilitas.

Selanjutnya dilakukan pengujian parsial. Variabel jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, tingkat pendidikan KRT, rasio ketersediaan SLB, dan jarak ketersediaan SLB signifikan memengaruhi partisipasi sekolah. Dengan demikian dapat diinterpretasikan kecenderungan anak penyandang disabilitas untuk berpartisipasi sekolah, dan dapat membentuk persamaan dengan variabel-variabel yang signifikan tersebut. Hasil pengujian parsial disajikan dalam Tabel 2.

Berdasarkan Tabel 2 diperoleh hasil bahwa variabel jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, tingkat pendidikan KRT, rasio ketersediaan SLB, dan jarak ketersediaan SLB signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas. Selanjutnya persamaan bilevel regresi logistik untuk partisipasi bersekolah anak penyandang disabilitas yang terbentuk adalah:

$$\begin{aligned} & \ln\left(\frac{\widehat{\pi}_{ij}}{1 - \widehat{\pi}_{ij}}\right) \\ &= -1,615 \\ &+ 1,184 \text{Kesulitan penglihatan}_{ij} \\ &+ 1,238 \text{Kesulitan pendengaran}_{ij} \\ &+ 0,821 \text{Kesulitan berjalan}_{ij} \\ &+ 1,249 \text{Kesulitan menggerakkan tangan}_{ij} \\ &+ 0,917 \text{Kesulitan konsentrasi}_{ij} \\ &+ 1,152 \text{Gangguan perilaku} \\ &\text{/emosional}_{ij} \\ &+ 0,293 \text{Kesulitan berbicara} \\ &\text{/berkomunikasi}_{ij} \\ &+ 1,543 \text{Kesulitan mengurus diri sendiri}_{ij} \\ &+ 0,945 \text{Seringkali kesulitan}_{ij} \\ &+ 1,916 \text{Sedikit kesulitan}_{ij} \\ &+ 0,132 \text{Perkotaan}_{ij} + 0,744 > \text{SMP}_{ij} \\ &+ 0,00008 \text{Rasio Ketersediaan SLB}_j \\ &- 0,00002 \text{Jarak Ketersediaan SLB}_j \\ &+ \widehat{u}_{0j} \end{aligned}$$

Dari persamaan bilevel regresi logistik dapat diketahui seberapa besar pengaruh dan kecenderungan dari variabel tingkat kesulitan disabilitas anak, jenis disabilitas anak, tingkat pendidikan KRT, daerah tempat tinggal, rasio ketersediaan SLB, dan jarak ketersediaan SLB terhadap partisipasi bersekolah anak penyandang disabilitas. Hasil pengujian parsial pada Tabel 2 menunjukkan bahwa tingkat kesulitan disabilitas anak berpengaruh signifikan terhadap partisipasi sekolah anak penyandang disabilitas. Anak penyandang disabilitas yang memiliki tingkat kesulitan disabilitas sedikit kesulitan memiliki kecenderungan 6,796 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang sering kesulitan dengan asumsi variabel lain konstan. Sementara itu anak penyandang disabilitas yang seringkali kesulitan memiliki kecenderungan 2,574 kali lebih besar untuk bersekolah dibandingkan dengan yang selalu kesulitan dengan asumsi variabel lain konstan.

Hasil ini menunjukkan bahwa semakin sulit/parah disabilitas yang dialami oleh seorang anak, maka partisipasi bersekolahnya akan semakin kecil. Hasil ini sejalan dengan hasil penelitian Lamichhane & Kawakatsu (2015) yang menunjukkan

bahwa terdapat pengaruh negatif parahnya disabilitas anak terhadap peluang anak dalam berpartisipasi sekolah. Hal ini menunjukkan bahwa anak penyandang disabilitas mengalami rintangan dan hambatan ketika akan mengakses pendidikan. Rintangan tersebut dapat berupa infrastruktur yang sulit diakses oleh penyandang disabilitas dan kurangnya dukungan dari institusi yang terkait dengan penyediaan pelayanan pendidikan. Penelitian oleh Lehr et. al (2004) dalam Wieringo (2015) juga mendukung hasil dalam penelitian ini. Hasil penelitian tersebut menunjukkan bahwa siswa penyandang disabilitas yang memiliki tingkat keparahan disabilitas yang lebih berat mengalami tingkat absen yang tinggi dalam bersekolah atau cenderung tidak berpartisipasi sekolah. Hasil penelitian Sugiarto (2015) menunjukkan bahwa semakin parah jenis disabilitas yang dialami oleh seorang anak maka akan menghambat anak penyandang disabilitas untuk bersekolah, . Kondisi ini disebabkan anak penyandang disabilitas yang mengalami kesulitan yang parah/berat memerlukan perhatian dan perlakuan yang lebih khusus dibandingkan dengan anak penyandang disabilitas yang mengalami kesulitan yang ringan.

Jenis disabilitas anak berpengaruh signifikan terhadap partisipasi sekolah anak penyandang disabilitas. Anak penyandang disabilitas yang mengalami kesulitan/gangguan untuk mengurus diri sendiri (seperti mandi, makan, berpakaian, buang air besar, buang air kecil) memiliki kecenderungan 4,677 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan. Selanjutnya anak penyandang disabilitas yang mengalami kesulitan menggunakan/menggerakkan tangan/jari memiliki kecenderungan 3,488 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan. Kemudian anak penyandang

disabilitas yang mengalami kesulitan/gangguan pendengaran memiliki kecenderungan 3,449 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan. Sementara itu anak penyandang disabilitas yang mengalami kesulitan/gangguan penglihatan memiliki kecenderungan 3,270 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan.

Anak penyandang disabilitas yang mengalami gangguan perilaku dan atau emosional memiliki kecenderungan 3,166 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan. Selanjutnya anak penyandang disabilitas yang mengalami kesulitan/gangguan dalam hal mengingat atau berkonsentrasi memiliki kecenderungan 2,502 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan. Sementara itu anak penyandang disabilitas yang mengalami kesulitan/gangguan berjalan/naik tangga memiliki kecenderungan 2,274 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan. Selanjutnya anak penyandang disabilitas yang mengalami kesulitan/gangguan berbicara dan atau memahami/berkomunikasi dengan orang lain memiliki kecenderungan 1,340 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang mengalami lebih dari satu kesulitan (disabilitas ganda) dengan asumsi variabel lain konstan.

Hasil yang diperoleh dalam penelitian ini sejalan dengan konsep WHO, WHO (2001) menjelaskan bahwa tingkat

partisipasi sekolah anak berbeda tergantung kepada jenis disabilitas yang dialami, anak yang mengalami disabilitas fisik umumnya memiliki tingkat partisipasi sekolah yang lebih baik dibandingkan dengan anak yang mengalami disabilitas intelektual atau disabilitas sensorik. Didukung pula dengan hasil penelitian Sugiarto (2015) bahwa partisipasi sekolah anak yang mengalami disabilitas fisik lebih besar dibandingkan dengan anak yang mengalami disabilitas mental.

Tingkat pendidikan kepala rumah tangga berpengaruh positif dan signifikan terhadap partisipasi sekolah anak penyandang disabilitas. Anak penyandang disabilitas yang memiliki KRT dengan tingkat pendidikan lebih dari SMP (minimal SMA) memiliki kecenderungan 2,105 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang memiliki KRT dengan tingkat pendidikan SMP ke bawah dengan asumsi variabel lain konstan. Hasil tersebut mengindikasikan bahwa semakin tinggi tingkat pendidikan yang dimiliki oleh seorang KRT maka akan menyebabkan tingginya pula kecenderungan anak penyandang disabilitas untuk berpartisipasi sekolah. Sejalan dengan hasil penelitian Lamichhane & Kawakatsu (2015) bahwa kepala rumah tangga yang memiliki tingkat pendidikan menengah ke atas memiliki pengaruh positif terhadap partisipasi sekolah anak. Jika orang tua memiliki level pendidikan yang tinggi maka partisipasi anak untuk tetap bersekolah juga turut meningkat (Huisman & Smits, 2014). Diperjelas oleh Lamichhane (2015) yang menyatakan bahwa pendidikan orang tua baik pendidikan ibu maupun pendidikan ayah memiliki pengaruh positif terhadap partisipasi sekolah anak, orang tua yang memiliki pendidikan yang tinggi akan mementingkan pendidikan anaknya supaya mendapatkan pendidikan yang tinggi pula. Hasil penelitian ini juga didukung oleh hasil penelitian Sugiarto (2015) bahwa pendidikan KRT memiliki pengaruh yang positif terhadap partisipasi sekolah anak penyandang disabilitas.

Daerah tempat tinggal berpengaruh signifikan terhadap partisipasi sekolah anak penyandang disabilitas. Terdapat hubungan yang positif antara daerah tempat tinggal terhadap partisipasi sekolah anak penyandang disabilitas. Anak penyandang disabilitas yang tinggal di perkotaan memiliki kecenderungan 1,141 kali lebih besar untuk bersekolah dibandingkan dengan anak penyandang disabilitas yang tinggal di perdesaan dengan asumsi variabel lain konstan. Sejalan dengan hasil penelitian Lamichhane (2015) yang menemukan bahwa hubungan yang negatif terhadap partisipasi sekolah terdapat pada anak yang bertempat tinggal pada daerah perdesaan (rural), orang yang tinggal di daerah perdesaan biasanya bekerja pada sektor pertanian, sehingga anak-anak dalam keluarga tersebut diminta untuk membantu bekerja di sektor pertanian, selain itu anak-anak yang tinggal di daerah perdesaan juga harus berhadapan dengan sekolah yang berjarak jauh dari rumah mereka.

Variabel pada level provinsi yakni rasio ketersediaan SLB per 10.000 anak penyandang disabilitas berpengaruh signifikan terhadap partisipasi sekolah anak penyandang disabilitas. Setiap kenaikan satu rasio ketersediaan SLB di suatu provinsi, kecenderungan anak penyandang disabilitas untuk bersekolah akan berubah sebesar 1,000077 kali dengan asumsi variabel lain konstan. Sejalan dengan hasil penelitian Huisman & Smits (2008) bahwa ketersediaan sekolah akan berpengaruh positif terhadap kecenderungan anak untuk berpartisipasi sekolah, ketika sekolah tidak tersedia maka anak-anak tidak akan bisa untuk bersekolah.

Jarak ketersediaan SLB berpengaruh signifikan terhadap partisipasi sekolah anak penyandang disabilitas. Terdapat pengaruh yang negatif antara jarak ketersediaan SLB terhadap partisipasi sekolah anak penyandang disabilitas. Setiap kenaikan satu rasio jarak ketersediaan SLB di suatu provinsi, kecenderungan anak penyandang disabilitas untuk bersekolah akan berubah sebesar 0,9999802 kali. Jarak ketersediaan sekolah berpengaruh negatif terhadap kecenderungan anak untuk tetap

berpartisipasi sekolah. Jika jarak menuju sekolah semakin jauh maka semakin partisipasi sekolah anak cenderung menurun (Huisman & Smits, 2008). Menurut Colclough, Rose, & Tembon (2000) jarak dari rumah menuju sekolah memiliki pengaruh terhadap kehadiran di sekolah pada seorang anak, semakin jauh jarak dari rumah menuju sekolah maka semakin rendah partisipasi sekolah anak. Didukung oleh Duze (2010) yang menjelaskan bahwa jarak dari rumah menuju ke sekolah yang jauh mempunyai dampak buruk terhadap kehadiran anak di sekolah.

KESIMPULAN DAN SARAN

1. Kesimpulan

Berdasarkan hasil dan pembahasan dalam penelitian ini maka diperoleh kesimpulan sebagai berikut:

1. Sebagian besar anak penyandang disabilitas usia 7-18 tahun di Indonesia berpartisipasi sekolah. Partisipasi bersekolah anak penyandang disabilitas yang lebih rendah terjadi pada anak penyandang disabilitas laki-laki, mengalami disabilitas ganda, memiliki tingkat kesulitan disabilitas yang selalu kesulitan, tinggal di daerah perdesaan, memiliki kepala rumah tangga perempuan, serta memiliki kepala rumah tangga dengan tingkat pendidikan paling tinggi SMP.
2. Partisipasi sekolah anak penyandang disabilitas dipengaruhi oleh variabel jenis disabilitas anak, tingkat kesulitan disabilitas anak, daerah tempat tinggal, serta tingkat pendidikan KRT. Kemudian variabel kontekstual yang signifikan memengaruhi partisipasi sekolah anak penyandang disabilitas adalah rasio ketersediaan SLB dan jarak ketersediaan SLB.
3. Kecenderungan berpartisipasi sekolah lebih besar pada anak penyandang disabilitas yang memiliki tingkat kesulitan disabilitas yang sedikit kesulitan, memiliki kesulitan

mengurus diri sendiri, memiliki kepala rumah tangga dengan tingkat pendidikan minimal SMA, tinggal di daerah perkotaan, tinggal dalam provinsi yang memiliki rasio ketersediaan SLB yang lebih besar, dan tinggal dalam provinsi yang memiliki jarak ketersediaan SLB yang lebih kecil.

2. Saran

Berdasarkan kesimpulan yang diperoleh dalam penelitian ini, beberapa saran yang dapat diberikan antara lain:

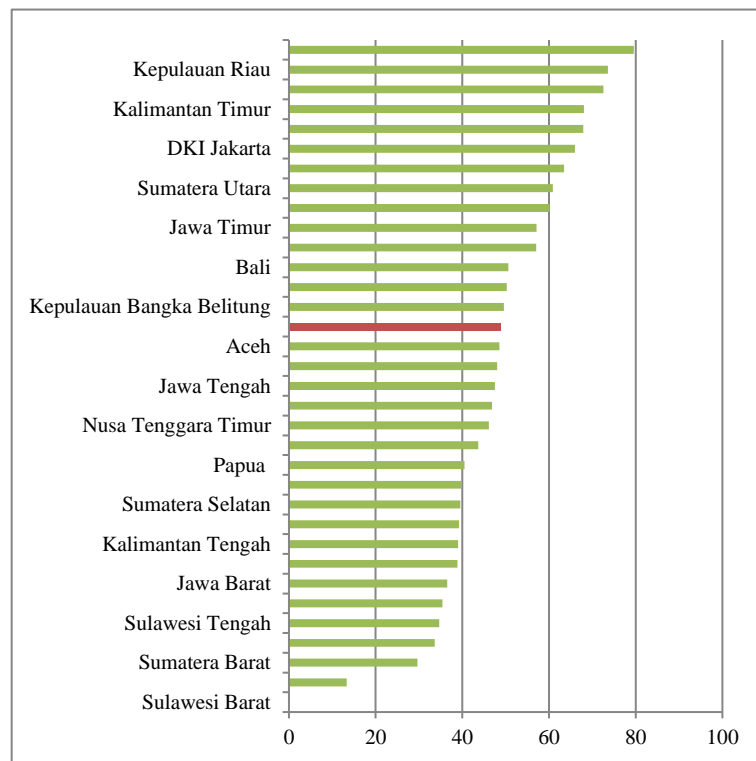
1. Pemerintah dapat memberikan dana pendidikan kepada anak penyandang disabilitas khususnya kepada anak penyandang disabilitas yang memiliki tingkat kesulitan disabilitas yang selalu kesulitan, mengalami disabilitas ganda, dan yang bertempat tinggal di perdesaan.
2. Orang tua anak penyandang disabilitas khususnya bagi orang tua yang mempunyai anak penyandang disabilitas yang memiliki tingkat kesulitan disabilitas yang selalu kesulitan, mengalami disabilitas ganda, dan yang bertempat tinggal di perdesaan dapat memberikan dukungan berupa dukungan material kepada anak mereka agar dapat bersekolah, serta untuk pemerintah dapat memberikan penyuluhan kepada orang tua anak penyandang disabilitas mengenai pentingnya pendidikan bagi anak mereka.
3. Pemerintah melalui Kemendikbud dapat meningkatkan penyediaan infrastruktur pendidikan bagi anak penyandang disabilitas pada wilayah yang rasio ketersediaan SLBnya masih kecil, dan dapat menyediakan infrastruktur pendidikan yang mudah diakses oleh anak penyandang disabilitas pada wilayah yang jarak ketersediaan SLBnya masih besar.
4. Keterbatasan penelitian ini adalah ketersediaan variabel yang terbatas karena menggunakan data sekunder. Bagi penelitian selanjutnya dapat dilakukan dengan mengkaji

permasalahan pendidikan anak penyandang disabilitas secara lebih mendalam berdasarkan faktor internal dan eksternal anak penyandang disabilitas menggunakan data primer dengan pendekatan kepada orang tua anak penyandang disabilitas.

DAFTAR PUSTAKA

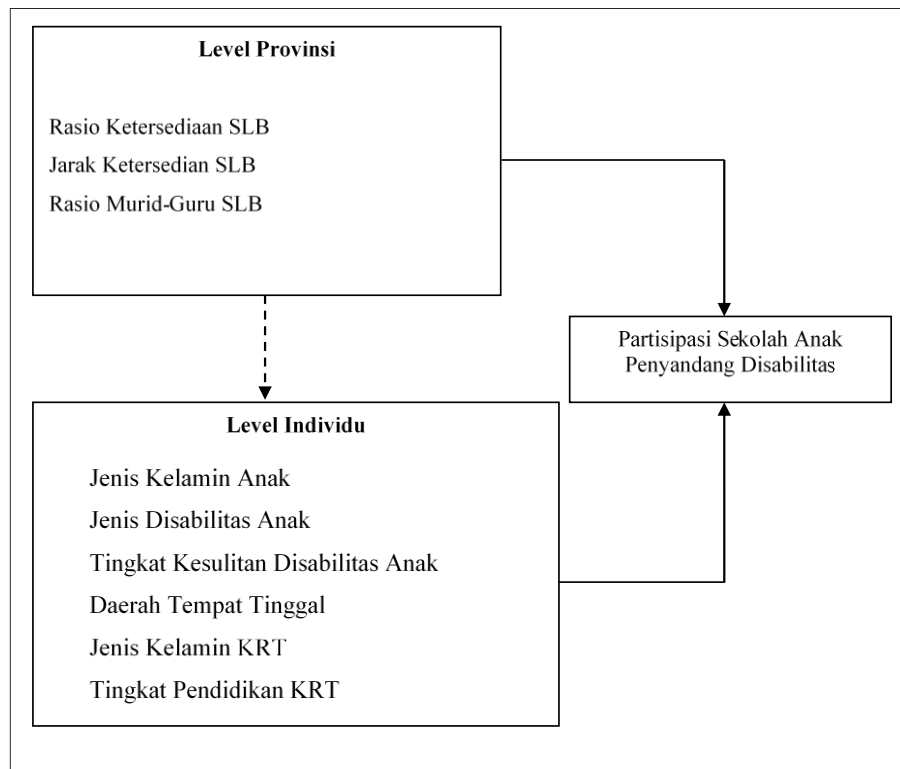
- Badan Pusat Statistik & Kementerian Sosial Republik Indonesia. (2013). *Profil Penyandang Masalah Kesejahteraan Sosial Indonesia Berdasarkan Susenas Tahun 2012*. Jakarta: BPS & Kemensos.
- Badan Pusat Statistik. (2012). *Statistik Pendidikan 2012*. Jakarta: BPS.
- Colclough, C., Rose, P., & Tembon, M. (2000). *Gender Inequalities in Primary Schooling: The Roles of Poverty and Adverse Cultural Practice*. IDS Working Paper 78.
- Duze C.O. (2010). *Average Distance Travelled to School by Primary and Secondary School Students in Nigeria and Its Effect on Attendance. An International Multi-Disciplinary Journal, Etiophia Vol.4 (4), Serial No.17, October, 2010 ISSN 2070-0083*.
- Hox, Joop J. (2010). *Multilevel Analysis: Techniques and Applications (Second Edition)*. New York: Routledge.
- Huisman, J. & Smits J. (2008). *Effects of Household and District-Level Factors on Primary School Enrollment in 30 Developing Countries. World Development Vol. 37, No. 1, pp. 179–193, 2009*.
- Huisman, J. & Smits, J. (2014). *Keeping children in school: effects of household and context characteristics on school dropout in 363 districts of 30 developing countries*. NiCE Working Paper,09:(105).
- Kementerian Dalam Negeri Republik Indonesia. (2015). *Buku Induk Kode dan Data Wilayah Administrasi Pemerintahan per Provinsi, Kabupaten/Kota dan Kecamatan Seluruh Indonesia*. Jakarta: Kemendagri.
- Kementerian Pendidikan dan Kebudayaan. (2016). *Statistik Sekolah Luar Biasa (SLB) 2015/2016*. Jakarta: Kemendikbud.
- Lamichhane, Kamal & Kawakatsu, Yoshito. (2015). *Disability and determinants of schooling: A case from Bangladesh. International Journal of Educational Development 40 (2015) 98-105*.
- Lamichhane, Kamal. (2015). *Disability, Education, and Employment in Developing Countries: From Charity to Investment*. India: Cambridge University Press.
- Nezlek, John B. (2008). *An Introduction to Multilevel Modeling for Social and Personality Psychology. Social and Personality Psychology Compass 2/2: 842-860*.
- Sugiarto. (2015). *Kesempatan Sekolah Anak dengan Disabilitas dan Faktor-Faktor yang Mempengaruhinya. Prosiding Simposium Nasional Riset Pendidikan II ISBN: 978-602-7807-58-7*.
- Undang-Undang Republik Indonesia Nomor 8 Tahun 2016 tentang Penyandang Disabilitas.
- Weiringo, Richard. (2015). *A Case Study of the Experiences of Students with Disabilities Who Did Not Complete High School*. [Disertasi]. Lynchburg: Liberty University.
- World Health Organization. (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO.
- World Health Organization. (2011). *World Report on Disability*. Geneva: WHO.

Lampiran 1. Persentase anak penyandang disabilitas usia 7-17 tahun yang masih bersekolah antarprovinsi di Indonesia tahun 2012



Sumber: Kemenkes & BPS, Publikasi Profil Penyandang Masalah Kesejahteraan Sosial Indonesia Tahun 2012

Lampiran 2. Kerangka pikir penelitian



Keterangan :

- > : Menunjukkan pengaruh
- - - -> : Menunjukkan hierarki

Lampiran 3. *Output* bilevel regresi logistik untuk partisipasi sekolah anak penyandang disabilitas tanpa variabel penjelas (Null Model)

```
. meqrlogit PartisipasiSekolahAnak || Provinsi:
```

```
Refining starting values:
```

```
Iteration 0: log likelihood = -6460.7752 (not concave)
Iteration 1: log likelihood = -6440.8821
Iteration 2: log likelihood = -6435.9487
```

```
Performing gradient-based optimization:
```

```
Iteration 0: log likelihood = -6435.9487
Iteration 1: log likelihood = -6433.6089
Iteration 2: log likelihood = -6433.4721
Iteration 3: log likelihood = -6433.4721
```

```
Mixed-effects logistic regression
Group variable: Provinsi
```

```
Number of obs = 10,267
Number of groups = 34
```

```
Obs per group:
```

```
min = 44
avg = 302.0
max = 895
```

```
Integration points = 7
Log likelihood = -6433.4721
```

```
Wald chi2(0) = .
Prob > chi2 = .
```

Partisipas~k	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	.7512836	.0488492	15.38	0.000	.6555409	.8470263

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Provinsi: Identity var(_cons)	.0583484	.0190616	.0307576	.1106891

```
LR test vs. logistic model: chibar2(01) = 96.31 Prob >= chibar2 = 0.0000
```

```
. estat icc
```

```
Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf. Interval]	
Provinsi	.0174267	.0055939	.0092626	.0325503

HIERARCHICAL BAYES SPASIAL UNTUK ESTIMASI PENGELUARAN PERKAPITA LEVEL KECAMATAN

Dwi Asih Septi Wahyuni¹

Badan Pusat Statistik Kabupaten Tapanuli Tengah
e-mail: ¹asih.septi@bps.go.id

Abstrak

Badan Pusat Statistik merilis angka pengeluaran perkapita setiap tahun untuk 514 kabupaten/kota di Indonesia. Angka pengeluaran perkapita diperoleh melalui Survei Sosial Ekonomi Nasional (Susenas). Susenas hanya dapat memberikan data yang akurat hingga level kabupaten/kota. Disisi lain, kebutuhan data hingga level kecamatan bahkan desa/kelurahan semakin meningkat guna memotret keberhasilan tujuan pembangunan berkelanjutan (*sustainable development goals*) setiap kabupaten/kota. Penelitian ini mengusulkan metode *small area estimation* dengan menggabungkan data survei dan sensus menggunakan model *hierarchical bayes* dengan mempertimbangkan keterkaitan antar wilayah (bobot spasial). Estimasi pengeluaran perkapita level kecamatan mampu memberikan hasil yang akurat dengan menggabungkan data pengeluaran perkapita sebagai variabel dependen dari Susenas dan data sensus dari Potensi Desa (Podes) sebagai variabel independen menggunakan model *hierarchical bayesspasial*.

Kata kunci: *small area estimation, hierarchical bayes, spasial, pengeluaran perkapita*

Abstract

The Central Bureau of Statistics releases annual per capita expenditure figures for 514 districts / cities in Indonesia. The per capita expenditure figures are obtained through the National Socio-Economic Survey (Susenas). Susenas can only provide accurate data down to the district / city level. On the other hand, the need for data up to the sub-district level and even villages / wards is increasing in order to capture the success of the sustainable development goals of each district / city. This study proposes a small area estimation method by combining survey and census data using a hierarchical Bayes model by considering the relationship between regions (spatial weight). The estimation of expenditure per capita at the sub-district level is able to provide accurate results by combining the per capita expenditure data as the dependent variable from Susenas and census data from the Village Potential Survey (Podes) as an independent variable using a bayesspasial hierarchical model.

Keywords: *small area estimation, hierarchical bayes, spatial, expenditure per capita*

PENDAHULUAN

Capaian keberhasilan pembangunan dapat diukur melalui sebuah data. Misalnya untuk mengetahui ketepatan sasaran program pengentasan kemiskinan, pemerintah harus memiliki data jumlah penduduk miskin dari tahun ke tahun. Selain itu, pemerintah juga memerlukan data penduduk miskin yang dapat disajikan hingga level administrasi level bawah seperti kecamatan dan desa/kelurahan.

Angka pengeluaran perkapita merupakan data dasar yang digunakan untuk menghitung garis kemiskinan dan jumlah penduduk miskin. Angka pengeluaran perkapita dihasilkan dari Susenas yang dilaksanakan oleh Badan Pusat Statistik (BPS) setiap tahun. Namun sayangnya Susenas hanya mampu menghasilkan angka estimasi hingga level kabupaten/kota. Disisi lain, tuntutan ketersediaan data hingga level administratif yang lebih kecil seperti kecamatan bahkan desa/kelurahan semakin meningkat. Oleh karena itu diperlukan alternatif metode guna menghasilkan data yang akurat pada level kecamatan bahkan desa/kelurahan.

Rao dan Molina (2015) menyebutkan bahwa untuk memperoleh data hingga level area kecil (*small area*), dapat menggunakan model estimasi tidak langsung. Model tersebut menggunakan pendekatan model statistik dengan “meminjam kekuatan” dalam menghasilkan estimasi untuk area kecil dari data survey dan menggunakan data tambahan yang tersedia pada level kecil dari data sensus. Arndt, dkk (2016) mengatakan bahwa keterbatasan informasi yang tersedia pada survei dan sensus dapat diatasi melalui metode *small area estimation* (SAE). Chandra dan Kaustav (2018) telah melakukan estimasi jumlah penduduk miskin di Kota Bihar di India pada level kecamatan menggunakan data dari *Household Consumer Expenditure Survey 2011-2012 of NSSO* dan *Population Census 2011*. Hasilnya menunjukkan bahwa estimasi pada level kecamatan melalui metode SAE memberikan hasil yang akurat dan representatif.

Wahyuni (2018) dalam penelitiannya menyebutkan bahwa dalam melakukan penaksiran parameter area kecil, metode *hierarchical bayes* dengan bobot spasial memberikan akurasi yang lebih baik jika dibandingkan dengan metode *hierarchical bayes* tanpa bobot spasial. Grassetti dan Laura (2018) menyatakan bahwa korelasi antara unit spasial dapat digunakan pada level kecamatan untuk struktur model hirarki. Dalam penelitian ini, data pengeluaran perkapita level kecamatan diperoleh melalui metode SAE dari data Susenas dan Podes dengan metode *hierarchical bayes* dengan mempertimbangkan keterkaitan antar wilayah (bobot spasial).

METODE HIERARCHICAL BAYES SPASIAL

Rao dan Molina (2015) menyebutkan bahwa esensi metode area kecil adalah penggunaan data tambahan yang tersedia pada level area kecil seperti data administratif atau data sensus sebelumnya. Data tersebut digunakan untuk membangun variabel prediktor untuk digunakan pada model statistik yang dapat digunakan untuk melakukan estimasi area kecil.

Model SAE diklasifikasikan menjadi dua tipe, yang pertama model level area yang menghubungkan estimasi langsung area kecil dengan kovariat area spesifik dan yang kedua model level unit yang menghubungkan nilai unit dari variabel penelitian untuk dikaitkan dengan kovariat unit spesifik dan juga kemungkinan kovariat area spesifik (Molina, Isabel, Balgobin Nandram, dan J.N.K Rao, 2014).

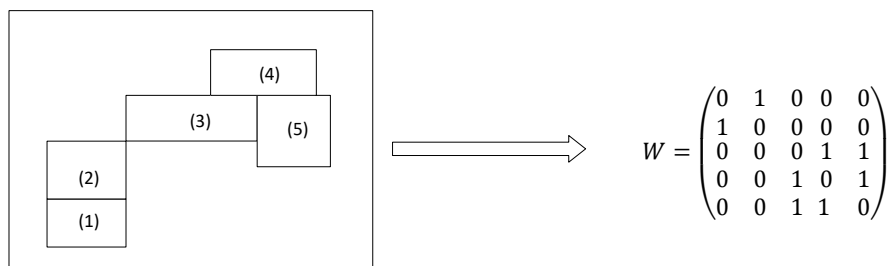
Estimasi langsung pengeluaran perkapita level kecamatan menggunakan data susenas tidak dapat dilakukan karena jumlah sampel yang hanya memberikan presisi hingga level kabupaten/kota. Marchetti dan Luca (2017) menyebutkan bahwa estimator langsung hanya digunakan pada unit area khusus, sehingga ketika jumlah sampel kecil maka estimasi langsung akan menghasilkan estimasi yang tidak dapat dipercaya serta menghasilkan varians yang besar. Oleh karena itu

diperlukan metode estimasi tidak langsung untuk menghasilkan angka pengeluaran perkapita level kecamatan dengan metode SAE.

Estimasi pengeluaran perkapita level kecamatan dalam penelitian ini diperoleh melalui metode *hierarchical bayes* spasial. Data pengeluaran perkapita diperoleh dari Susenas, namun variabel independen yang akan digunakan tidak tersedia pada data Susenas sehingga model SAE yang digunakan dalam penelitian ini menggunakan pendekatan model level area. Variabel independen yang digunakan berasal dari data Podes 2014. Penentuan variabel independen berdasarkan pada kajian literature dan penelitian-penelitian sebelumnya. Selain itu kelima variabel independen memiliki hubungan erat dengan variabel dependen yakni pengeluaran perkapita. Berikut variabel independen yang digunakan dalam penelitian ini:

1. Rasio keluarga pengguna listrik PLN per kecamatan
2. Rasio SD/MI dan SMP/MTs Negeri per 10.000 penduduk per kecamatan
3. Rasio SMA/MA dan SMK Negeri per 10.000 penduduk per kecamatan
4. Rasio tenaga medis per 10.000 penduduk per kecamatan
5. Jumlah bantuan pemerintah kabupaten per kecamatan

Bobot spasial yang digunakan dalam penelitian ini adalah *rook contiguity*. Lessage (1999) menyebutkan bahwa *rook contiguity* merupakan salah satu matriks ketergantungan spasial yang menetapkan bahwa bernilai 1 jika wilayah $-i$ berhimpit sisinya dengan wilayah $-j$ dan bernilai 0 (nol) jika tidak berhimpit sisinya dengan wilayah $-j$. Berikut ilustrasi dari *rook contiguity*:



Gambar 1. Ilustrasi *Rook Contiguity*

Pemodelan *hierarchical bayes* spasial yang digunakan dalam penelitian ini menggunakan model spasial yang diterapkan oleh Rao dan Molina (2015) sebagai berikut:

1. Hirarki 1

$$y_i | \theta_i, \sigma_v^2 \sim N_p(\theta_i, \psi_i) \dots\dots\dots (1)$$

Dimana ψ_i adalah matriks diagonal dengan elemen σ_i^2/n_i .

- a) y_i menunjukkan pengeluaran perkapita di kecamatan ke i .
- b) θ_i merupakan parameter dari variabel y_i yang berdistribusi normal. θ_i juga merupakan parameter sasaran yang akan diestimasi pada penelitian ini.

Model ini merupakan distribusi sampel variabel pengeluaran perkapita dari data Susenas 2017.

2. Hirarki 2

$$\theta_i | \beta, \sigma_v^2 \sim N(x_i^T \beta, \sigma_v^2 D^{-1}) \dots\dots\dots (2)$$

Dimana:

- a) $x_i = (x_{1i}, \dots, x_{pi})^T$ adalah variabel independen
- b) $\beta = (\beta_1, \dots, \beta_p)^T$ adalah vektor $p \times 1$
- c) $D = \lambda R + (1 - \lambda)I$
- d) R adalah matriks pembobot spasial berdimensi $m \times m$ dengan anggota r_{ii} adalah banyaknya tetangga area ke- i , $r_{il} = 1$, jika area i adalah tetangga area l dan 0 (nol) lainnya dan $i \neq l$, $m =$ banyaknya area kecil
- e) λ adalah parameter autokorelasi spasial, bernilai $0 \leq \lambda \leq 1$
- f) $\sigma_v^2 =$ varians antar area kecil
- g) $I =$ matriks identitas berdimensi m

Model ini menggambarkan pola hubungan antara pengeluaran perkapita dengan data pendukung dari Podes 2014

3. Distribusi Prior

Prior untuk parameter model $(\beta, \lambda, \sigma_v^2, \sigma_1^2, \dots, \sigma_m^2)$ saling bebas, $f(\beta) \propto 1$

$$\frac{1}{\sigma_v^2} \sim \text{gamma}(a, b); a \geq 0, b \geq 0$$

- a) v_i merupakan pengaruh acak yang diasumsikan mengikuti distribusi normal dengan rata-rata 0 (nol) dan varians σ_v^2 . Dalam melakukan penaksiran y_i distribusi prior konjugat yang tepat adalah *invers gamma*.
- b) Distribusi prior dari σ_v^2 yakni invers gamma. Nilai a dan b pada invers gamma dapat ditentukan pada nilai mendekati 0 (nol) sebagai bentuk tidak dimilikinya informasi awal (Zhou dan You, 2008).
- c) $f(\beta) \propto 1$ menunjukkan bahwa distribusi prior untuk β dalam model *hierarchical bayes* adalah *flat prior* yakni prior yang memiliki densitas serba sama *equally likely/uniformly distributed*. Jenis prior ini digunakan karena β hanya mempunyai nilai-nilai pada *range* yang tertentu saja.

Bukhari (2015) menyatakan bahwa dari hasil simulasi, model *hierarchical bayes* spasial lebih baik dari model *hierarchical bayes* tanpa bobot spasial karena memberikan nilai RB dan RRMSE yang lebih kecil. Oleh karena itu, dalam penentuan model terbaik dalam penelitian ini adalah dengan melihat nilai *relative bias* (RB) dan *relative root mean square error* (RRMSE) terkecil pada model *hierarchical bayes* spasial. Berikut dengan rumusan RB dan RRMSE:

$$1. RB_i = \frac{1}{k} \frac{\sum_{k=1}^K (\hat{Y}_i^{(k)} - \bar{Y}_i)}{\bar{Y}_i} \dots\dots\dots (3)$$

$$2. RRMSE = \frac{\sqrt{\frac{1}{k} \sum_{k=1}^K (\hat{Y}_i^{(k)} - \bar{Y}_i)^2}}{\bar{Y}_i} \dots\dots\dots (4)$$

Metode SAE dalam penelitian ini menggunakan model *hierarchical bayes* spasial melalui proses simulasi. Simulasi dilakukan dengan menggunakan data pengeluaran perkapita sebanyak 20 kecamatan hasil Susenas. Adapun variabel independen yang digunakan sebanyak 5 variabel berasal dari Pendataan

Podes. Simulasi dilakukan menggunakan *software* R. Data dibangkitkan dari distribusi posterior $\sigma_v^2 \sim IG(a, b)$, dimana *IG* adalah invers gamma.

Berikut tahap simulasi model *hierarchical bayes* spasial:

1. Menyiapkan data Susenas 2017 dan Podes 2014.
2. Menduga pengeluaran perkapita masing-masing kecamatan menggunakan metode estimasi langsung (*direct estimation*).
 - a. Dari data Susenas 2017, dihitung estimasi langsung pengeluaran perkapita setiap kecamatan.
 - b. Melakukan uji normalitas dengan metode Anderson Darling. Uji anderson darling digunakan karena data yang digunakan dalam penelitian adalah data kuantitatif. Selain itu uji anderson darling juga dapat digunakan untuk menguji kenormalan berbagai macam sebaran sebaran data yakni sebaran normal, lognormal, eksponensial, weibull, dan logistik.
3. Melakukan eksplorasi data untuk melihat keterkaitan antara variabel dependen pengeluaran perkapita dengan keseluruhan variabel independen.
4. Membentuk model SAE dengan metode *hierarchical bayes* spasial untuk pengeluaran perkapita setiap kecamatan ke- i dengan menggunakan informasi dari variabel penyerta
5. Menambahkan matriks ketergantungan spasial *rook contiguity*.
6. Mengolah data dan model menggunakan algoritma *Gibbs Sampling* MCMC melalui *software* R. Algoritma *Gibbs Sampling* MCMC digunakan karena mampu menghasilkan nilai posterior melalui proses penghitungan integral yang rumit dan multidimensi. Algoritma *gibbs sampling* mampu menghasilkan nilai estimasi parameter dengan algoritma yang paling sederhana.
 - a. Memeriksa konvergensi rantai markov yang dihasilkan melalui diagnosis *trace plot*, *density plot*, dan *autocorrelation plot*.

- b. Membandingkan RB dan RRMSE estimasi langsung dengan estimasi SAE *hierarchical bayes* spasial.
- c. Melakukan proses iterasi sebanyak 150.000 kali
- d. Memilih model terbaik.
- e. Memperoleh estimasi pengeluaran perkapita level kecamatan dengan metode *hierarchical bayes* spasial.

Bentuk *gibbs sampling* untuk model SAE dengan metode *hierarchical bayes* spasial sebagai berikut:

- i. $[\theta_i | \beta, \lambda, \sigma_v^2, y_i] \sim MVN[\Lambda y, (1 - \Lambda)X\beta, \Lambda E]$ dimana $\Lambda = (E^{-1} + D/\sigma_v^2)^{-1}E^{-1}$ dengan $E = \{\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_1^2\}$ dan $X = (x_1, \dots, x_m)^T$
- ii. $[\beta | \theta, \lambda, \sigma_v^2, y_i] \sim MVN[(X'DX)^{-1}X'D\theta, \sigma_v^2(X'DX)^{-1}]$
- iii. $[\lambda | \theta, \beta, \sigma_v^2] \sim [\lambda R + (1 - \lambda)I]^{\frac{1}{2}} x \exp\{-\frac{1}{2\sigma_v^2}(\eta - X\beta)'[\lambda R + (1 - \lambda)I](\theta - X\beta)\}$
- iv. $[\sigma_v^2 | \theta, \beta, \lambda] \sim IG[\frac{m}{2} + a, \frac{1}{2}(\theta - X\beta)'D(\theta - X\beta) + b]$

Dimana:

- a) MVN adalah multivariate normal
- b) $\Lambda = (\Lambda_1, \dots, \Lambda_p)$ adalah vector $p \times 1$
- c) E adalah matriks diagonal dengan elemen $\sigma_i^2 = \tilde{\sigma}_1^2$

Berdasarkan bentuk gibbs sampling tersebut, penaksiran parameter $\theta, \beta, \sigma_v^2$ dapat dibangkitkan secara langsung dari (i), (ii), dan (iv) melalui algoritma *gibbs sampling* karena ketiga parameter memiliki distribusi yang jelas yakni normal dan invers gamma.

SIMULASI MODEL

Data pengeluaran perkapita level kecamatan hasil estimasi langsung dari data Susenas dilakukan uji normalitas dengan uji Anderson Darling menggunakan software R. Hasil uji normalitas menunjukkan dengan tingkat signifikansi 5 persen, data pengeluaran perkapita berdistribusi normal dengan nilai statistik uji anderson darling sebesar 0,69472 dan *p-value* sebesar 0,05879. Oleh karena itu pemodelan *hierarchical bayes* spasial menggunakan distribusi normal.

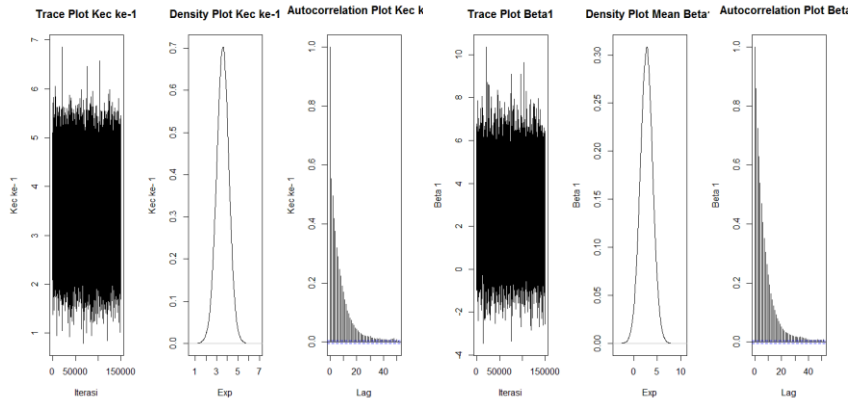
Sebelum proses simulasi model, dilakukan uji korelasi hubungan antara variabel dependen yang diperoleh dari data Susenas dan variabel independen yang diperoleh dari data Podes. Tabel 1 menunjukkan hasil korelasi pearson antara variabel Y dan X1, X2, X3, X4, X5. Hasil uji korelasi menunjukkan bahwa variabel Y

Tabel 1. *Correlations*

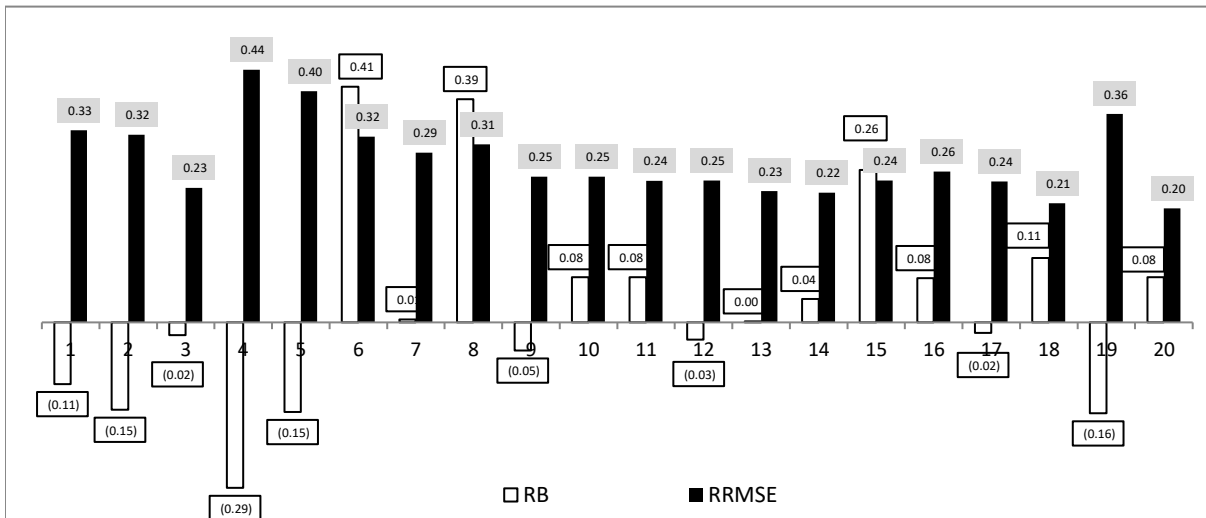
		Y	X1	X2	X3	X4	X5
Y	Pearson Correlation	1	.610**	.733**	.650**	.652**	-.279
	Sig. (2-tailed)		.004	.000	.002	.002	.233
	N	20	20	20	20	20	20
X1	Pearson Correlation	.610**	1	.437	.557*	.490*	-.615**
	Sig. (2-tailed)	.004		.054	.011	.028	.004
	N	20	20	20	20	20	20
X2	Pearson Correlation	.733**	.437	1	.701**	.524*	-.144
	Sig. (2-tailed)	.000	.054		.001	.018	.544
	N	20	20	20	20	20	20
X3	Pearson Correlation	.650**	.557*	.701**	1	.656**	-.396
	Sig. (2-tailed)	.002	.011	.001		.002	.084
	N	20	20	20	20	20	20
X4	Pearson Correlation	.652**	.490*	.524*	.656**	1	-.117
	Sig. (2-tailed)	.002	.028	.018	.002		.624
	N	20	20	20	20	20	20
X5	Pearson Correlation	-.279	-.615**	-.144	-.396	-.117	1
	Sig. (2-tailed)	.233	.004	.544	.084	.624	
	N	20	20	20	20	20	20

** . *Correlation is significant at the 0.01 level (2-tailed).*

* . *Correlation is significant at the 0.05 level (2-tailed).*



Gambar 2. Trace plot, density plot, dan autocorrelation plot



Gambar 3. Hasil RB dan RRMSE Pengeluaran Perkapita Level Kecamatan Metode Hierarchical Bayes Spasial

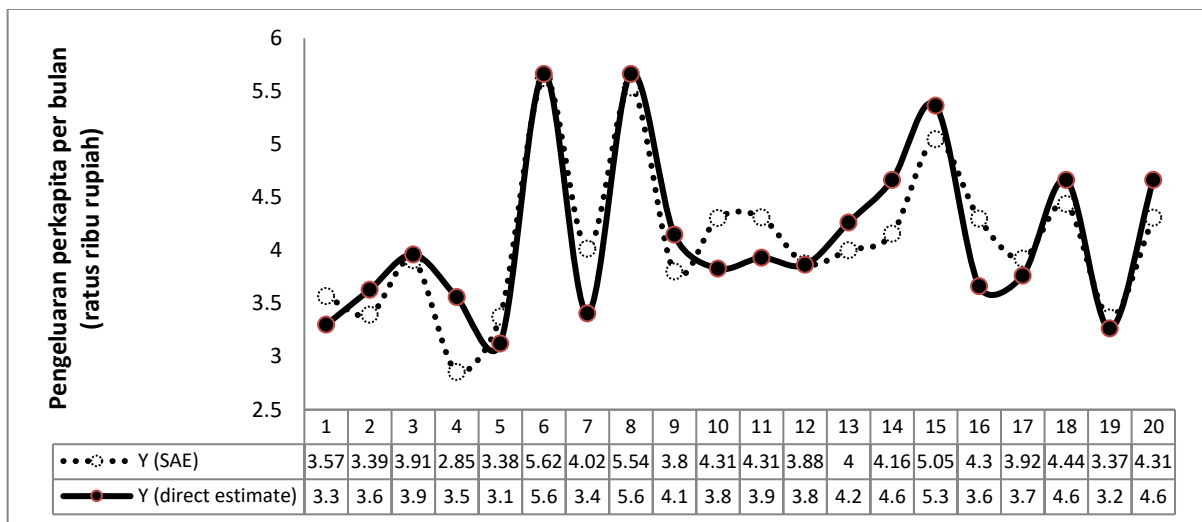
berhubungan erat dengan variabel X1, X2, X3, X4. Oleh karena itu dalam proses pemodelan variabel independen yang digunakan hanya X1, X2, X3, X4.

Karakteristik suatu model dapat diketahui melalui sebuah simulasi. Dalam penaksiran parameter pengeluaran perkapita menggunakan *hierarchical bayes* spasial digunakan inisiasi awal parameter $a = b = 0,01$, dan $\sigma_v^2 = 0,001$. Iterasi yang dilakukan dalam proses pemodelan sebanyak 15.000 kali. Proses iterasi dilakukan hingga mencapai titik stasioner. Kestasioneran dapat dilihat dari *trace plot*, *density plot*, dan *autocorrelation plot* berikut:

Gambar 2 menunjukkan bahwa proses *burn in* telah selesai pada *trace plot* sehingga distribusi posterior telah konvergen dan memberikan nilai yang cukup stabil karena sudah tidak membentuk

pola tertentu. *Density plot* (plot kepekatan peluang) menunjukkan bahwa pola distribusi dari penduga parameter yang diamati cenderung simetris. *Autocorrelation plot* menunjukkan bahwa nilai-nilai autokorelasi mendekati satu dan selanjutnya nilai-nilai berkurang menuju nol sehingga dapat dikatakan bahwa pada rantai terdapat korelasi penaksiran pengeluaran perkapita antar iterasi. Korelasi tersebut menunjukkan bahwa algoritma sudah berada di daerah distribusi target.

Setiap kali proses iterasi, diperoleh nilai baru ke $-k$ untuk parameter β, σ^2, Y_{HBS} dari kecamatan ke-1 hingga ke-20. Nilai penaksiran parameter diperoleh dari nilai rata-rata rantai markov yang stasioner. Gambar 3 menunjukkan hasil RB dan RRMSE estimasi pengeluaran perkapita



Gambar 4. Perbandingan Pengeluaran Perkapita Level Kecamatan Hasil Estimasi Langsung dan Hasil SAE *Hierarchical Bayes* Spasial

level kecamatan menggunakan *hierarchical bayes* spasial.

Gambar 3 menunjukkan perbedaan hasil RB dan RRMSE pengeluaran perkapita 20 kecamatan menggunakan metode *hierarchical bayes* spasial. Terlihat bahwa nilai RB dan RRMSE pengeluaran perkapita level kecamatan metode *hierarchical bayes* spasial memberikan nilai yang sangat kecil (kurang dari nol). Hal ini menunjukkan bahwa nilai estimasi hasil *hierarchical bayes* spasial memberikan hasil yang akurat.

Gambar 4 menunjukkan perbandingan angka pengeluaran perkapita hasil estimasi langsung dan hasil pemodelan SAE menggunakan *hierarchical bayes* spasial. Nilai estimasi yang dihasilkan melalui metode *hierarchical bayes* spasial cukup bervariasi, beberapa diantaranya lebih tinggi dibandingkan nilai estimasi langsung dan beberapa lainnya lebih rendah. Perbedaan yang terjadi menunjukkan adanya koreksi dari model *hierarchical bayes* spasial terhadap nilai estimasi langsung. Koreksi ini dilakukan karena lemahnya akurasi dari kurangnya sampel. Selain itu, pengaruh korelasi spasial mengakibatkan pola garis halus (fluktuasi kecil).

KESIMPULAN DAN SARAN

Dalam penelitian ini, Susenas 2017 dan Podes 2014 digunakan untuk

melakukan estimasi pengeluaran perkapita level kecamatan di Kabupaten Tapanuli Tengah. Susenas didesain untuk menghasilkan estimasi akurat pada level kabupaten/kota, sedangkan Podes dapat memberikan estimasi yang akurat hingga level kecamatan. Estimasi pengeluaran perkapita level kecamatan dapat dilakukan dengan memanfaatkan kelebihan data survei yakni Susenas dan data sensus yakni Podes. Metode tersebut merupakan teknik *small area estimation* sehingga diperoleh estimasi pengeluaran perkapita level kecamatan. Metode *small area estimation* dilakukan melalui pendekatan metode *hierarchical bayes* dengan mempertimbangkan keterkaitan antar wilayah (bobot spasial).

Proses pemilihan variabel independen yang digunakan dalam pemodelan sangat mempengaruhi akurasi hasil estimasi. Simulasi model *hierarchical bayes* spasial dilakukan melalui algoritma *gibbs sampling*. Pemeriksaan konvergensi rantai markov dilakukan melalui diagnosis *trace plot*, *density plot*, dan *autocorrelation plot*. Proses iterasi dilakukan sebanyak 150.000 kali.

Estimasi pengeluaran perkapita level kecamatan yang dihasilkan melalui pendekatan SAE dan metode *hierarchical bayes* spasial memberikan hasil yang akurat karena menghasilkan nilai RB dan RRMSE yang kecil dimana nilai RB yang dihasilkan berada diantara -0,29 hingga

0,41 sedangkan nilai RRMSE yang dihasilkan berada diantara 0,20 hingga 0,44. Oleh karena itu dapat dikatakan bahwa pemodelan pengeluaran perkapita level kecamatan menggunakan *hierarchical bayes* spasial memberikan akurasi data yang baik dilihat dari nilai *relative bias* dan *relative root mean square error* yang kecil.

DAFTAR PUSTAKA

- Arndt, Channing.dkk. 2016. *Poverty Mapping Based on Firsr Order Dominance With An Example From Mozambique. Journal of International Development*, (28), 3-21.
- Bukhari, Ari Shobri. 2015. Pendugaan Area Kecil Komponen Indeks Pendidikan dalam IPM di Kabupaten Indramayu dengan Metode *Hierarchical Bayes* Berbasis Spasial. Tesis Sekolah Universitas Padjadjaran Bandung. Bandung. 166 hlm.
- Chandra, Hukum dan Kaustav Aditya. 2018. *Localised Estimates and Spatial Mapping of Poverty Incidence in The State of Bihar in India-An Application of Small Area Estimation Technique. Plos one*, 13(6), 1-14.
- Grassetti, Luca dan Laura Rizzi. 2018. *The determinants of Individual Health Care Expenditure in the Italian Region of Friuli Venezia Giulia: Evidence From a Hierarchical Spatial Model Estimation. Empir Econ. Springer-Verlag GmnH Germany*
<https://doi.org/10.1007/s00181-017-1372-9>, 1-23.
- Guadarrama, Maria, Isabel Molina dan J.N.K Rao. 2015. *A Comparison of Small Area Estimation Methods for Poverty Mapping. UC3M Working Papers Statistics and Econometrics*15-05. *Universidad Carlos III De Madrid*, 1-25.
- Lesage, James P. 1999. *The Theory and Practice of Spatial Econometrics*. Department of Economics. University of Toledo.
- Marchetti, Stefano dan Luca Secondi, 2017. *Estimates of Household Consumption Expenditure at Provincial Level in Italy by Using Small Area Estimation Methods: "Real" Comparisons Using Purchasing Power Parities. Springer Science & Business Media Dordrecht (131)*, 215-234.
- Molina, Isabel, Balgobin Nandram, dan J.N.K Rao. 2014. *Small Area Estimation of General Parameters With Application to Poverty Indicators: A Hierarchical Bayes Approach. The Annals of Applied Statistics Vol 8 No. 2*, 852-885.
- Rao, JNK dan Molina (2015). *Small Area Estimation*. New Jersey. John Wiley & Sons.
- Wahyuni, Dwi Asih Septi. 2018. *Spatial Hierarchical Bayes Estimation of Mean Years of Schooling.IOP Conf. Series: Material Science and Engineering 300(1)*, 1-6.
- Zhou, Qian M dan Yong You. 2008. *Hierarchical Bayes Small Area Estimation For The Canadian Community Health Survey.Canada: SSC Annual Meeting. Proceeding of The Survey Methods Section*.

PERBANDINGAN KLASIFIKASI ANALISIS DISKRIMINAN FISHER DAN METODE NAIVE BAYES

Aliftha Ainurrochmah¹, Memi Nor Hayati², Andi M. Ade Satriya³

^{1,2,3}Universitas Mulawarman

e-mail: ¹aliftha.aaa@gmail.com, ²meminorhayati@fmipa.unmul.ac.id, ³andiade06@yahoo.com

Abstrak

Klasifikasi adalah suatu teknik pembentukan model dari data yang telah diketahui kelompok klasifikasinya. Model tersebut digunakan untuk mengklasifikasikan obyek baru. Analisis diskriminan Fisher merupakan teknik multivariat untuk memisahkan obyek-obyek dalam kelompok yang berbeda. Naive Bayes merupakan pengklasifikasian berdasarkan probabilitas dan teorema Bayes dengan asumsi independensi yang kuat. Penelitian ini bertujuan melakukan perbandingan tingkat akurasi klasifikasi dengan analisis diskriminan Fisher dan metode naive Bayes pada status pembayaran premi nasabah asuransi. Data yang digunakan memiliki 4 variabel bebas yaitu pendapatan, usia, masa pembayaran premi dan besar pembayaran premi. Hasil kesalahan akurasi dengan menggunakan nilai APER (*Apparent Rate Error*) menunjukkan bahwa metode naive Bayes memiliki tingkat akurasi yang lebih tinggi sebesar 15,38% daripada analisis diskriminan Fisher sebesar 46,15% dalam menganalisis status pembayaran premi nasabah asuransi.

Kata kunci: klasifikasi, analisis diskriminan Fisher, naive Bayes, asuransi.

Abstract

Classification is a technique to form a model of data that is already known to its classification group. The model was formed will be used to classify new objects. Fisher discriminant analysis is multivariate technique to separate objects in different groups. Naive Bayes is a classification technique based on probability and Bayes theorem with assumption of independence. This research has a goal to compare the level of classification accuracy between Fisher's discriminant analysis and Naive Bayes method on the insurance premium payment status customer. The data used four independent variables that is income, age, premium payment period and premium payment amount. The results of misclassification using the APER (Apparent Rate Error) indicate that the naive Bayes method has a higher level of accuracy is 15,38% than Fisher's discriminant analysis is 46,15% on the insurance premium payment status customer.

Keywords: classification, Fisher discriminant analysis, naive Bayes, insurance.

PENDAHULUAN

Perkembangan yang pesat di bidang ilmu dan teknologi menuntut adanya kemampuan manusia dalam mempertimbangkan segala kemungkinan sebelum mengambil keputusan dan tindakan. Salah satu pertimbangan pengambilan keputusan yang penting adalah persoalan manajemen pemilihan nasabah bagi perusahaan asuransi. Pemilihan nasabah bagi perusahaan asuransi sangat penting karena akan berdampak pada kelancaran administrasi perusahaan tersebut.

Salah satu jenis asuransi yang paling diminati masyarakat ialah asuransi jiwa. Menurut Pasal 1 Undang-Undang Nomor 40 Tahun 2014, asuransi jiwa adalah jasa penanggulangan risiko yang memberikan pembayaran kepada pemegang polis, tertanggung atau pihak lain yang berhak ketika yang tertanggung meninggal dunia atau tetap hidup dalam jangka waktu tertentu yang telah diatur dalam perjanjian. Fungsi dari asuransi jiwa sendiri yaitu memberikan jaminan tersedianya sejumlah uang tunai atau memberi penghasilan bagi keluarga (ahli waris) tertanggung jika yang tertanggung tersebut meninggal dunia.

Pada proses administrasi pengelolaan asuransi jiwa, pihak perusahaan asuransi jiwa pun tidak ingin mengalami kerugian dari nasabah. Kriteria calon pemilik polis asuransi jiwa dapat dilihat dari umur, pendapatan, status pekerjaan, masa pembayaran, besar pembayaran premi dan lain-lain (Prakoso, 1994). Berdasarkan kriteria tersebut, perusahaan dapat menentukan apakah calon nasabah tersebut termasuk kategori pembayaran premi lancar atau tidak lancar. Hal ini dilakukan untuk mengantisipasi adanya nasabah yang mengalami pembayaran premi tidak lancar. Pengelompokan nasabah dalam kategori pembayaran premi lancar atau tidak lancar dengan kriteria tertentu dapat dilakukan menggunakan analisis statistika yaitu dengan metode klasifikasi.

Terdapat beberapa macam metode pengklasifikasian yang dapat digunakan yaitu *decision tree*, *neural network*, *k-*

nearest neighbor, *naive Bayes* dan lain sebagainya. Selain metode analisis tersebut, terdapat metode analisis lain dimana pada proses pengklasifikasiannya memperhatikan asumsi tertentu yaitu analisis diskriminan (Artha, 2016).

Berdasarkan uraian tersebut, penulis tertarik melakukan penelitian ilmiah dengan judul “Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode *Naive Bayes*”.

METODE

1. Tinjauan Referensi

Penelitian mengenai analisis diskriminan Fisher dan metode *naive Bayes* telah banyak dilakukan oleh peneliti. Hal ini terkait dengan pentingnya mengetahui klasifikasi suatu obyek dengan beberapa karakteristik tertentu dan obyek tersebut sebelumnya belum diketahui klasifikasi. Banyaknya permintaan untuk menjadi nasabah asuransi menuntut adanya pertimbangan perusahaan asuransi dalam memilih nasabah. Pemilihan nasabah bagi perusahaan asuransi sangat penting karena akan berdampak pada kelancaran administrasi perusahaan tersebut. Kriteria calon pemilik polis asuransi jiwa dapat dilihat dari umur, pendapatan, status pekerjaan, masa pembayaran, besar pembayaran premi dan lain-lain.

Menurut Widarjono (2015) analisis multivariat berasal dari kata *multy* dan *variate* yang berarti alat analisis lebih dari dua variabel. Analisis multivariat berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada suatu objek. Analisis multivariat sering digunakan untuk memecahkan permasalahan dalam penelitian yang sifatnya sangat kompleks.

Menurut Towa dan Tairas (2002) klasifikasi adalah pengelompokan yang sistematis dari obyek, gagasan, buku dan benda-benda lain ke dalam kelompok atau golongan tertentu berdasarkan ciri-ciri yang sama. Klasifikasi merupakan suatu pekerjaan menilai obyek data untuk memasukkannya ke dalam kelompok tertentu dari sejumlah kelompok yang tersedia. (Prasetyo, 2014).

Tujuan dari penelitian ini adalah mengetahui bagaimana tingkat akurasi dan hasil klasifikasi status pembayaran premi nasabah asuransi PT. Prudential Life Tahun 2018 dengan status pembayaran premi lancar atau tidak lancar menggunakan analisis diskriminan Fisher dan metode *naive* Bayes. Menurut Triandaru (2008), pada dasarnya asuransi jiwa dipengaruhi oleh pendapatan, usia, masa pembayaran premi, dan besar pembayaran premi. Oleh karena itu pada penelitian ini variabel bebas yang digunakan adalah pendapatan, usia, masa pembayaran premi, dan besar pembayaran premi.

2. Metode Analisis

Data Training dan Data Testing

Data *training* digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk memprediksi kelompok dari data baru yang belum pernah ada. Data *testing* digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang ada pada data *testing* seharusnya tidak boleh ada pada data *training* sehingga dapat diketahui apakah model *classifier* sudah tepat atau belum dalam melakukan klasifikasi (Witten dan Eibe, 2011).

Jumlah data *training* dan data *testing* dapat dihitung sebagai berikut:

$$\text{Jumlah data } training = \text{proporsi data } training \times N \quad (1)$$

$$\text{Jumlah data } testing = N - \text{jumlah data } training \quad (2)$$

dengan:

N = jumlah seluruh data (data *training* + data *testing*)

Analisis Diskriminan

Analisis diskriminan adalah teknik multivariat untuk memisahkan obyek-obyek dalam himpunan yang berbeda dan mengalokasikan obyek baru dalam himpunan-himpunan tersebut. Analisis diskriminan merupakan teknik menganalisis data yang mempunyai satu variabel terikat merupakan kategori

sedangkan variabel bebas merupakan metrik.

Asumsi pada analisis diskriminan menurut Supranto (2004) adalah variabel bebas berdistribusi normal (*multivariate normality*) dan matriks varian kovarian antar kelompok adalah sama.

1. Distribusi Normal Multivariat

Menurut Johnson dan Wichern (2002), pengujian data berdistribusi normal multivariat dapat menggunakan sebuah metode yang didasarkan pada ukuran jarak kuadrat tergeneralisasi yaitu jarak mahalanobis dengan membuat Q-Q *plot*. Pengujian data berdistribusi multivariat normal dengan membuat Q-Q *plot* antara jarak mahalanobis d_j^2 dengan kuantil chi-kuadrat $q_j = \chi^2_{\left(p, \frac{(j-0,5)}{n}\right)}$. Apabila *plot*

d_j^2 dan q_j membentuk garis lurus berarti data pengamatan memenuhi asumsi distribusi normal multivariat. Selain menggunakan Q-Q *plot*, pengujian distribusi normal multivariat dapat dilakukan dengan pengujian Kolmogorov Smirnov dengan hipotesis sebagai berikut:

H_0 : Data berdistribusi normal multivariat

H_1 : Data tidak berdistribusi normal multivariat

menggunakan taraf signifikansi yaitu α dengan statistik uji sebagai berikut:

$$D_{hitung} = \text{Sup} |S(d_j^2) - F_0(d_j^2)| \quad (3)$$

dengan:

$S(d_j^2)$ = fungsi distribusi kumulatif dari

$$d_j^2$$

$F_0(d_j^2)$ = fungsi distribusi kumulatif

Chi-Kuadrat

Kriteria pengujian distribusi normal multivariat ini yaitu H_0 ditolak jika nilai D_{hitung} lebih besar dari nilai D pada tabel Kolmogorov Smirnov atau jika p -value kurang dari α .

2. Kesamaan Matriks Varians Kovarian

Menurut Johnson dan Wichern (2002), uji yang digunakan untuk mengetahui kesamaan matriks varian

kovarian adalah uji *Box's M*. Prosedur pada uji *Box's M* yaitu menggunakan pendekatan sebaran *Chi-Square* dengan derajat bebas $\frac{1}{2}(k-1)k(p+1)$ dengan p adalah banyaknya variabel bebas dan k adalah banyaknya kelompok. Hipotesis matriks varian kovarian sebagai berikut:

Hipotesis :

$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$ (matriks varian kovarian antar kelompok sama)

H_1 : sedikitnya ada sepasang matriks varian kovarian yang berbeda antar kelompok

Statistik uji :

$$C = (1-u)M$$

$$= (1-u) \sum_{i=1}^k (n_i - 1) \ln |\mathbf{S}| - \sum_{i=1}^k (n_i - 1) \ln |\mathbf{S}_i| \quad (4)$$

Jika nilai $C \geq \chi^2_{\alpha, \frac{1}{2}(k-1)p(p+1)}$ artinya

tolak H_0 pada taraf nyata α yang berarti matriks varian kovarian dari k kelompok adalah berbeda

Analisis Diskriminan Fisher

Analisis diskriminan yang menggunakan pendekatan Fisher merupakan analisis diskriminan memiliki matriks varian kovarian antar kelompok yang sama (Johnson dan Wichern, 2007).

Fungsi diskriminan Fisher dapat ditunjukkan sebagai berikut:

$$\hat{y} = \mathbf{b}^T \mathbf{X} \quad (5)$$

(1×1) (1×p) (p×1)

Pada klasifikasi menggunakan analisis diskriminan Fisher, jika fungsi diskriminan Fisher $\hat{y} = \mathbf{b}^T \mathbf{X} \geq \hat{m}$ maka obyek pengamatan akan masuk ke dalam kelompok pertama. Jika $\hat{y} = \mathbf{b}^T \mathbf{X} < \hat{m}$ maka obyek pengamatan akan masuk ke dalam kelompok kedua, dengan nilai \hat{m} sebagai berikut:

$$\hat{m} = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{S}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \quad (6)$$

Metode Naive Bayes

Teori keputusan Bayes adalah pendekatan statistik *fundamental* dalam pengenalan pola (*pattern recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan peluang yang ditimbulkan dalam keputusan-keputusan tersebut (Santosa, 2017).

Persamaan dari teorema Bayes pada umumnya adalah sebagai berikut:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (7)$$

dengan:

$P(A|B)$ = peluang terjadinya A dengan syarat B telah terjadi

$P(A)$ = peluang terjadinya A

$P(B|A)$ = peluang terjadinya B dengan syarat A telah terjadi

$P(B)$ = peluang terjadinya B

Asumsi independensi *naive* membuat syarat peluang menjadi sederhana sehingga perhitungan menjadi mungkin dilakukan. Selanjutnya penjabaran $P(Y | X_1, X_2, \dots, X_p)$ dapat disederhanakan menjadi:

$$P(Y | X_1, X_2, \dots, X_p) = P(Y) \times P(X_1 | Y) \times P(X_2 | Y) \times P(X_3 | Y) \times \dots \times P(X_p | Y) \quad (8)$$

$$= P(Y) \prod_{g=1}^p P(X_g | Y)$$

Pada umumnya teorema *naive* Bayes mudah dihitung untuk nilai pengamatan variabel bebas bertipe kategorik, namun untuk nilai pengamatan variabel bebas dengan tipe numerik (non kategorik) ada perlakuan khusus sebelum diproses menggunakan *naive* Bayes yaitu dengan cara sebagai berikut:

a Melakukan diskritisasi pada setiap nilai pengamatan variabel bebas kontinu dan mengganti nilai pengamatan tersebut dengan nilai interval diskrit.

b Mengasumsikan bentuk tertentu dari distribusi peluang untuk nilai pengamatan kontinu dan memperkirakan parameter

distribusi dengan data *training*. Distribusi Gaussian biasanya dipilih untuk memrepresentasikan peluang bersyarat pada nilai pengamatan kontinu pada sebuah kelompok $P(Y | X_g)$ adalah sebagai berikut:

$$P(X = x_g | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_g - \mu_i)^2}{2\sigma_i^2}\right) \quad (9)$$

dengan:

$i = 1, 2, \dots, k$

$g = 1, 2, \dots, p$

$x_g =$ nilai variabel bebas ke- g

$y_i =$ kelompok ke- i

$\sigma_i^2 =$ variansi kelompok ke- i

$\mu_i =$ rata-rata kelompok ke- i

3. Pengukuran Tingkat Akurasi

Evaluasi fungsi klasifikasi dilakukan dengan cara membuat tabulasi antara *actual group* dan *predicted group* yang diperoleh dari fungsi klasifikasi dari analisis diskriminan dan metode *naive Bayes* yang selanjutnya dilakukan perhitungan proporsi pengamatan yang salah klasifikasi. Diharapkan proporsi pengamatan yang salah diklasifikasikan tersebut bisa sekecil mungkin. Menghitung nilai APER dapat dilakukan melalui tabel klasifikasi sebagai berikut:

keterangan :

$f_{11} =$ jumlah obyek dari c_1 tepat

Tabel 1. Tabel Klasifikasi

Actual Membership	Predicted membership		Total
	\hat{c}_1	\hat{c}_2	
c_1	f_{11}	f_{12}	A
c_2	f_{21}	f_{22}	B
Total	C	D	E

diklasifikasikan sebagai \hat{c}_1

$f_{12} =$ jumlah obyek dari c_1 salah diklasifikasikan sebagai \hat{c}_2

$f_{21} =$ jumlah obyek dari c_2 salah

diklasifikasikan sebagai \hat{c}_1

$f_{22} =$ jumlah obyek dari c_2 tepat

diklasifikasikan sebagai \hat{c}_2

dengan perhitungan nilai APER sebagai berikut:

$$\begin{aligned} \text{APER} &= \frac{\text{jumlah obyek salah klasifikasi}}{\text{jumlah prediksi dilakukan}} \times 100\% \\ &= \frac{f_{12} + f_{21}}{f_{11} + f_{12} + f_{21} + f_{22}} \times 100\% \end{aligned} \quad (10)$$

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan data nasabah asuransi PT. Prudential *Life* Jalan MT Haryono Samarinda pada Tahun 2018. Adapun variabel terikat adalah status pembayaran premi nasabah asuransi dengan kategori lancar dan tidak lancar dengan variabel bebas adalah pendapatan, usia, masa pembayaran premi dan besar pembayaran premi.

1. Membagi Data *Training* dan Data *Testing*

Data nasabah asuransi dibagi menjadi data *training* dan data *testing* dengan proporsi pembagian 80:20. Adapun perhitungan pembagian data *training* dan data *testing* dengan menggunakan Persamaan (1) dan Persamaan (2) sebagai berikut:

$$\begin{aligned} \text{Jumlah data } \textit{training} &= \text{Proporsi data } \textit{training} \times N \\ &= 80\% \times 130 \\ &= 104 \end{aligned}$$

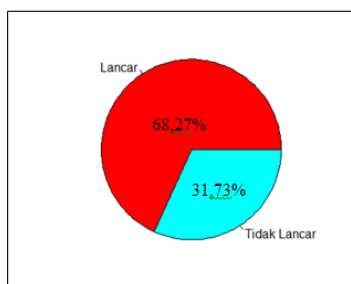
$$\begin{aligned} \text{Jumlah data } \textit{testing} &= N - \text{jumlah data } \textit{training} \\ &= 130 - 104 \\ &= 26 \end{aligned}$$

Berdasarkan perhitungan diperoleh hasil bahwa 104 data pertama hasil randomisasi data untuk masing-masing variabel penelitian berfungsi sebagai data *training*, sedangkan 26 data terakhir untuk masing-masing variabel penelitian berfungsi sebagai data *testing* yang nantinya digunakan untuk melakukan

proses klasifikasi dan pengukuran tingkat akurasi klasifikasi analisis diskriminan Fisher dan metode *naive* Bayes.

2. Analisis Statistik Deskriptif

Analisis statistika deskriptif ini dilakukan untuk mengetahui gambaran umum data nasabah asuransi PT. Prudential *Life* Tahun 2018. Data yang digunakan pada proses analisis statistika deskriptif adalah data *training*. Gambaran umum pada mengenai status pembayaran premi nasabah asuransi dengan status pembayaran premi lancar dan tidak lancar dapat dilihat pada Gambar 1.



Gambar 1. Persentase status pembayaran premi

Berdasarkan Gambar 1 menunjukkan persentase status pembayaran premi nasabah asuransi PT. Prudential *Life* Tahun 2018. Dari 104 data nasabah asuransi dapat diketahui bahwa 68,27% atau sebanyak 71 orang nasabah asuransi memiliki status pembayaran premi lancar dan 31,73% atau sebanyak 33 orang nasabah asuransi memiliki status pembayaran premi tidak lancar. Dari data tersebut dapat diketahui bahwa nasabah asuransi dengan status pembayaran premi lancar lebih banyak dibandingkan dengan nasabah asuransi dengan status pembayaran premi tidak lancar.

3. Klasifikasi Status Pembayaran Analisis Diskriminan Fisher

Analisis diskriminan adalah teknik multivariat untuk memisahkan obyek-obyek dalam kelompok berbeda dan mengalokasikan obyek baru dalam kelompok-kelompok tersebut. Adapun tahapan-tahapan dalam proses analisis diskriminan sebagai berikut:

a. Pengujian Asumsi Normal Multivariat

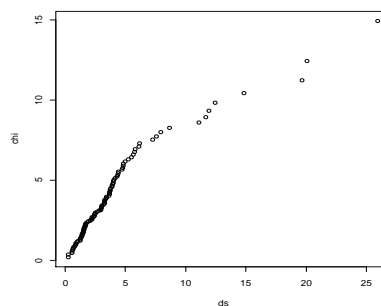
Uji normal multivariat digunakan untuk mengetahui apakah data status pembayaran premi nasabah asuransi dengan keempat variabel bebas berdistribusi normal multivariat atau tidak. Adapun pengujian hipotesis normal multivariat pada Persamaan (3) dengan taraf signifikansi sebesar 5% dengan daerah kritis H_0 ditolak jika nilai $D_{hitung} < D_{(104;0,05)}$ pada tabel Kolmogorov Smirnov atau jika $p - value < 0,05$.

Hasil pengujian menggunakan uji Kolmogorov Smirnov dapat dilihat pada Tabel 2.

Berdasarkan Tabel 2 diperoleh keputusan bahwa H_0 gagal ditolak yang artinya data nasabah asuransi dengan Tabel 2. Uji Kolmogorov Smirnov

D_{hitung}	D_{tabel}	$p - value$	Keputusan
0,125	0,133	0,3908	H_0 gagal ditolak

keempat variabel bebas berdistribusi normal multivariat.



Gambar 2. Q-Q plot uji normal multivariat

Pada Gambar 2 terlihat bahwa *plot* cenderung membentuk garis lurus, sehingga dapat disimpulkan bahwa data nasabah asuransi berdistribusi normal multivariat.

b. Pengujian Kesamaan Varians Kovarians

Uji kesamaan matriks varian kovarian antar kelompok 1 yaitu nasabah asuransi dengan status pembayaran premi lancar dan kelompok 2 yaitu nasabah asuransi dengan status pembayaran premi tidak lancar dilakukan dengan menggunakan uji *Box's M*. Adapun pengujian kesamaan matriks varian kovarian pada Persamaan (4) dengan daerah Kritis H_0 ditolak jika nilai

$C \geq \chi^2_{0,05;\frac{1}{2}(2-1)4(4+1)}$ atau jika $p\text{-value} < 0,05$.

Hasil pengujian menggunakan uji *Box's M* dapat dilihat pada Tabel 3.

Berdasarkan Tabel 3 diperoleh keputusan bahwa H_0 gagal ditolak yang artinya matriks varian kovarian antar Tabel 3. Uji *Box' M*

C	$\chi^2_{0,05;10}$	Keputusan
15,96	18,31	H_0 gagal ditolak

kelompok data nasabah asuransi adalah sama.

c. Analisis Diskriminan Fisher

Klasifikasi status pembayaran premi data nasabah asuransi menggunakan analisis diskriminan Fisher dilakukan dengan mengetahui fungsi diskriminan Fisher menggunakan data *training* terlebih dahulu. Jika fungsi diskriminan Fisher sudah diketahui, kemudian dapat dilakukan proses pengklasifikasian status pembayaran premi data nasabah asuransi menggunakan data *testing*.

Nilai koefisien dari masing-masing variabel sebagai berikut:

$$\mathbf{b} = \begin{bmatrix} 6,4458 \times 10^{-7} \\ -1,8449 \times 10^{-1} \\ 1,7133 \times 10^{-1} \\ -2,2186 \times 10^{-6} \end{bmatrix}$$

Nilai batasan kelompok \hat{m} sebesar $\hat{m} = -6,2615$ sehingga diperoleh nilai fungsi diskriminan Fisher sebagai berikut:

$$y = 6,4458 \times 10^{-7} X_1 - 1,8449 \times 10^{-1} X_2 + 1,7133 \times 10^{-1} X_3 - 2,2186 \times 10^{-6} X_4$$

Pada data *testing* pertama diperoleh nilai fungsi diskriminan Fisher sebagai berikut:

$$\begin{aligned} \hat{y}_1 &= 6,4458 \times 10^{-7} (6.000.000) \\ &\quad - 1,8449 \times 10^{-1} (40) \\ &\quad + 1,7133 \times 10^{-1} (6) \\ &\quad - 2,2186 \times 10^{-6} (3.000.000) \end{aligned}$$

$$\hat{y}_1 = -9,14$$

Sehingga dapat diketahui nilai fungsi diskriminan Fisher pada data *testing* pertama sebesar -9,14. Berdasarkan aturan

analisis diskriminan Fisher, jika $\hat{y} = \mathbf{b}^T \mathbf{X} \geq -6,2615$ maka data obyek pengamatan akan masuk ke dalam kelompok pertama yaitu nasabah asuransi memiliki status pembayaran premi lancar dan jika $\hat{y} = \mathbf{b}^T \mathbf{X} < -6,2615$ maka obyek pengamatan akan masuk ke dalam kelompok kedua yaitu nasabah asuransi memiliki status pembayaran premi tidak lancar. Pada data *testing* pertama dapat diketahui nilai $\hat{y}_1 = -9,14 < \hat{m} = -6,2615$ maka obyek pengamatan data *testing* pertama dapat diklasifikasikan ke dalam kelompok kedua yaitu nasabah asuransi memiliki status pembayaran premi tidak lancar.

4. Klasifikasi Status Pembayaran Premi Metode *Naive Bayes*

Tahapan klasifikasi metode *naive Bayes* yang pertama yaitu menghitung probabilitas awal (*prior*) pada kedua kelompok menggunakan data *training*. Adapun nilai probabilitas awal (*prior*) setiap kelompok sebagai berikut:

a. Kelompok pertama (nasabah asuransi dengan status pembayaran premi lancar)

Pada 104 data *training* yang ada, terdapat 71 nasabah asuransi yang memiliki status pembayaran premi lancar sehingga nilai probabilitas awal (*prior*) kelompok pertama sebagai berikut:

$$P(Y_1) = \frac{71}{104} = 0,6827$$

b. Kelompok kedua (nasabah asuransi dengan status pembayaran premi tidak lancar)

Pada 104 data *training* yang ada, terdapat 33 nasabah asuransi yang memiliki status pembayaran premi tidak lancar sehingga nilai probabilitas awal (*prior*) kelompok kedua yaitu sebagai berikut:

$$P(Y_2) = \frac{33}{104} = 0,3173$$

Perhitungan nilai probabilitas setiap variabel bebas pada kedua kelompok dilakukan pada data *testing* berdasarkan rata-rata dan variansi dari data *training*. Pada data *testing* pertama diketahui nasabah asuransi memiliki pendapatan (X_1) sebesar

Rp.6.000.000,00, usia (X_2) 40 tahun, masa pembayaran premi (X_3) selama 6 tahun dan besar pembayaran premi (X_4) sebesar Rp.3.000.000,00. Keempat nilai variabel bebas pada data *testing* kemudian di nilai probabilitas setiap variabel bebas pada kedua kelompok. Perhitungan nilai probabilitas setiap variabel bebas pada kedua kelompok mengasumsikan variabel bebas berdistribusi Gaussian menggunakan Persamaan (9) sebagai berikut:

a. Pendapatan (X_1)

Adapun nilai probabilitas variabel pendapatan data *testing* pertama pada setiap kelompok sebagai berikut:

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi lancar pada variabel pendapatan diketahui memiliki nilai rata-rata sebesar Rp. 10.239.437 dan nilai variansi sebesar $1,8813 \times 10^{13}$ maka probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi lancar dapat dihitung sebagai berikut,

$$P(X_1 = 6.000.000 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(1,8813 \times 10^{13})}} \times \exp\left(-\frac{(6.000.000)-(10.239.437)^2}{2(1,8813 \times 10^{13})}\right) = 5,7061 \times 10^{-8}$$

Sehingga probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi lancar sebesar $5,7061 \times 10^{-8}$.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel pendapatan diketahui memiliki nilai rata-rata sebesar Rp. 10.696.970 dan nilai variansi sebesar $1,5968 \times 10^{13}$ maka probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi tidak lancar dapat dihitung sebagai berikut,

$$P(X_1 = 6.000.000 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(1,5968 \times 10^{13})}} \times \exp\left(-\frac{(6.000.000)-(10.696.970)^2}{2(1,5968 \times 10^{13})}\right) = 5,0047 \times 10^{-8}$$

Sehingga probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi tidak lancar sebesar $5,0047 \times 10^{-8}$.

b. Usia (X_2)

Adapun nilai probabilitas variabel usia data *testing* pertama pada setiap kelompok sebagai berikut:

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi lancar pada variabel usia diketahui memiliki nilai rata-rata sebesar 37,8591 dan nilai variansi sebesar 72,6942, maka probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_2 = 40 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(72,6942)}} \times \exp\left(-\frac{(40)-(37,8591)^2}{2(72,6942)}\right) = 0,0453$$

Sehingga probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi lancar sebesar 0,0453.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel usia diketahui memiliki nilai rata-rata usia sebesar 52,1515 dan nilai variansi sebesar 52,4451, maka probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_2 = 40 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(52,4451)}} \times \exp\left(-\frac{(40)-(52,1515)^2}{2(52,4451)}\right) = 1,3483 \times 10^{-2}$$

Sehingga probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi tidak lancar sebesar $1,3483 \times 10^{-2}$.

c. Masa Pembayaran Premi (X_3)

Adapun nilai probabilitas variabel masa pembayaran premi data *testing* pertama pada setiap kelompok sebagai berikut:

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi lancar pada variabel masa pembayaran premi diketahui memiliki nilai rata-rata sebesar 8,5352 dan nilai variansi sebesar 5,0237 maka probabilitas variabel masa pembayaran premi 6 tahun dengan syarat status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_3 = 6 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(5,0237)}} \times \exp\left(-\frac{(6)-(8,5352)^2}{2(5,0237)}\right) = 0,0939$$

Sehingga probabilitas masa pembayaran premi 6 tahun dengan syarat status pembayaran premi lancar sebesar 0,0939.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel masa pembayaran premi diketahui memiliki nilai rata-rata sebesar 9,0909 dan nilai variansi sebesar 1,8352 maka probabilitas variabel masa pembayaran premi 6 tahun dengan syarat status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_3 = 6 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(1,8352)}} \times \exp\left(-\frac{(6)-(9,0909)^2}{2(1,8352)}\right) = 0,0218$$

Sehingga probabilitas variabel masa pembayaran premi 6 tahun dengan syarat status pembayaran premi tidak lancar sebesar 0,0218.

d. Besar Pembayaran Premi (X_4)

Adapun nilai probabilitas variabel besar pembayaran premi data *testing* pertama pada setiap kelompok sebagai berikut:

Berdasarkan data *training* pembayaran premi dengan status

pembayaran premi lancar pada variabel besar pembayaran premi diketahui memiliki nilai rata-rata sebesar Rp. 1.633.803 dan nilai variansi sebesar $1,5968 \times 10^{13}$ maka probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_4 = 3.000.000 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(1,5968 \times 10^{13})}} \times \exp\left(-\frac{(3.000.000)-(1.633.803)^2}{2(1,5968 \times 10^{13})}\right) = 1,7494 \times 10^{-7}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi lancar sebesar $1,7494 \times 10^{-7}$.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel besar pembayaran premi diketahui memiliki nilai rata-rata sebesar Rp. 3.969.697 dan nilai variansi sebesar $3,8428 \times 10^{12}$ maka probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi tidak lancar, dihitung sebagai berikut,

$$P(X_4 = 3.000.000 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(3,8428 \times 10^{12})}} \times \exp\left(-\frac{(3.000.000)-(3.969.697)^2}{2(3,8428 \times 10^{12})}\right) = 1,8012 \times 10^{-7}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi tidak lancar sebesar $1,8012 \times 10^{-7}$.

Adapun perhitungan perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada kedua kelompok sebagai berikut:

a. Kelompok pertama (nasabah asuransi dengan status pembayaran premi lancar)

$$P(Y_1 | X_1, X_2, X_3, X_4)$$

$$\begin{aligned}
&= P(Y_1) \times P(X_1 | Y_1) \times P(X_2 | Y_1) \times P(X_3 | Y_1) \times \\
&\quad P(X_4 | Y_1) \\
&= P(Y_1) \times P(X_1 = 6.000.000 | Y_1) \times \\
&\quad P(X_2 = 40 | Y_1) \times P(X_3 = 6 | Y_1) \times \\
&\quad P(X_4 = 3.000.000 | Y_1) \\
&= (0,6827) \times (5,7061 \times 10^{-8}) \times (0,0453) \times \\
&\quad (0,0939) \times (1,7494 \times 10^{-7}) \\
&= 2,9022 \times 10^{-17}
\end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi lancar dengan syarat pendapatan sebesar Rp. 6.000.000,00, usia 40 tahun, masa pembayaran premi selama 6 tahun dan besar pembayaran premi Rp 3.000.000,00 adalah sebesar $2,9022 \times 10^{-17}$

b. Kelompok kedua (nasabah asuransi dengan status pembayaran premi tidak lancar)

$$\begin{aligned}
&P(Y_2 | X_1, X_2, X_3, X_4) \\
&= P(Y_2) \times P(X_1 | Y_2) \times P(X_2 | Y_2) \times P(X_3 | Y_2) \times \\
&\quad P(X_4 | Y_2) \\
&= P(Y_2) \times P(X_1 = 6.000.000 | Y_2) \times \\
&\quad P(X_2 = 40 | Y_2) \times P(X_3 = 6 | Y_2) \times \\
&\quad P(X_4 = 3.000.000 | Y_2) \\
&= (0,3173) \times (5,0047 \times 10^{-8}) \times (1,3483 \times 10^{-2}) \\
&\quad (0,0218) \times (1,8012 \times 10^{-7}) \\
&= 8,4135 \times 10^{-19}
\end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi tidak lancar dengan syarat pendapatan sebesar Rp. 6.000.000,00, usia 40 tahun, masa pembayaran premi selama 6 tahun dan besar pembayaran premi Rp. 3.000.000,00 adalah sebesar $8,4135 \times 10^{-19}$

Berdasarkan perhitungan perkalian probabilitas awal dan probabilitas setiap variabel bebas (*posterior*) pada kedua kelompok dapat diketahui bahwa kelompok yang memiliki *posterior* terbesar adalah nasabah asuransi dengan kelompok status pembayaran premi lancar sebesar $2,9022 \times 10^{-17}$ dibandingkan *posterior* nasabah asuransi dengan kelompok status pembayaran premi tidak lancar sebesar $8,4135 \times 10^{-19}$, sehingga dapat disimpulkan data *testing* pertama yaitu obyek dengan

nilai-nilai variabel pendapatan sebesar Rp. 6.000.000,00, usia 40 tahun, masa pembayaran premi selama 6 tahun dan besar pembayaran premi Rp. 3.000.000,00 diklasifikasikan masuk ke dalam kelompok pertama yaitu nasabah asuransi dengan status pembayaran premi lancar.

5. Pengukuran Tingkat Akurasi Klasifikasi

Pada proses klasifikasi menggunakan analisis diskriminan Fisher, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 4. Tanda (*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan analisis diskriminan Fisher.

Tabel 4. Hasil Klasifikasi Analisis Diskriminan Fisher

Klasifikasi Awal Status Pembayaran Premi	Prediksi Klasifikasi Analisis Diskriminan Fisher		Total
	Lancar	Tidak Lancar	
Lancar	8	12*	20
Tidak Lancar	0*	6	6
Total	8	18	26

Berdasarkan Tabel 4 dapat diketahui bahwa status pembayaran premi nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan analisis diskriminan Fisher diperoleh hasil yaitu dari 20 nasabah yang memiliki status pembayaran premi lancar terdapat 8 nasabah tepat diklasifikasikan memiliki status pembayaran premi lancar dan sisanya 12 nasabah tidak tepat diklasifikasikan. Terdapat 6 nasabah yang memiliki status pembayaran premi tidak lancar serta tepat diklasifikasikan memiliki status pembayaran premi tidak lancar sehingga diperoleh nilai APER sebagai berikut:

$$\begin{aligned}
\text{APER} &= \frac{12+0}{26} \times 100\% \\
&= \frac{12}{26} \times 100\% \\
&= 46,15\%
\end{aligned}$$

Pada proses klasifikasi menggunakan metode *naive* Bayes, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 5. Tanda (*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan metode *naive* Bayes.

Nilai multikolinieritas (VIF) variabel bebas yang ada diperoleh bahwa pada variabel pendapatan (X_1) sebesar 1,1493, variabel usia (X_2) sebesar 1,1035, variabel masa pembayaran premi (X_3) sebesar 1,1897 dan variabel besar pembayaran premi (X_4) sebesar 1,0155, karena nilai VIF pada keempat variabel yang ada kurang dari 10 maka diindikasikan tidak mengalami multikolinieritas antar variabel bebas yang ada serta di peroleh standar error residual sebesar 0,3551.

Tabel 5. Hasil Klasifikasi Metode *Naive* Bayes

Klasifikasi Awal Status Pembayaran Premi	Prediksi Klasifikasi Metode <i>Naive</i> Bayes		Total
	Lancar	Tidak Lancar	
Lancar	18	2*	20
Tidak Lancar	2*	4	6
Total	20	6	26

Berdasarkan Tabel 5 dapat diketahui bahwa status pembayaran premi nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan metode *naive* Bayes diperoleh hasil yaitu dari 20 nasabah yang memiliki status pembayaran premi lancar terdapat 18 nasabah tepat diklasifikasikan memiliki status pembayaran premi lancar dan sisanya 2 nasabah tidak tepat diklasifikasikan. Pada 6 nasabah yang memiliki status pembayaran premi tidak lancar terdapat 4 nasabah tepat diklasifikasikan memiliki status pembayaran premi tidak lancar dan sisanya 2 nasabah tidak tepat diklasifikasi sehingga diperoleh nilai APER sebagai berikut:

$$\begin{aligned} \text{APER} &= \frac{2+2}{26} \times 100\% \\ &= \frac{4}{26} \times 100\% \\ &= 15,38\% \end{aligned}$$

Tabel 6. Hasil Pengukuran Tingkat Akurasi

Metode	APER
Analisis Diskriminan Fisher	46,15%
Metode <i>Naive</i> Bayes	15,38%

Pada Tabel 6 dapat diketahui bahwa pada analisis diskriminan Fisher menunjukkan kesalahan klasifikasi dalam memprediksi klasifikasi sebesar 46,15% dan pada metode *naive* Bayes menunjukkan kesalahan klasifikasi dalam memprediksi klasifikasi sebesar 15,38%. Hal ini menunjukkan bahwa tingkat akurasi pengklasifikasian metode *naive* Bayes lebih baik dibandingkan dengan analisis diskriminan Fisher.

KESIMPULAN DAN SARAN

1. Kesimpulan

1. Pengklasifikasian status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan analisis diskriminan Fisher diperoleh hasil yaitu dari 26 nasabah, terdapat 12 nasabah yang tidak tepat diklasifikasikan.
2. Pengklasifikasian status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan metode *naive* Bayes diperoleh hasil yaitu dari 26 nasabah, terdapat 4 nasabah yang tidak tepat diklasifikasikan.
3. Hasil pengukuran tingkat akurasi klasifikasi menggunakan APER status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2018 dengan analisis diskriminan Fisher diperoleh hasil kesalahan klasifikasi sebesar 46,15% dan metode *naive* Bayes sebesar 15,38%. Hal ini menunjukkan bahwa metode *naive* Bayes memberikan ketepatan prediksi klasifikasi yang lebih baik

2. Saran

Sebaiknya dalam penelitian selanjutnya dapat menggunakan metode pengklasifikasian lainnya seperti *Support Vector Machine* (SVM), *Probabilistic Neural Network* (PNN), analisis diskriminan Kernel atau metode pengklasifikasian lainnya serta dapat dikembangkan dengan data yang memiliki kategori lebih dari dua pada variabel terikat dan variabel bebas yang lebih banyak.

DAFTAR PUSTAKA

- Artha, C. A. 2016. *Perbandingan Hasil Klasifikasi Menggunakan Regresi Logistik dan Analisis Diskriminan Kuadratik Pada Kasus Pengklasifikasian Jurusan Di SMA Negeri 8 Samarinda Tahun Ajaran 2014/2015*. Jurnal Eksponensial, Vol. 7(2), 179-186.
- Johnson, R. A., dan D. W Winchern. 2002. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Prakoso, D. 1994. *Asuransi Indonesia*. Semarang: Dahara Prize.
- Prasetyo, E. 2014. *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Santosa, B. 2000. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- Towa, H. dan Tairas, J. N. B. (2002). *Pengantar Klasifikasi Persepuluhan Dewey*. Jakarta: BPK Gunung Mulya.
- Triandaru, S. dan Budisantoso, T. (2008). *Bank dan Lembaga Keuangan Lainnya*. Jakarta: Salemba Empat.
- Widarjono, A. 2015. *Analisis Multivariat Terapan dengan Program SPSS, AMOS, dan SMARTPLS Edisi Kedua*. Yogyakarta: UPM STIM YKPN.
- Witten, dkk. 2011. *Data Mining: Pratical Machine Learning Tools and Techniques, Third Edition*. Burlington: Morgan Kaufmann Publishers.

PERAMALAN TIME SERIES MENGGUNAKAN GAUSSIAN KERNEL PCA DAN AUTOREGRESSIVE

Kasiful Aprianto¹

Badan Pusat Statistik Provinsi Sulawesi Barat
e-mail: ¹apriantokasiful@gmail.com

Abstrak

Peramalan time series menjadi bagian penting dalam pengambilan keputusan karena mampu memberikan gambaran atau kejadian mendatang berupa data prediksi. Paper ini menawarkan gaussian kernel PCA dan Autoregressive (KPCA-AR) sebagai metode peramalan untuk melakukan prediksi data. Metode KPCA-AR bekerja dengan melakukan pemetaan data ke dimensi yang lebih tinggi menggunakan kernel dengan distribusi gaussian. Setelah itu dilakukan transformasi data dengan PCA agar dimensi yang dihasilkan dapat direduksi dengan varian maksimum sehingga tidak mengurangi karakteristik data secara signifikan. Data inilah yang kemudian digunakan untuk melakukan peramalan menggunakan autoregressive. Paper ini juga membandingkan beberapa metode peramalan lainnya seperti ARIMA, ANN, SVM, dan Eksponensial Smoothin. Hasil menunjukkan bahwa KPCA-AR secara umum mampu memberikan prediksi yang baik dan bisa digunakan sebagai alternative dari metode perhitungan yang ada dilihat dari kelebihan ataupun kekurangannya.

Kata kunci: *Gaussian, Kernel, PCA, Autoregressive, Forecasting*

Abstract

Time series forecasting becomes an important part in decision making because it is able to provide a picture or future event in the form of predictive data. This paper offers the Gaussian kernel PCA and Autoregressive (KPCA-AR) as a forecasting method to predict data. The KPCA-AR method works by mapping data to a higher dimension using a kernel with a gaussian distribution. After that the transformation of data with PCA is done so that the resulting dimensions can be reduced with maximum variants so as not to significantly reduce the characteristics of the data. This data is then used to forecast using autoregressive. This paper also compares several other forecasting methods such as ARIMA, ANN, SVM, and Exponential Smoothing as a comparison. The results show that KPCA-AR in general is able to give good predictions and can be used as an alternative to existing calculation methods seen from the advantages or disadvantages.

Keywords: *Gaussian, Kernel, PCA, Autoregressive, Forecasting*

PENDAHULUAN

Peramalan yaitu kegiatan memprediksi sesuatu yang akan terjadi kedepan dengan berbagai pertimbangan dan kemungkinan berdasarkan fakta yang terjadi hari ini dan sebelumnya. Dengan memprediksi kejadian mendatang menggunakan data historis yang tersedia, pemilik kepentingan mampu mendapatkan gambaran umum kedepan dan menetapkan kebijakan agar tujuan tepat sasaran. Peramalan time series telah banyak digunakan untuk beberapa kasus prediksi seperti pertanian, iklim, ekonometrik, finansial, dan komunikasi.

Beberapa teknik peramalan yang dilakukan sebelumnya seperti SVM, Neural Network, exponential smoothing, ARIMA, dan beberapa teknik lainnya telah berhasil memberikan prediksi yang baik sesuai pada percobaan yang pernah dilakukan [1][2]. Selanjutnya dilakukan sejumlah percobaan seperti penggabungan beberapa metode untuk meningkatkan akurasi peramalan. Diantaranya seperti ARIMA-ANN, dimana dilakukan penggabungan antara metode ARIMA dan Neural network [3], lalu penggabungan metode algoritma genetika dan SVM [4], dimana kedua hasil memberikan peningkatan akurasi. Pada penelitian ini penulis akan melakukan kombinasi antara Gaussian kernel, PCA, dan autoregressive, untuk selanjutnya akan disingkat dengan KPCA-AR.

KPCA-AR bertujuan untuk menghilangkan asumsi normalitas suatu data sehingga bisa diprediksi menggunakan autoregressive, dimana syarat agar bisa dilakukan peramalan dengan metode autoregressive yaitu data terdistribusi secara normal. Hal tersebut dapat diatasi dengan melakukan transformasi data dari linear menjadi nonlinear. Untuk beberapa kasus, jika suatu data tidak terpenuhi asumsi normalitas, maka harus di transformasi, berupa polynomial, log, atau berbagai bentuk lainnya. Salah satunya yaitu dengan menggunakan kernel. Kernel sendiri merupakan teknik transformasi dengan melakukan proyeksi data ke dimensi yang lebih tinggi. Teknik ini juga

digunakan pada SVM sehingga data mampu dipisah secara linear dengan dimensi yang berbeda. Kenyataannya, dimensi yang dihasilkan kernel cenderung besar dan membuat proses komputasi semakin besar. Untuk itu dilakukan pengurangan dimensi tanpa mengurangi makna atau karakteristik dari data tersebut secara signifikan. Salah satu metodenya adalah PCA. PCA sendiri merupakan teknik menyederhanakan data dengan transformasi yang mampu membentuk koordinat baru dengan varian maksimum.

Gaussian Kernel

Kernel adalah suatu cara melakukan pemetaan data dari data ruang inisial ke ruang fitur. Salah satu contoh fungsi kernel adalah kernel gaussian.

$$k(\bar{x}, \bar{x}') = \exp\left(\frac{-(\bar{x} - \bar{x}')^2}{2\sigma^2}\right)$$

dengan \bar{x} dan \bar{x}' adalah suatu data vektor yang menjelaskan sebuah titik dengan dimensi n . Terlihat bahwa $(\bar{x} - \bar{x}')$ merefleksikan jarak antar titik, dan σ merupakan suatu variabel bebas.

Misalkan $A = \{\bar{x}^1 \dots \bar{x}^m\}$ dimana dataset A terdiri dari titik ke-1 sampai titik ke- m , masing-masing berdimensi n vector, maka fungsi K dipetakan seperti berikut.

$$k(\bar{x}^i, \bar{x}^j): \mathbb{R}^n \cdot \mathbb{R}^n \rightarrow \mathbb{R}$$

Fungsi di atas merupakan kernel jika dan hanya jika matriks yang dihasilkan adalah matriks Gram yang dapat dihasilkan dari inner product. Sebagai bukti, misal fungsi kernel untuk titik i dan j , dapat dijelaskan dalam rumus berikut:

$$\begin{aligned} k(\bar{x}^i, \bar{x}^j) &= \exp\left[\frac{-(\bar{x}^i - \bar{x}^j)^2}{2\sigma^2}\right] \\ &= \exp\left[\frac{-\{(\bar{x}^{iT} \cdot \bar{x}^i) + (\bar{x}^{jT} \cdot \bar{x}^j) - 2(\bar{x}^{iT} \cdot \bar{x}^j)\}}{2\sigma^2}\right] \\ &= \exp\left[-\frac{\bar{x}^{iT} \cdot \bar{x}^i}{2\sigma^2}\right] \cdot \exp\left[-\frac{\bar{x}^{jT} \cdot \bar{x}^j}{2\sigma^2}\right] \cdot \exp\left[\frac{\bar{x}^{iT} \cdot \bar{x}^j}{\sigma^2}\right] \end{aligned}$$

Terlihat bahwa dari fungsi yang diberikan, titik i dan j berdimensi n terpetakan ke sebuah koordinat baru dari fungsi $k(x^i, x^j)$. Semua kombinasi titik dari sebuah matriks A dapat digambarkan sebagai berikut:

$$K = \begin{bmatrix} k(\bar{x}^1, \bar{x}^1) & \dots & k(\bar{x}^1, \bar{x}^n) \\ \vdots & \ddots & \vdots \\ k(\bar{x}^n, \bar{x}^1) & \dots & k(\bar{x}^n, \bar{x}^n) \end{bmatrix}$$

yang mana matriks ini bisa dihasilkan dalam bentuk *inner product*.

Principal Component Analysis

Principal Component Analysis adalah analisis yang mentransformasikan variable lama menjadi variable baru dengan mereduksi sejumlah variabel tanpa harus kehilangan sebagian besar variabel aslinya. Algoritma dari PCA dapat dijelaskan sebagai berikut:

1. Mencari rata-rata untuk setiap variabel, lalu menghitung kovarian
2. Mencari nilai eigen dan vector eigen dari matriks kovariansi empiric yang diperoleh
3. Menghitung proporsi variansi dari masing-masing PC beserta nilai akumulasi untuk q-PC pertama

Autoregressive

Autoregresi merupakan bentuk model regresi dengan mengamati nilai sekarang berdasarkan nilai-nilai sebelumnya dengan selang waktu tertentu. Peramalan autoregresi dapat ditulis sebagai berikut:

$$x_t = b_0 + b_1x_{(t-1)} + \dots + b_nx_{(t-n)}$$

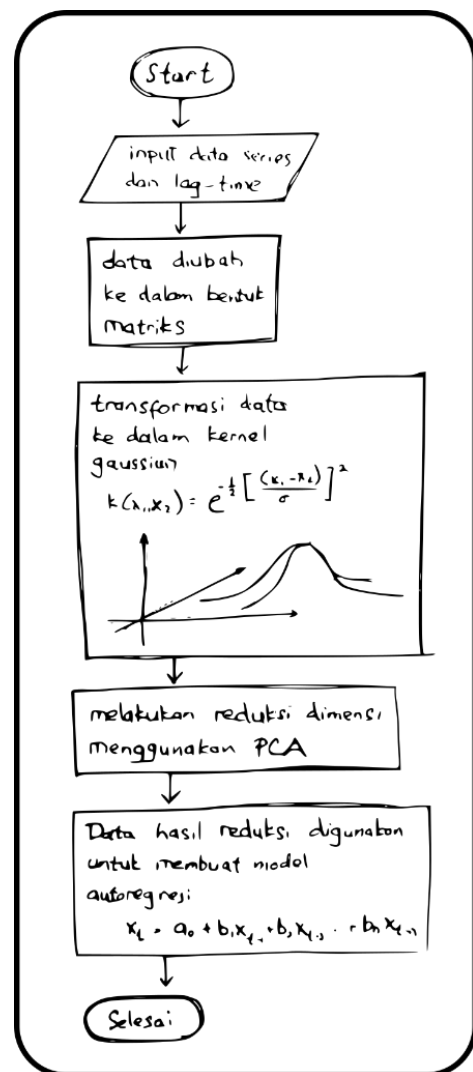
atau

$$x_t = b_0 + \sum_{i=1}^n b_nx_{(t-n)}$$

Proses yang digunakan umumnya adalah regresi linear yang variable independennya adalah data pada waktu sebelumnya.

METODE PENELITIAN

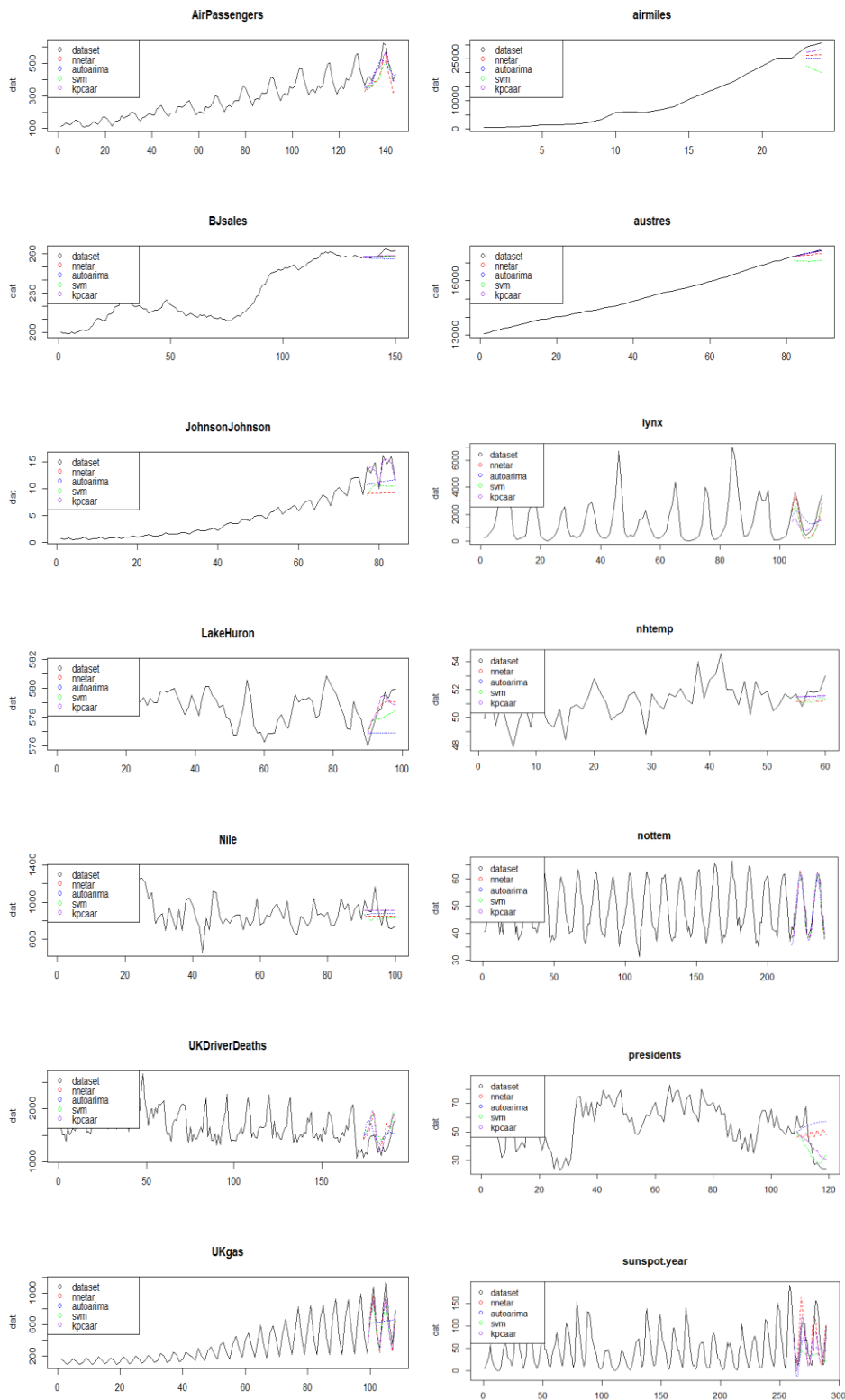
Alur kerja KPCA-AR terlihat seperti pada gambar 1. Algoritma KPCA-AR ini dilakukan dengan beberapa tahapan. Tahapan pertama, data yang akan dilakukan peramalan diubah ke dalam bentuk matriks dengan jumlah kolom sebanyak n-lag waktu. Pada peramalan autoregressive, data yang telah menjadi matriks ini kemudian dibuat sebuah model dengan metode regresi linier. Perbedaannya pada peramalan ini adalah data di transformasi terlebih dahulu ke dalam bentuk kernel gaussian.



Gambar 1. Diagram Kerangka Pikir Penelitian

Setelah data di transformasi ke kernel gaussian, dilakukan reduksi dimensi menggunakan PCA. Dilakukannya reduksi dimensi pada data yang telah di transformasi bertujuan untuk mengurangi beban komputasi pada perhitungan tanpa mengurangi nilai dari data yang ada secara signifikan.

Selanjutnya dilakukan pengujian metode KPCA-AR. Pengujian dilakukan dengan menggunakan beberapa data (<http://www.eio.upc.edu/~pau/cms/rdata/datasets.html>). Pengujian dilakukan dengan membandingkan akurasi hasil peramalan KPCAAR dengan beberapa metode yang umum digunakan seperti SVM, ANN, dan ARIMA. Parameter KPCAAR seperti lag_time dan sigma dipilih secara manual, dilakukan beberapa kali percobaan, lalu



Gambar 6. Grafik Peramalan Masing-Masing Metode

dipilih hasil yang terbaik untuk dilakukan perbandingan.

HASIL DAN PEMBAHASAN

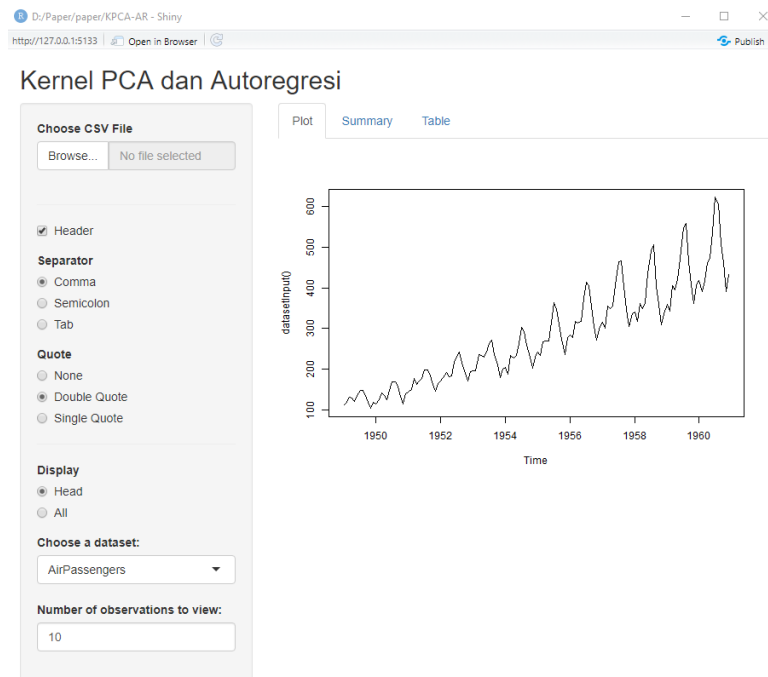
Algoritma KPCAAR dibuat dengan menggunakan Bahasa pemrograman R.

Untuk ANN dan ARIMA, menggunakan function “nnetar” dan “auto.arima” dari library “Forecast”. Sedangkan untuk SVM menggunakan function “svm” dari library “e1071”.

Dari percobaan data yang dilakukan, diperoleh hasil seperti pada Tabel 1 dan

Tabel 1. MAPE untuk Setiap Metode

Dataset	nnetar	autoarima	svm	kpcaar	lag_time
AirPassengers	15.94372321	7.441935163	10.7463359	7.864531784	12
BJsales	0.862745266	1.322658056	0.923395793	0.84982028	12
JohnsonJohnson	31.6261733	19.62258008	33.73841313	5.762246031	12
LakeHuron	0.103826264	0.312135633	0.174800607	0.107171035	3
Nile	13.77403041	13.6604378	13.74082359	13.46892396	6
UKDriverDeaths	18.63374415	18.45316595	20.98493376	18.1515543	12
Ukgas	12.36497035	53.62159405	15.59796615	12.10370445	6
airmiles	11.94842902	15.16495802	28.74914311	6.716193478	3
austres	0.523319045	0.192920899	2.373558751	0.097649517	6
lynx	33.0095407	70.94256864	39.05565862	42.13972704	9
nhtemp	1.451560646	1.13602354	1.299848029	1.043626014	3
nottem	4.126216051	4.928378424	4.02801591	3.707257135	6
presidents	51.78385659	65.73650925	20.26320618	22.51766703	12
sunspot.year	88.79173087	68.53967663	73.37327652	72.87463634	7



Gambar 3. Aplikasi KPCAAR menggunakan R

Gambar 2. Dari tabel 1, 9 dari 14 dataset yang diberikan menunjukkan bahwa KPCAAR dapat memberikan nilai MAPE yang lebih kecil dari metode peramalan lainnya.

Pada Gambar 2 dan Tabel 1 terlihat bahwa peramalan yang dihasilkan oleh KPCAAR memberikan nilai MAPE yang lebih kecil dibandingkan dengan yang lainnya, diantaranya yaitu BJsales, JohnsonJohnson, Nile, UKDriverDeaths, Ukgas, airmiles, austres, nhtemp, dan nottem.

KESIMPULAN DAN SARAN

KPCAAR berhasil dibuat menggunakan bahasa pemrograman R, sedangkan untuk algoritma lainnya menggunakan library “forecast” dan “e1071”. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa KPCAAR secara umum mampu memberikan hasil peramalan yang baik, terlihat dari MAPE yang dihasilkan lebih kecil dari beberapa metode lainnya yang digunakan untuk dataset yang

sudah diberikan. Hasil juga dipengaruhi oleh lag_time dan parameter sigma yang diberikan. Oleh karena itu, perlu adanya penelitian lanjutan untuk mengetahui nilai lag_time dan sigma yang optimal agar memberikan hasil yang semakin baik.

DAFTAR PUSTAKA

- [1] A. Luqman, A. Kusworo, W. Achmad. *Support Vector Machine untuk Prakiraan Beban Listrik Harian pada Sektor Industri, Jurnal Teknologi Elektro dan Kejuruan*, vol. 24, no. 2, 2015, pp. 23-28.
- [2] L. Tri. *Perbandingan Algoritma Neural Network (NN) dan Support Vector Machines (SVM) dalam Peramalan Penduduk Miskin di Indonesia, Jurnal Online ICT STMIK IKMI*, vol. 10, no. 1, 2013, pp. 49-57.
- [3] R. Santosha, K. Naveena, S. Subedar, S. Abhishek. *Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India, International Journal of Current Microbiology and Applied Sciences*, vol. 6, 2017, pp. 1721 – 1726.
- [4] M. Sung-Hwan, L. Jumin, H. Ingoo. *Hybrid genetic algorithm and support vector machines for bankruptcy prediction, Expert Systems with Applications*, vol. 3, 2006, pp. 652-660.

NOSQL MODEL DATA WAREHOUSE METADATA SURVEI DINAMIS

STUDI KASUS: SURVEI RUMAH TANGGA

Lutfi Rahmatuti Maghfiroh¹, Ibnu Santoso²

Politeknik Statistika STIS
e-mail: ¹lutfirm@stis.ac.id, ²ibnu@stis.ac.id

Abstrak

Suatu survei yang dilakukan antar periode cenderung mengalami perubahan metadata. Meski demikian, seluruh data dan metadata tersebut haruslah disimpan secara terintegrasi untuk keperluan data retrieval yang efisien. Tantangan yang muncul dengan perubahan metadata ini adalah ketika dilakukan suatu query untuk data lebih dari satu periode. Satu query untuk data satu periode tentu tidak efisien. Di sisi lain, satu query untuk data beberapa periode dapat menyebabkan terjadinya masalah konsistensi data karena adanya perubahan metadata. Pada penelitian ini, kami melakukan studi lebih lanjut terhadap beberapa model metadata data warehouse yang dapat mengakomodasi perubahan metadata secara dinamis seperti data warehouse tradisional, Multi Version Data Warehouse (MVDW) dan Manajemen metadata menggunakan model non-relasional seperti NoSQL. Dari hasil penelitian sebelumnya didapatkan bahwa model non-relasional pada database NoSQL menghasilkan waktu query yang lebih cepat dan penyimpanan yang lebih efisien dibandingkan model relasional, namun model tersebut belum dapat menangani model survei rumah tangga yang berbeda-beda. Penelitian ini bertujuan untuk menghasilkan model data warehouse yang dapat mengelola perubahan metadata pada survei rumah tangga yang berbeda-beda secara dinamis. Dari hasil pengujian didapatkan hasil bahwa model yang diusulkan telah dapat menangani perubahan metadata dari model survei rumah tangga yang berbeda-beda dengan waktu dan tempat penyimpanan yang tidak jauh berbeda dari hasil penelitian sebelumnya.

Kata kunci: Metadata, Data warehouse, MVDW, NoSQL

Abstract

A survey conducted between periods tends to experience changes in metadata. However, all data and metadata must be stored in an integrated manner for data retrieval efficiency. There are difficulty in query performed for data in between periods because of metadata changes. One query for one period data is certainly not efficient. On the other hand, one query for data for several periods can cause data consistency problems due to changes in metadata. In the latest study, we conduct further studies on several data warehouse metadata models that can accommodate dynamic metadata changes such as traditional data warehouse, Multi Version Data Warehouse (MVDW) and metadata management using non-relational models such as NoSQL. From the results of previous studies it was found that the non-relational model in the NoSQL database resulted in faster query time and more efficient storage than the relational model, but the model had not been able to handle different household survey models. This study aims to produce a data warehouse model that can manage changes in metadata in household surveys that vary dynamically. The results of the test show that the proposed model has been able to handle changes in metadata from household survey models that vary with time and place of storage that are not much different from the results of previous studies.

Keywords: Metadata, Data warehouse, MVDW, NoSQL

PENDAHULUAN

Salah satu tugas dari Badan Pusat Statistik (BPS) adalah mendiseminasikan data hasil survei dan sensus yang telah dilaksanakan. Diseminasi data hasil survei selain dipublikasikan dalam bentuk naskah buku cetak juga disebarluaskan dalam bentuk raw data jika diperlukan oleh pengguna data. Pengguna data BPS yang membutuhkan raw data dapat mengajukan permohonan secara online lewat aplikasi Silastik (<https://silastik.bps.go.id>) setelah sebelumnya menelusuri kuesioner, layout file, dan sampel data di <https://s.bps.go.id/layoutdata>.

Untuk mendapatkan raw data hasil survei/sensus yang utuh dan mudah dipahami, pengguna data harus mencocokkan terlebih dahulu antara kuesioner, layout file, dan sampel data untuk data periode tertentu. Setelah itu, petugas akan melakukan disclosure control untuk menghilangkan identitas individual pada raw data. Setelah itu, raw data yang telah melewati proses disclosure control dapat diserahkan kepada pengguna data.

SOP seperti ini tidak menjadi masalah jika pengguna data hanya membutuhkan raw data untuk satu periode saja. Masalah dapat muncul jika pengguna data membutuhkan data untuk multi periode atau data time series. Adanya perubahan data dan metadata untuk survei setiap periodenya menyebabkan pengguna data juga harus menelusuri kuesioner, layout file, dan sampel data untuk periode survei yang berbeda. Demikian juga petugas harus melakukan query secara terpisah untuk mendapatkan raw data survei periode berbeda karena database juga disimpan secara terpisah. Tabel 1 dan tabel 2 menunjukkan contoh kasus perbedaan data dan metadata untuk blok yang sama di survei yang sama di tahun yang berbeda.

Kasus yang pertama adalah adanya perbedaan penamaan variabel. Sebagai contoh, untuk pertanyaan umur pada periode yang satu adalah di rincian 5 disimpan dalam variabel B4AR5 sedangkan untuk periode yang lain adalah di rincian 407 disimpan dalam variabel R405. Hal ini

akan menimbulkan masalah dan kesulitan bagi pengguna dalam mencari nama variabel yang sama karena dalam dua periode survei saja variabel yang digunakan sudah berberda. Belum lagi jika ada perbedaan variabel lagi di periode yang lain.

Kasus selanjutnya adalah adanya perbedaan pengkodean responden untuk jawaban yang sama dalam periode yang berbeda. Sebagai contoh, rincian 8 untuk periode 1 untuk jawaban “Ya” dan “Tidak” adalah kode 1 dan kode 2, sedangkan pada rincian 408 periode 2 kode untuk jawaban tersebut adalah kode 1 dan kode 5. Masalah yang hampir sama dengan kasus sebelumnya akan timbul lagi pada kasus ini, namun dengan item yang berbeda.

Selain itu, ada juga kasus tidak ada kontinuitas suatu variabel. Suatu variabel mungkin tidak ditanyakan lagi di periode survei yang lain. Sebagai contoh, rincian 10 dan 11 pada periode 1 tidak ditanyakan pada survei periode 2 dan terdapat perbedaan konsep definisi pada rincian 14-15 periode 1 dengan rincian 412-413 periode 2. Perubahan metadata ini berpotensi menimbulkan masalah inefisiensi dan juga potensi kesalahan data yang diambil pada proses data retrieval.

Namun bisa saja, kasus pertanyaan yang belum muncul di periode survei sebelumnya akan muncul di periode survei yang lain. Kasus lain yang berdampak hampir sama adalah bisa saja suatu pertanyaan muncul di suatu periode survei, kemudian tidak muncul di suatu periode survei yang lain, kemudian muncul kembali di suatu periode survei.

Untuk keperluan data retrieval yang efisien, data survei dan metadatanya untuk setiap periode seharusnya disimpan secara terintegrasi dalam suatu data warehouse. Data Warehouse (DW) adalah suatu tempat penyimpanan data yang terintegrasi dan berorientasi subyek. Penggunaan data warehouse untuk menyimpan dan menyusun data berdasarkan subyek bertujuan untuk mempermudah data retrieval dan juga analisis data. Namun, penyimpanan data pada data warehouse dengan perubahan metadata masih menjadi

Tabel 1. MAPE untuk Setiap Metode

IV. A. KETERANGAN ANGGOTA RUMAH TANGGA																	
No. urut	Nama anggota rumah tangga (Tulis siapa saja yang biasanya tinggal dan makan di rula ini baik dewasa, anak-anak maupun bayi)	Hubungan dengan kepala rumah tangga	Jenis Kelamin 1. Laki-laki 2. Perempuan	Umur (Tahun)	Status kawinan	Apakah menjadi korban kejahatan dalam setahun terakhir?	Jika Kol. (7) berkode 1 sd 6, Apakah dilaporkan ke Polisi? 1. Ya 2. Tidak	Berapa kali bepergian * selama 3 bulan kalender? Jika tidak bepergian isikan "00"	Jika bepergian (Kol. (9)≠ 00) Tujuan utama bepergian terakhir	Provinsi tujuan utama bepergian yang terakhir	Apakah mempunyai akte kelahiran dari kantor catatan sipil? Boleh saya melihatnya?	Jika Kol. (12) berkode 3 atau 4 (tidak punya/TT), Apa alasan utamanya [Jawaban jangan dibacakan!]	Apakah pernah mengikuti pendidikan pra sekolah?	Jika Kol. (14) berkode 1 atau 2, jenis pendidikan pra sekolah	Apakah mengikuti pendidikan pra sekolah angkutan yang biasa digunakan untuk sekolah	Anggota ruta berumur 0 - 6 tahun	Anggota ruta berumur 3-6 tahun dan Kolom 14 berkode 1 atau 2
(1)	(2)	[Kode] (3)	(4)	(5)	[Kode] (6)	(7)	(8)	(9)	[Kode] (10)	[Kode] (11)	[Kode] (12)	(13)	(14)	[Kode] (15)	(16)	[Kode] (17)	

Tabel 2. Layout kuesioner dan metadata untuk survei x di periode 2

BLOK IV. KETERANGAN DEMOGRAFI, AKTE KELAHIRAN, DAN PENDIDIKAN ANAK USIA DINI (PAUD)												
NAMA ANGGOTA RUMAH TANGGA												
(Tuliskan terlebih dahulu semua orang yang biasanya tinggal dan makan di rumah tangga ini mulai dari kepala rumah tangga, pasangannya, anak yang belum menikah, anak yang sudah menikah, menantu, cucu, orang tua/mertua, famili lain, pembantu, dan lainnya)												
Pastikan seluruh anggota rumah tangga tercatat dan tidak ada yang terlewat												
No Urut	402	403	404	405	406	407	408	409	410	411	412	413
		APAKAH HUBUNGAN (nama) DENGAN KEPALA RUMAH TANGGA? (Kode)	APAKAH STATUS PERKAWINAN (nama)? 1. Belum kawin 2. Kawin 3. Cerai hidup 4. Cerai mati	APAKAH (nama) LAKI-LAKI ATAU PEREMPUAN? 1. Laki-laki 2. Perempuan	KAPAN (nama) DILAHIRKAN? Bulan Tahun	BERAPAKAH UMUR (nama)? (tahun) Umur harus diisi, jika ≥ 97 tahun, tulis 97	Jika status perkawinannya kawin (cek 404 berkode 2) APAKAH SUAMI/ISTRI (nama) BIASANYA TINGGAL DI RUMAH TANGGA INI? 1. Ya 5. Tidak	Jika pernah kawin (cek 404 berkode 2 s.d. 4) PADA UMUR BERAPA (nama) MELANGSUNGKAN PERKAWINAN PERTAMA?	Cek 407, untuk ART berumur 0-17 tahun MILIKI AKTE KELAHIRAN DARI KANTOR CATATAN SIPIL? BOLEH SAYA MELIHATNYA? (Kode)	Cek 407, untuk ART berumur 0-10 tahun Jika pernah/masih (412 berkode 1, 2 atau 3), APA JENIS PENDIDIKAN YANG DITERIMA? (Kode)		

tantangan dalam proses data retrieval yang efisien.

Penelitian ini bertujuan untuk menganalisis model data warehouse yang

cocok untuk metadata survei rumah tangga dinamis dan melakukan ujicoba terhadap model yang dihasilkan dari sisi validitas, kecepatan, dan kebutuhan tempat

penyimpanan. Tujuan praktis dari penelitian ini adalah supaya nantinya pengguna data tidak perlu repot untuk mencocokkan kuesioner, layout file, dan sampel data untuk data survei yang berbeda periode. Demikian juga untuk petugas nantinya tidak perlu melakukan query secara terpisah untuk mendapatkan data survei yang berbeda periode. Cukup satu query saja untuk mendapatkan data yang diperlukan meskipun berbeda periode.

Sebagaimana dijelaskan pada penelitian sebelumnya pada penelitian Maghfiroh & Baskara (2017), untuk menangani perubahan metadata dalam data warehouse, Kimbal & Ross (2013) memperkenalkan konsep slowly changing dimensions (SCD) untuk melacak perubahan. Namun, teknik ini masih memiliki keterbatasan yang hanya dapat menangani perubahan pada atribut struktur anggota pada model multidimensional. Teknik ini tidak dapat menangani perubahan pada struktur atribut non-anggota seperti menambahkan atribut, menghapus atribut. Body dkk (2002) memperkenalkan model multidimensional temporal untuk menangani perubahan struktur Data Warehouse. Namun, pendekatan ini terbatas hanya menggunakan satu tabel fakta untuk menyimpan semua versi data. Hal ini membatasi perubahan skema dan hanya mendukung perubahan struktur pada tabel dimensi.

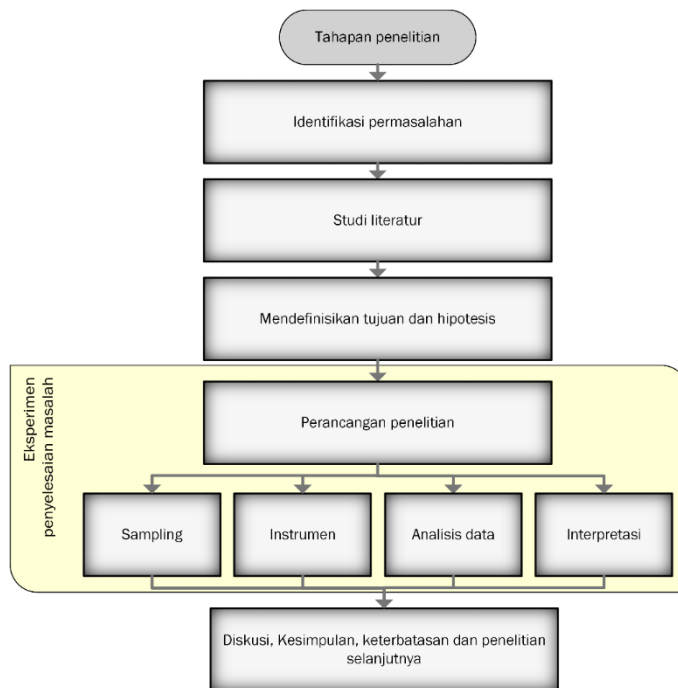
Bebel dkk (2006) menangani perubahan data warehouse menggunakan multiversion data warehouse (MVDW) dengan menyediakan format model untuk membuat MVDW. Penelitian ini dilanjutkan oleh Wrembel dan Morzy (2006), yang menambahkan penanganan definisi dan melakukan query di MVDW. Penelitian lainnya oleh Wrembel dan Morzy (2006) mengusulkan query pada MVDW dan menggunakan MVDW untuk menangani metadata pada penelitian Wrembel dan Babel (2007). Model yang diusulkan diimplementasikan menggunakan model relasional pada Oracle PL/SQL.

Kelemahan dari model relasional adalah adanya asumsi bahwa data telah terstruktur dan disimpan dalam bentuk baris dan kolom tabel. Tabel memiliki jumlah kolom yang sama dan tipe data yang sama pula (Parker dkk, 2013). Model ini belum dapat menangani data survei dengan berbagai struktur dengan perubahan variabel. NoSQL (*Not Only SQL*) dapat digunakan untuk mengatasi masalah ini karena model NoSQL fleksibel. Menurut Fitri (2013), Database NoSQL juga menyediakan mekanisme yang lebih mudah dan sederhana untuk menyimpan dan mengambil data dibandingkan database relasional. Beberapa penelitian lanjutan yang dilakukan oleh Chevalier dkk mengusulkan implementasi Data warehouse dan model multidimensional menggunakan NoSQL. Namun, penelitian tersebut belum dapat menangani perubahan metadata. Selain itu tidak terdapat cara menangani perubahan konsep dan definisi variabel yang ada dalam survei.

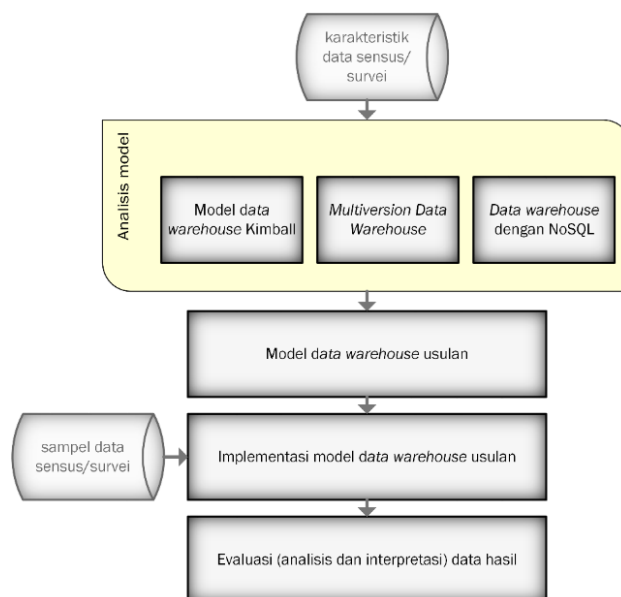
Maghfiroh & Baskara (2017) sudah membangun model DW dengan NoSQL untuk dapat menangani perubahan data dan metadata survei. Namun ternyata model yang diusulkan masih belum dinamis atau belum bisa menangani perubahan data dan metadata yang terjadi pada survei dengan model yang berbeda, baru hanya satu model survei saja.

METODE PENELITIAN

Tahapan penelitian ini dimulai dari identifikasi permasalahan, kemudian melakukan studi literatur, mendefinisikan tujuan dan hipotesis. Desain penelitian yang digunakan adalah experimental design yang dalam hal ini terbagi dalam pemilihan kasus data, instrumen yang digunakan, analisis data dan interpretasi sebagaimana ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

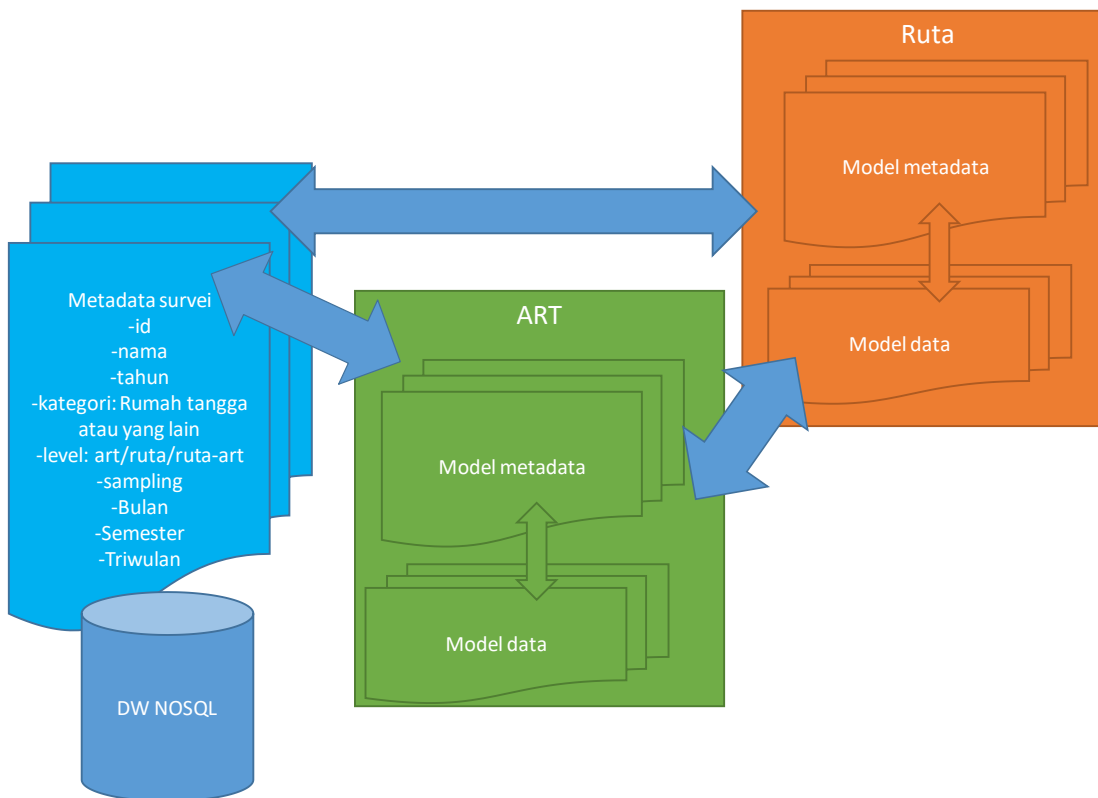


Gambar 2. Tahapan Eksperimental Desain

Dalam *experimental design* ini, tahapan yang dilakukan adalah mempelajari karakteristik data sensus/survei, kemudian melakukan analisis model untuk tiga model utama Data warehouse yang ada yaitu Model Data warehouse Kimball, *Multiversion Data Warehouse*, dan Data warehouse dengan NoSQL. Dari ketiga model tersebut kami telaah mana yang kiranya paling sesuai sebagai dasar usulan model data warehouse. Kemudian model data warehouse usulan diimplementasikan

pada sampel data dan metadata survei kemudian dilakukan evaluasi (analisis dan interpretasi) hasilnya. Hasil evaluasi terhadap model usulan akan dibandingkan dengan eksperimen yang sama pada model sebelumnya. Dengan demikian kami dapat mengevaluasi model yang kami usulkan. Proses dalam tahap *experimental design* ini ditunjukkan pada gambar 2.

HASIL DAN PEMBAHASAN



Gambar 3. Metadata Survei, Metadata Ruta, Metadata ART, dan DB NoSQL

1. Model Data Warehouse

Untuk menangani perubahan metadata dan data yang ada, kami merancang model DW dengan NoSQL berbasis dokumen. Model yang kami rancang terdiri dari 3 Bagian. Bagian yang pertama menyimpan informasi tentang survei, yang tersimpan dalam Koleksi Metadata Survei.

Bagian kedua menyimpan informasi tentang rumah tangga, yang tersimpan dalam Koleksi Metadata Rumah Tangga dan Koleksi Data Rumah Tangga. Koleksi Metadata Rumah Tangga menyimpan metadata responden rumah tangga dan perubahan yang terjadi. Koleksi Data Rumah Tangga menyimpan data responden rumah tangga dan perubahan yang terjadi.

Bagian ketiga menyimpan informasi tentang anggota rumah tangga, yang tersimpan dalam Koleksi Metadata Anggota Rumah Tangga dan Koleksi Data Anggota Rumah Tangga. Koleksi Metadata Anggota Rumah Tangga menyimpan metadata responden anggota rumah tangga dan perubahan yang terjadi. Koleksi Data Anggota Rumah Tangga menyimpan data

responden anggota rumah tangga dan perubahan yang terjadi.

Jika survei hanya terdiri dari data rumah tangga (level: ruta), maka metadata dan data yang terisi hanya untuk Koleksi Metadata Survei, Koleksi Metadata Rumah Tangga dan Koleksi Data Rumah Tangga. Jika survei hanya terdiri dari data anggota rumah tangga (level: art), maka metadata dan data yang terisi hanya untuk Koleksi Metadata Survei, Koleksi Metadata Anggota Rumah Tangga dan Koleksi Data Anggota Rumah Tangga. Jika survei terdiri dari data rumah tangga dan anggota rumah tangga (level: ruta-art), maka semua koleksi metadata dan data harus terisi. Gambar 3 mengilustrasikan hubungan antara metadata survei, metadata rumah tangga, dan metadata anggota rumah tangga dalam database NoSQL.

2. Metadata Survei

Metadata survei menyimpan informasi tentang:

- id
- nama
- tahun
- kategori: Rumah tangga atau yang lain

- level: art/ruta/ruta-art
- sampling
- Bulan
- Semester
- Triwulan

3. Model Data Rumah Tangga

Model Data Rumah Tangga menyimpan hasil respon survei untuk rumah tangga. Kode pertanyaan yang ada akan dijadikan kolom dan nilai jawaban akan menjadi isi dari kolom tersebut. Model Data Rumah Tangga dan Model Metadata Rumah Tangga akan terhubung dengan kolom id_survei. Gambar 4 menunjukkan model data Rumah Tangga yang digunakan.

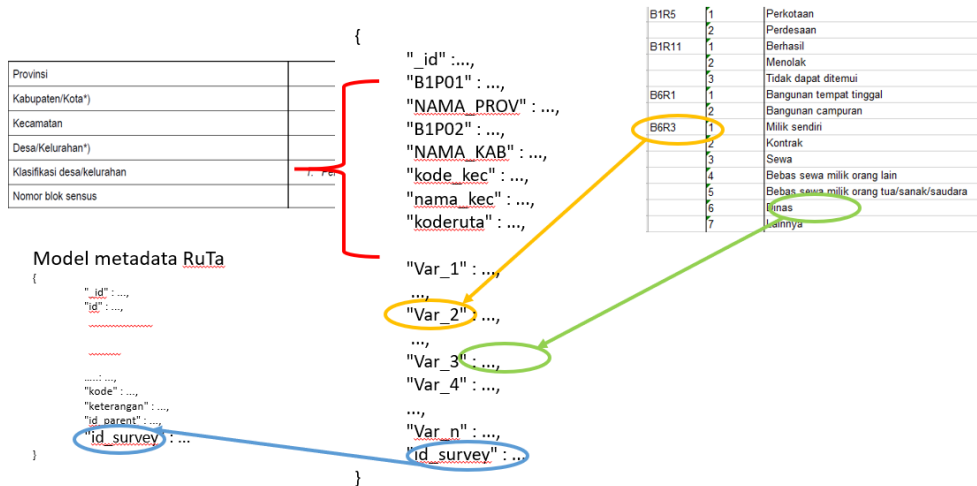
4. Model Metadata Rumah Tangga

Model Metadata Rumah Tangga menyimpan informasi metadata yang ada dan perubahan metadata yang terjadi antar

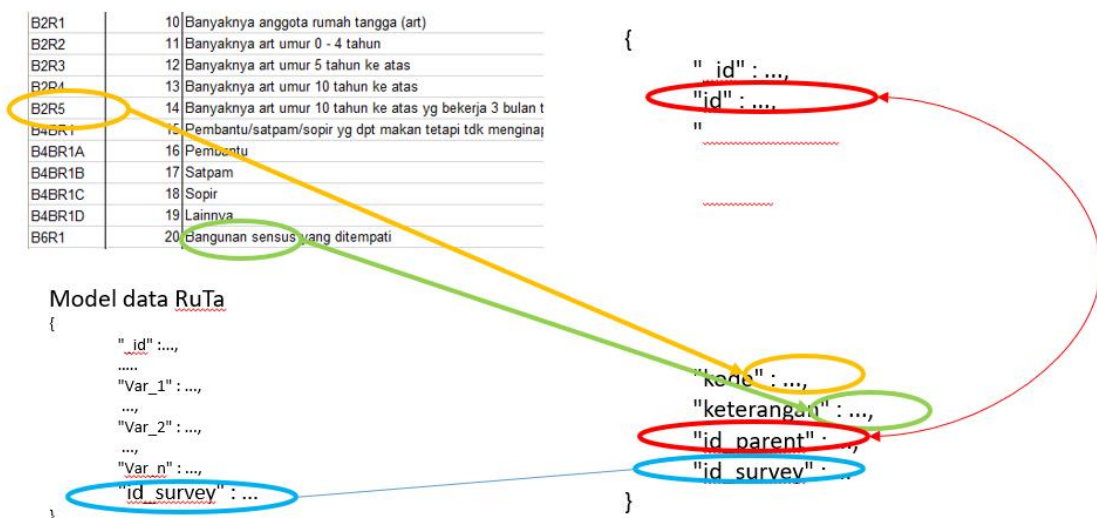
survei yang sama pada periode yang berbeda. Informasi yang disimpan di Model ini adalah kode variabel, keterangan dan kode induk dari kode tersebut di periode survei sebelumnya. Model Data Rumah Tangga dan Model Metadata Rumah Tangga akan terhubung dengan kolom id_survei. Gambar 5 menunjukkan model metadata rumah tangga.

5. Model Data Anggota Rumah Tangga

Model Data Anggota Rumah Tangga menyimpan hasil respon survei untuk anggota rumah tangga. Kode pertanyaan yang ada akan dijadikan kolom dan nilai jawaban akan menjadi isi dari kolom tersebut. Model Data Anggota Rumah Tangga dan Model Metadata Anggota Rumah Tangga akan terhubung dengan kolom id_survei. Di model ini juga terdapat kode ruta yang menghubungkan Model



Gambar 4. Model Data Rumah Tangga



Gambar 5. Model Metadata Rumah Tangga

Data Anggota Rumah Tangga dan Model Data Rumah Tangga jika ada sebagaimana ditunjukkan pada gambar 6.

6. Model Metadata ART

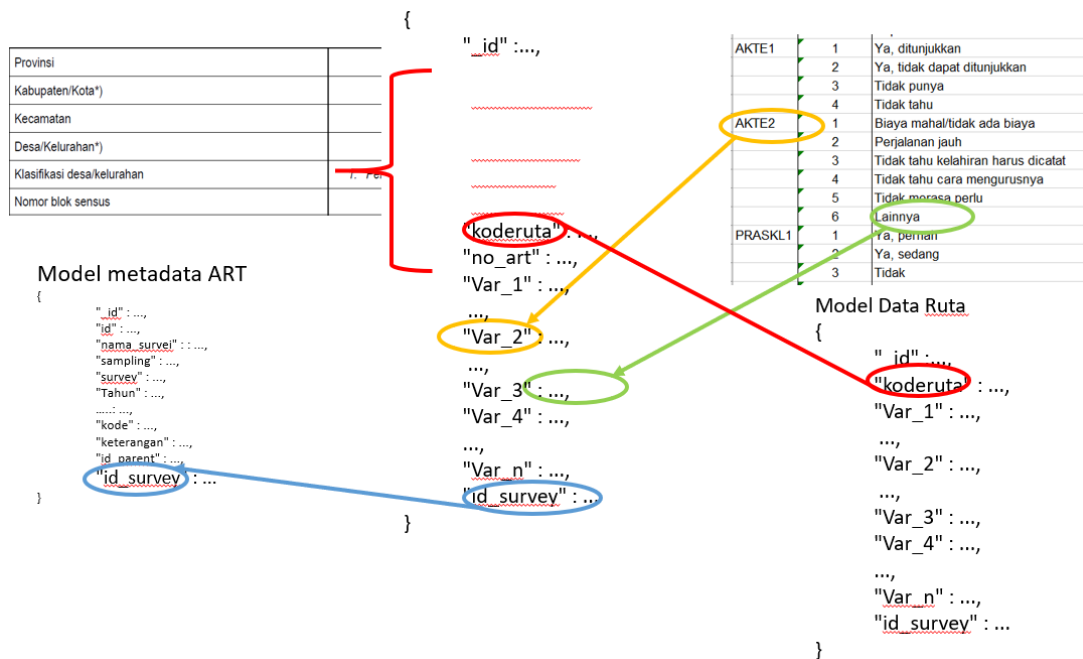
Model Metadata Anggota Rumah Tangga menyimpan informasi metadata yang ada dan perubahan metadata yang terjadi antar survei yang sama pada periode yang berbeda. Informasi yang disimpan di Model ini adalah kode variabel, keterangan dan kode induk dari kode tersebut di periode survei sebelumnya. Model Data Anggota Rumah Tangga dan Model Metadata Anggota Rumah Tangga akan terhubung

dengan kolom id_survei sebagaimana ditunjukkan pada gambar 7.

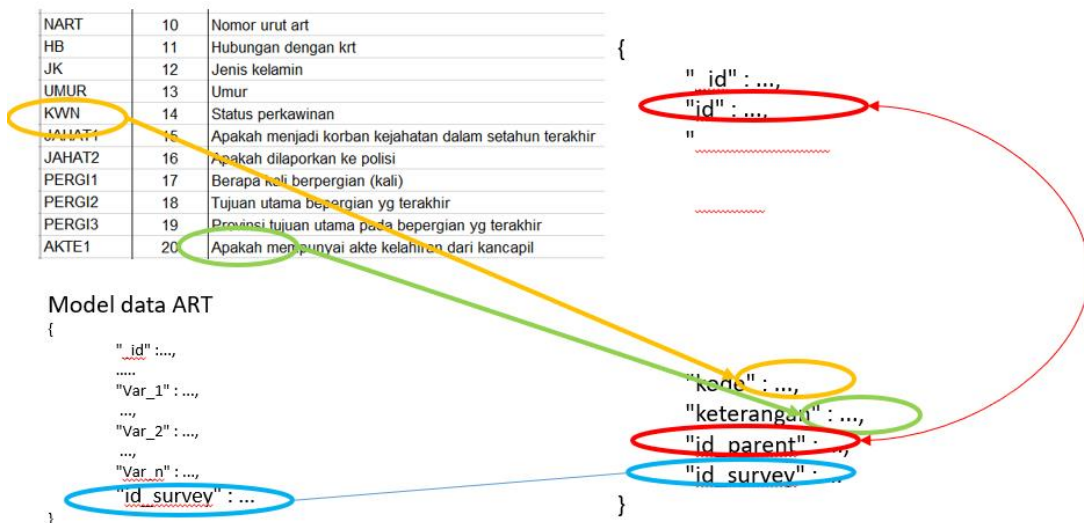
7. Algoritma

Untuk mendapatkan data yang ada dengan perubahan metadata yang ada, kami merancang algoritma. Input yang digunakan adalah query yang mengambil data anggota rumah tangga (art), rumah tangga (ruta) atau keduanya.

Jika query hanya mengambil data ruta maka daftar nama atribut akan di ekstrak dari query. Dari daftar nama atribut yang ada, kemudian kami cek koleksi metadata ruta untuk melihat perubahan metadata



Gambar 6. Model Data Anggot Rumah Tangga



Gambar 7. Model Metadata Anggot Rumah Tangga

yang ada. Dari metadata, kami mendapatkan daftar nama atribut yang baru yang dapat kita gunakan di query. Selanjutnya dari query dengan nama atribut yang baru tersebut kami ambil data ke koleksi data ruta untuk mendapatkan data yang diperlukan.

Jika query hanya mengambil data art maka daftar nama atribut akan di ekstrak dari query. Dari daftar nama atribut yang ada, kemudian kami cek koleksi metadata art untuk melihat perubahan metadata yang ada. Dari metadata, kita mendapatkan daftar nama atribut yang baru yang dapat kita gunakan di query. Selanjutnya dari query dengan nama atribut yang baru tersebut kita ambil data ke koleksi data art untuk mendapatkan data yang diperlukan.

Jika query mengambil data ruta dan art, setelah hasil query didapatkan, kami juga mengambil kode ruta dari koleksi data ruta untuk digunakan dalam query yang dijalankan dalam koleksi data art. Dengan demikian query akan mendapatkan data ruta dan art. Alur yang digunakan ditunjukkan pada gambar 8. Algoritma usulan untuk mendapatkan data dengan model yang diusulkan dapat dilihat di gambar 10. Algoritma ini lebih kompleks dibandingkan dengan algoritma pada model sebelumnya. Algoritma dari model sebelumnya dapat dilihat pada gambar 9.

Untuk mengimplementasikan model yang sebelumnya dan yang kami buat, kami menggunakan MongoDB versi 3.2. Untuk mengimplementasikan algoritma, kami menggunakan Bahasa pemrograman java dengan Netbeans 8.2. skema pengujian yang dilakukan dapat dilihat pada gambar 11.

8. Uji Coba

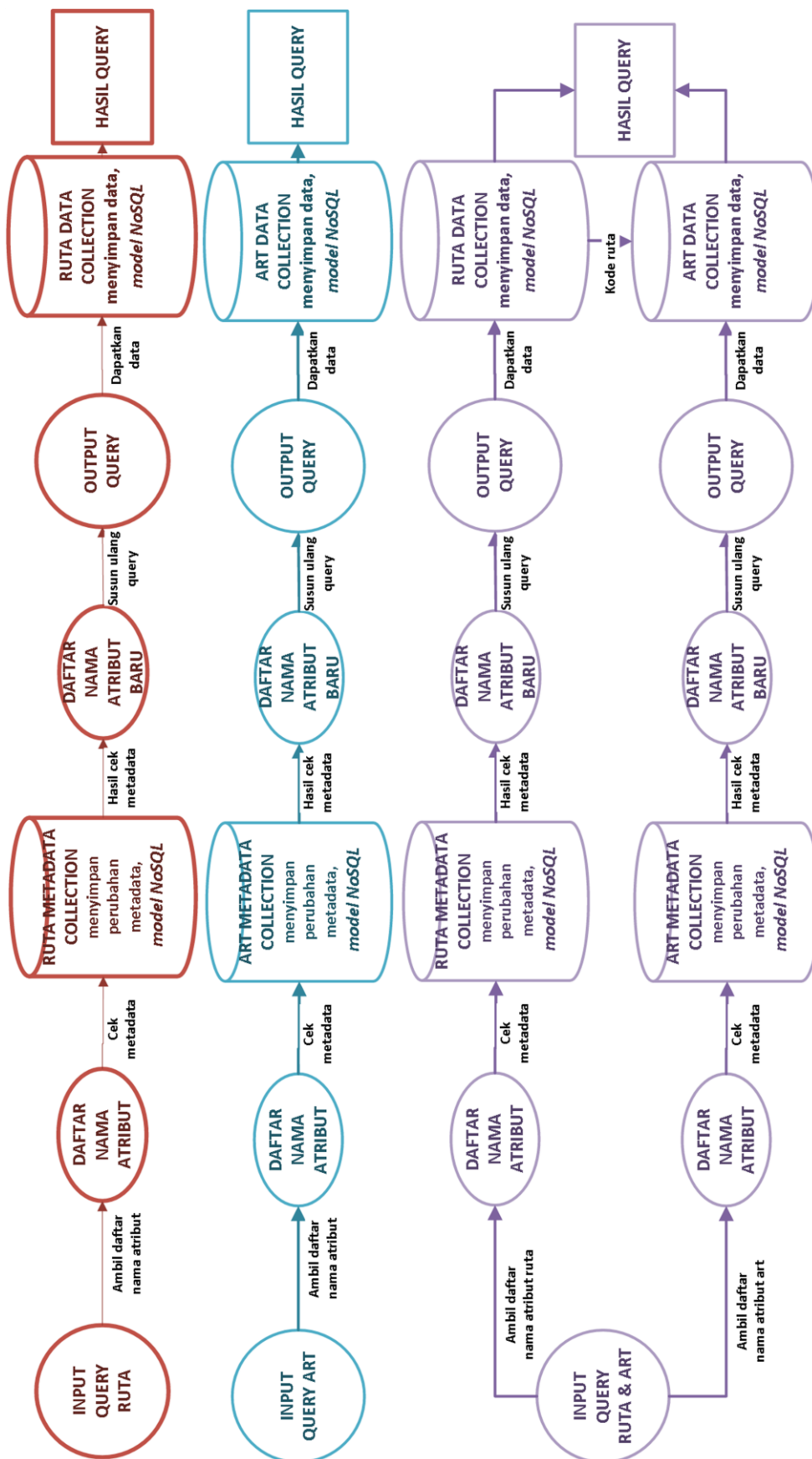
Untuk melakukan validasi terhadap model yang kami rancang, maka kami melakukan eksperimen pengujian terhadap model yang kami rancang dan model sebelumnya. Kami melakukan eksperimen dengan lingkungan perangkat lunak sistem operasi Windows 10 64 bit dan perangkat keras prosesor Inter ® Core ™ i7-8550U CPU @1.88GHz 1.99GHz dengan RAM

8GB dan penyimpanan 128TB SSD+1TB HDD.

Kami menjalankan 11 kasus query yang harus mendapatkan hasil yang sama antara model yang sebelumnya dengan model yang kami buat. Kami menggunakan data Survei Angkatan Kerja Nasional tahun 2010, 2012 dan 2015. Kami melakukan iterasi sebanyak 1000 kali terhadap algoritma yang dijalankan untuk setiap query.

- Kasus 1: Menampilkan kode hubungan dengan KRT dan jenis kelamin
- Kasus 2: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan jenis kelamin
- Kasus 3: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan umur untuk penduduk laki-laki
- Kasus 5: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan pendidikan terakhir
- Kasus 6: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan pendidikan terakhir untuk penduduk laki-laki
- Kasus 7: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan pendidikan terakhir untuk penduduk laki-laki diurutkan berdasarkan kode kabupaten/kota
- Kasus 8: Menampilkan jumlah penduduk dikelompokkan berdasarkan kabupaten/kota
- Kasus 9: Menampilkan jumlah penduduk dikelompokkan berdasarkan kabupaten/kota dan jenis kelamin
- Kasus 10: Menampilkan jumlah penduduk dikelompokkan berdasarkan kabupaten/kota dan jenis kelaminnya untuk penduduk yang pendidikan terakhirnya S1
- Kasus 11: Menampilkan jumlah penduduk dikelompokkan jenis kelamin dan pendidikannya

Hasil dari eksperimen tersebut kami bandingkan hasilnya untuk mengevaluasi

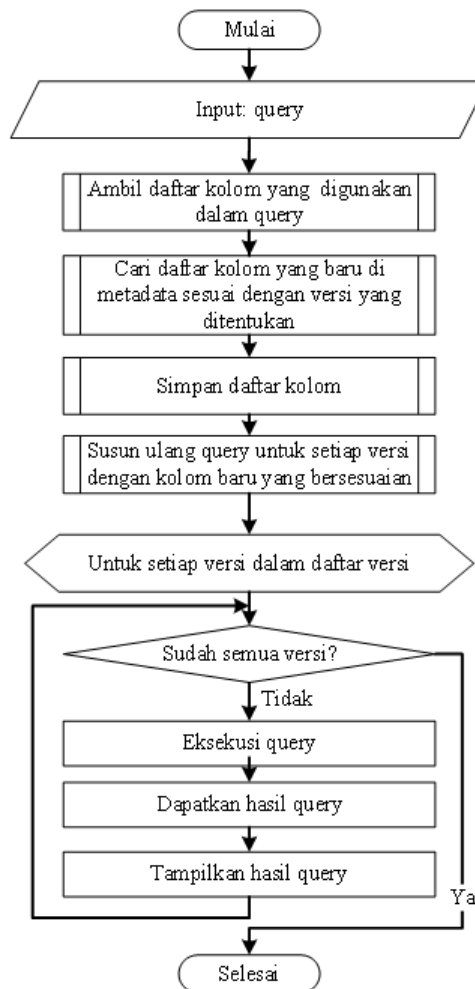


Gambar 8. Alur mendapatkan data dengan adanya perubahan data dan metadata

model yang kami buat terhadap model yang sebelumnya.

9. Hasil Uji Coba

Dari hasil eksperimen, kami melakukan evaluasi terhadap ukuran data yang tersimpan dan waktu untuk



Gambar 9. Algoritma mendapatkan data dari Maghfiroh & Baskara (2017)

menjalankan query dari iterasi yang dilakukan dengan algoritma yang diusulkan.

Berdasarkan implementasi data yang ada dengan model yang sebelumnya dan model yang diusulkan. Ukuran data dari kedua model tersebut tidak jauh berbeda, bahkan hampir sama. Perbandingan ukuran data ditunjukkan pada gambar 12.

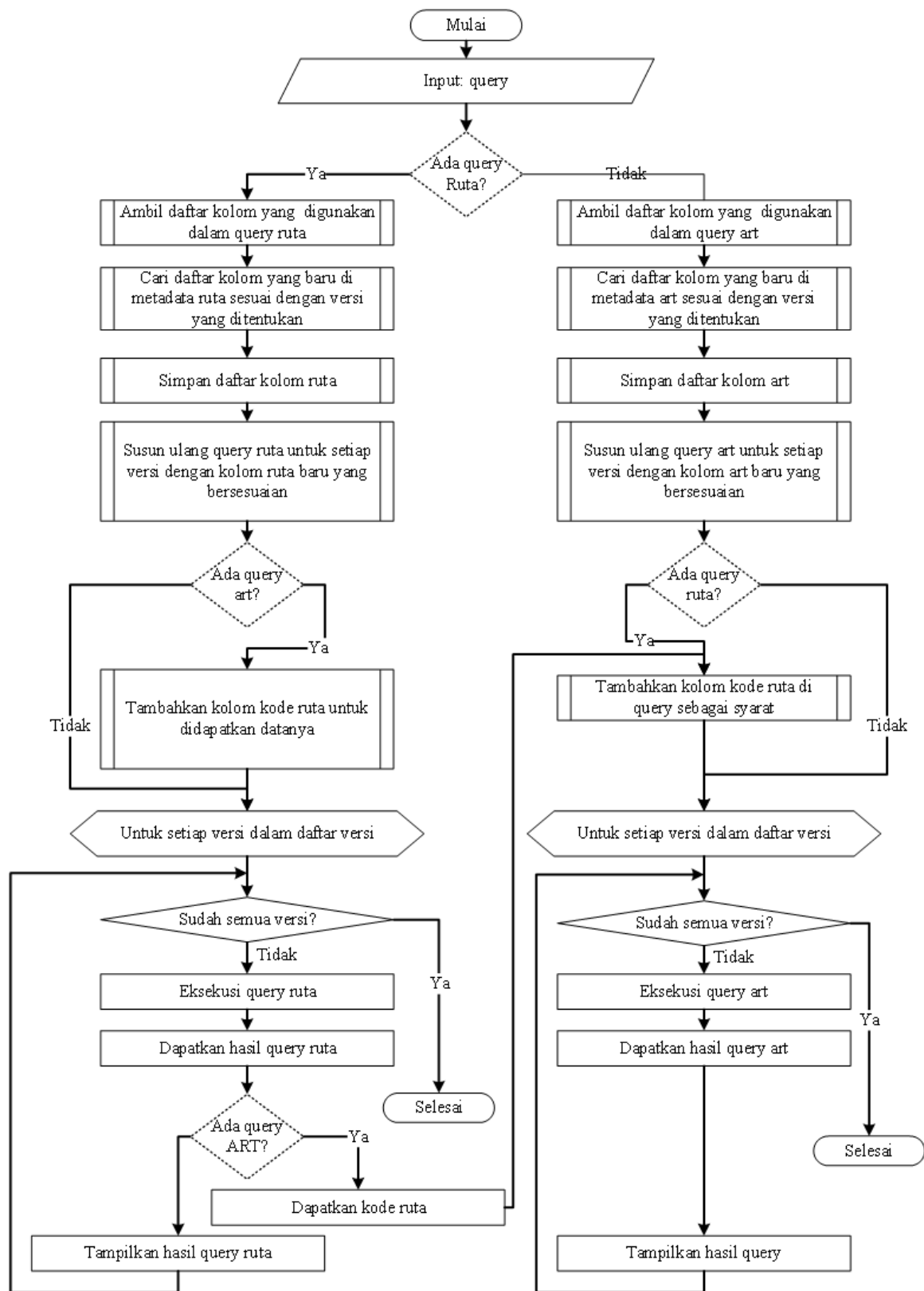
Dengan 11 kasus, baik dilihat secara keseluruhan maupun satu per satu kasus, waktu untuk query dengan model dan algoritma usulan cenderung sedikit lebih tinggi dibandingkan dengan model dan algoritma sebelumnya. Hal ini dapat dimungkinkan karena model dan algoritma yang diusulkan lebih kompleks dibandingkan dengan model dan algoritma usulan. Model dan algoritma yang lebih kompleks ini juga diakibatkan karena adanya kebutuhan untuk mengakomodasi model survei yang berbeda-beda, dibandingkan dengan model dan algoritma

sebelumnya yang hanya mengakomodasi satu model survei saja. Hasil secara umum dapat dilihat pada gambar 13 dan hasil per kasus dapat dilihat pada gambar 14 - 24

Berdasarkan hasil ujicoba, meskipun ukuran data yang hampir sama dan waktu query yang cenderung sedikit lebih tinggi, namun model dan algoritma yang diusulkan dianggap lebih baik karena dapat menangani perubahan metadata dan data dari model survei yang berbeda-beda, dibandingkan dengan model dan algoritma sebelumnya yang hanya bisa menangani satu model survei.

KESIMPULAN DAN SARAN

Untuk menangani perubahan metadata dan data yang ada, kami merancang model DW dengan NoSQL berbasis dokumen. Model yang kami rancang terdiri dari 3 Bagian, Metadata



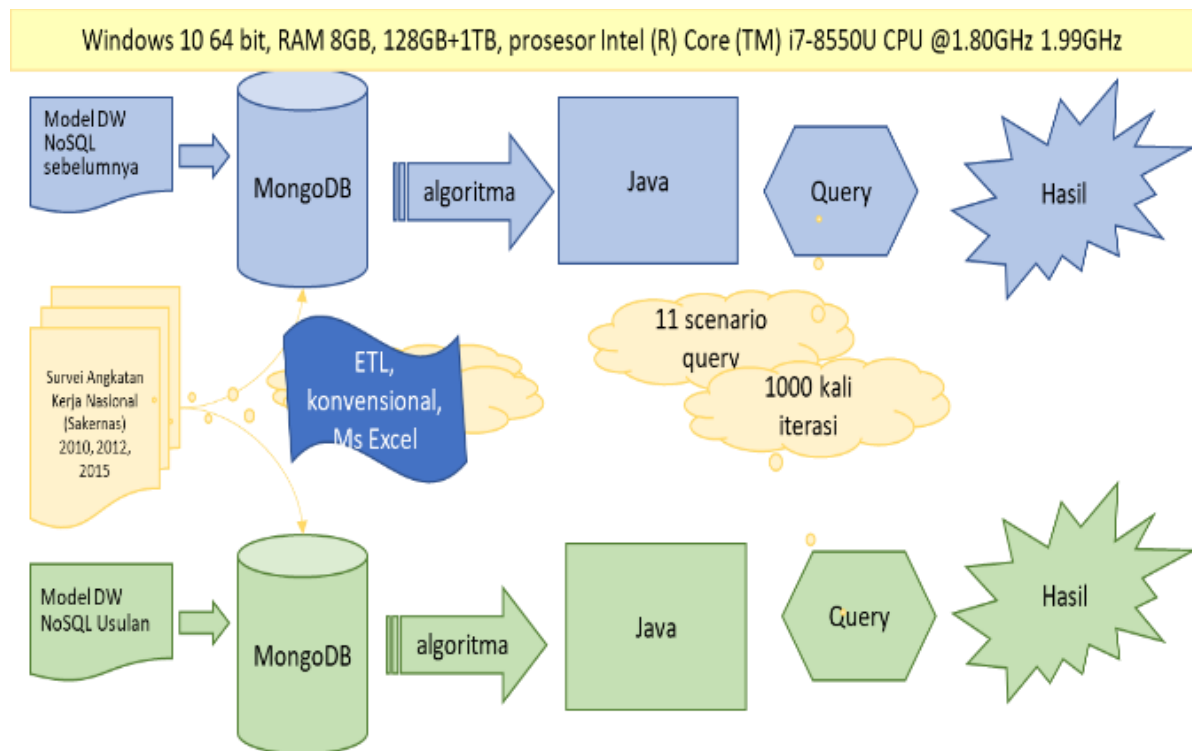
Gambar 10. Algoritma usulan untuk mendapatkan data

Survei, Data dan Metadata Rumah Tangga serta Data dan Metadata Anggota Rumah Tangga

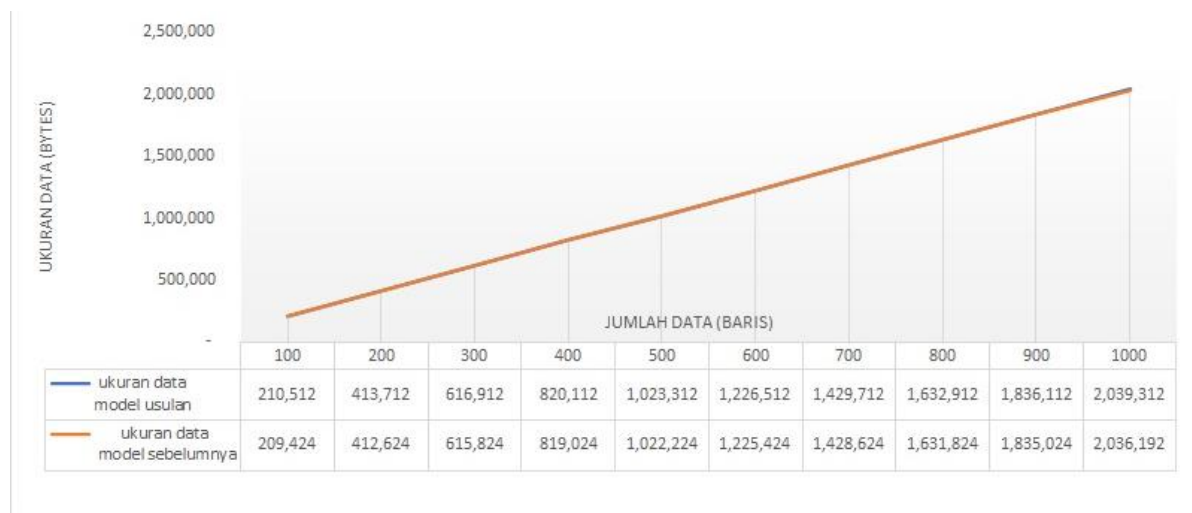
Untuk mendapatkan data yang ada dengan perubahan metadata yang ada, kami merancang algoritma. Input yang

digunakan adalah query yang mengambil data anggota rumah tangga (art), rumah tangga (ruta) atau keduanya.

Ukuran data dari kedua model tersebut tidak jauh berbeda bahkan hampir sama dan waktu untuk query dengan model



Gambar 11. Skenario Pengujian



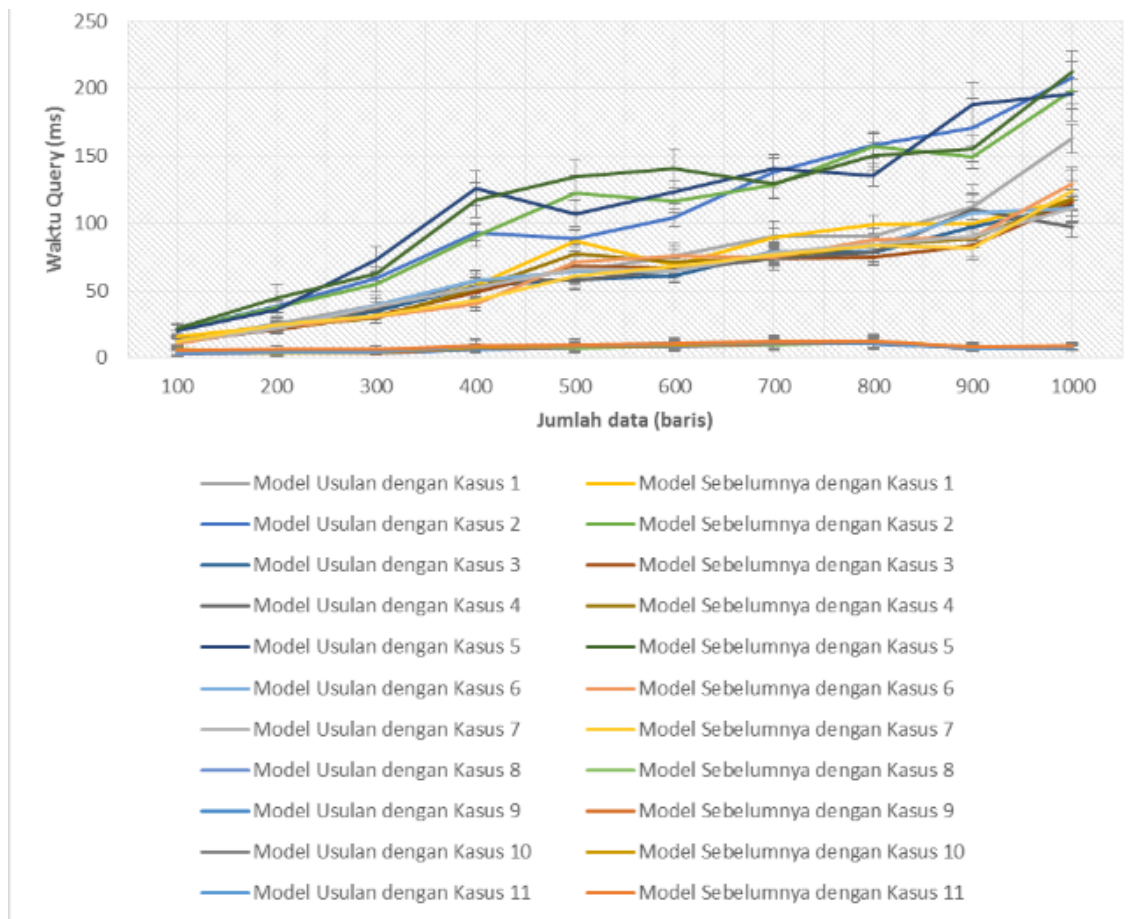
Gambar 12. Perbandingan Ukuran Data dari Implementasi dengan Model Usulan dan Model Sebelumnya

dan algoritma usulan cenderung sedikit lebih tinggi dibandingkan dengan model dan algoritma sebelumnya.

Meskipun model yang diusulkan sudah dapat mengakomodasi kebutuhan, namun waktu yang dibutuhkan untuk menjalankan query masih sedikit lebih tinggi. Hal ini dikarenakan algoritma yang digunakan memang lebih kompleks dibandingkan dengan algoritma yang digunakan dalam penelitian sebelumnya. Oleh karena itu, untuk penelitian selanjutnya kami menyarankan untuk

melakukan penelitian lebih lanjut untuk merancang algoritma yang lebih sederhana sehingga dapat menjalankan query dengan lebih cepat.

Selain itu, penelitian yang kami lakukan baru terbatas pada survei berbasis rumah tangga. Untuk penelitian selanjutnya, dapat dilakukan penelitian untuk perancangan model DW dinamis untuk model survei dengan basis yang lain atau bahkan model yang dapat digunakan semua model survei tanpa ada batasan.

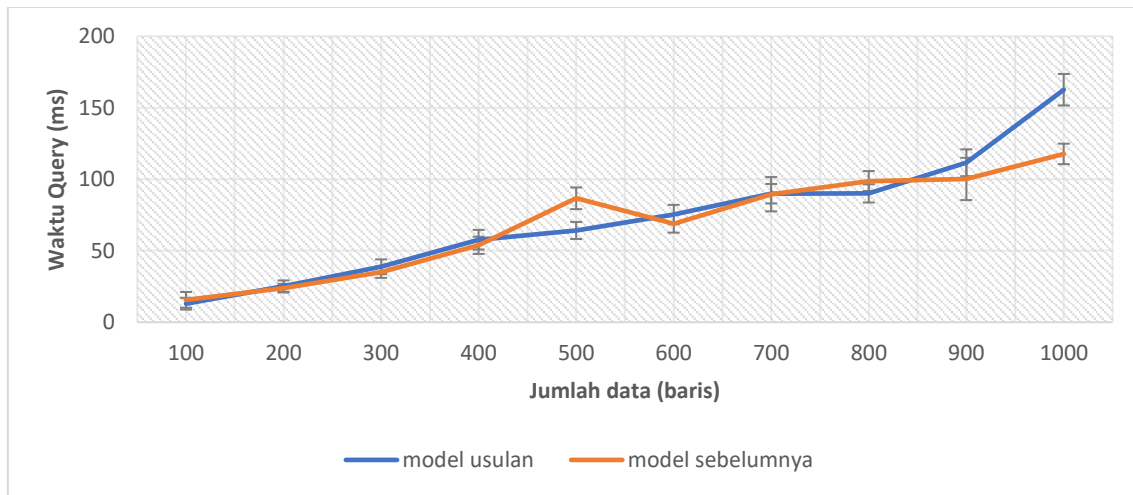


Gambar 13. Waktu Query

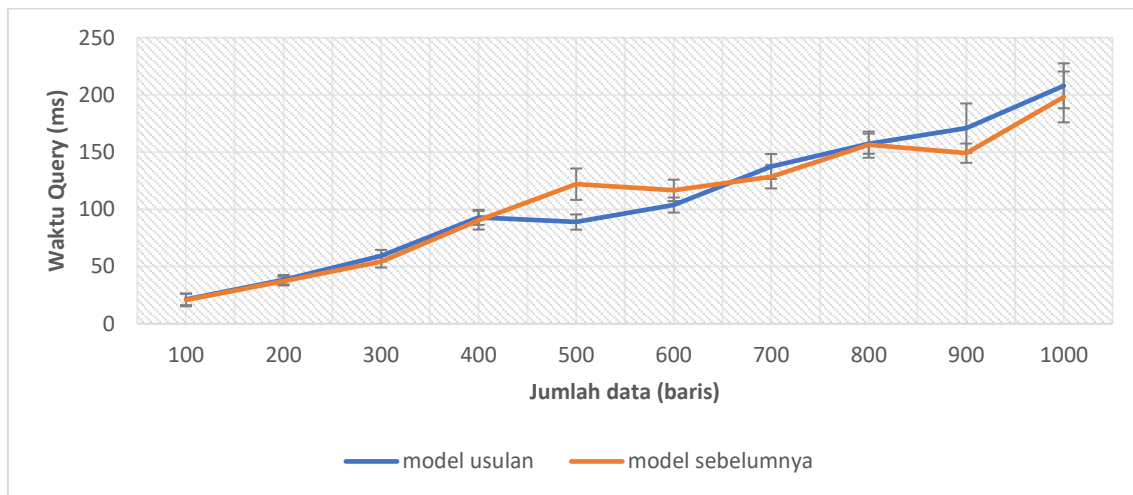
DAFTAR PUSTAKA

- B. Bećbel, Z. Królikowski, and R. Wrembel. 2006. Formal approach to modelling a multiversion data warehouse. in *Bulletin of the Polish Academy of Sciences, Technical Sciences*, vol. 54, pp. 51–62.
- M. Body et al. 2002. A Multidimensional and Multiversion Structure for OLAP Applications, in *Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP*, ser. DOLAP '02. New York, NY, USA: ACM, pp. 1–6.
- L.R. Maghfiroh and I.G.B. Baskara 2017. Survei Data and Metadata Modelling using Document Oriented NoSQL. *The International Conference on Data and Information Science ICoDis:2017*.
- R. Kimball and M. Ross 2013 *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Indianapolis: Wiley.
- R. Wrembel and B. Bećbel. 2007. Metadata Management in a Multiversion Data Warehouse. in *Journal on Data Semantics VIII*, ser. Lecture Notes in Computer Science, S. Spaccapietra et al., Eds. Springer Berlin Heidelberg, no. 4380, pp. 118–157.
- R. Wrembel and T. Morzy. 2006. Managing and Querying Versions of Multiversion Data Warehouse. in *Advances in Database Technology EDBT 2006*, ser. Lecture Notes in Computer Science, Y. Ioannidis et al., Eds. Springer Berlin Heidelberg, no. 3896, pp. 1121–1124.
- M. Chevalier et al. 2016. Document-oriented Models for Data Warehouses - NoSQL Document-oriented for Data Warehouses. in *Proceedings of the 18th International Conference on Enterprise Information Systems*. Rome, Italy: SCITEPRESS - Science and Technology Publications, pp. 142–149.

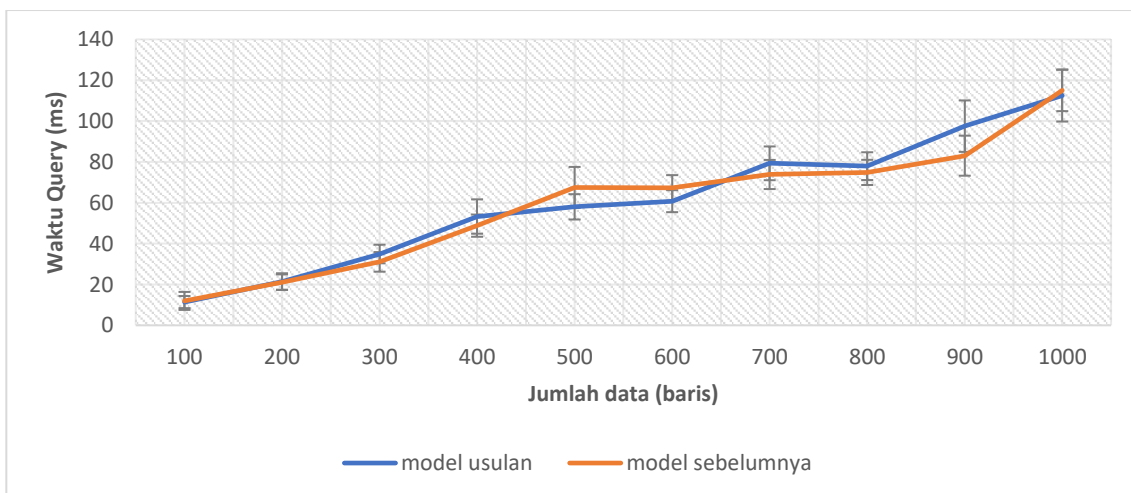
- . 2015. Implementing Multidimensional Data Warehouses into NoSQL. in *Proceedings of the 17th International Conference on Enterprise Information Systems*. Barcelona, Spain: SCITEPRESS - Science and Technology Publications, pp. 172–183.
- . 2015. Implementation of Multidimensional Databases with DocumentOriented NoSQL. in *Big Data Analytics and Knowledge Discovery*, ser. *Lecture Notes in Computer Science*, S. Madria and T. Hara, Eds. Springer International Publishing, no. 9263, pp. 379–390.
- . 2015. Implementation of Multidimensional Databases in ColumnOriented NoSQL Systems. in *Advances in Databases and Information Systems*, ser. *Lecture Notes in Computer Science*, M. Tadeusz, P. Valduriez, and L. Bellatreche, Eds. Springer International Publishing, pp. 79–91.
- . 2015. How Can We Implement a Multidimensional Data Warehouse Using NoSQL?. in *Enterprise Information Systems*, ser. *Lecture Notes in Business Information Processing*, S. Hammoudi et al., Eds. Springer International Publishing, no. 241, pp. 108–130.
- M. O. Fitri. 2013. Trend Penggunaan Nosql Untuk Basis Data Non Relasional. *Fakultas Sains dan Teknologi UIN ALauddin*, vol. 7, No 1 2013: Januari.
- Z. Parker, S. Poe, and S. V. Vrbsky. 2013. Comparing NoSQL MongoDB to an SQL DB. in *Proceedings of the 51st ACM Southeast Conference*, ser. *ACMSE '13*. New York, NY, USA: ACM, pp. 5:1–5:6.



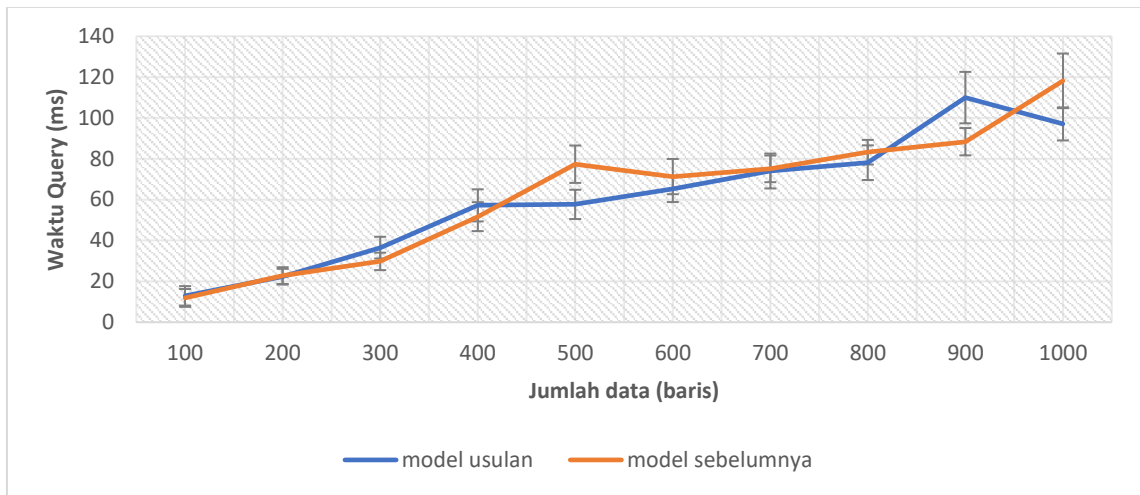
Gambar 14. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 1



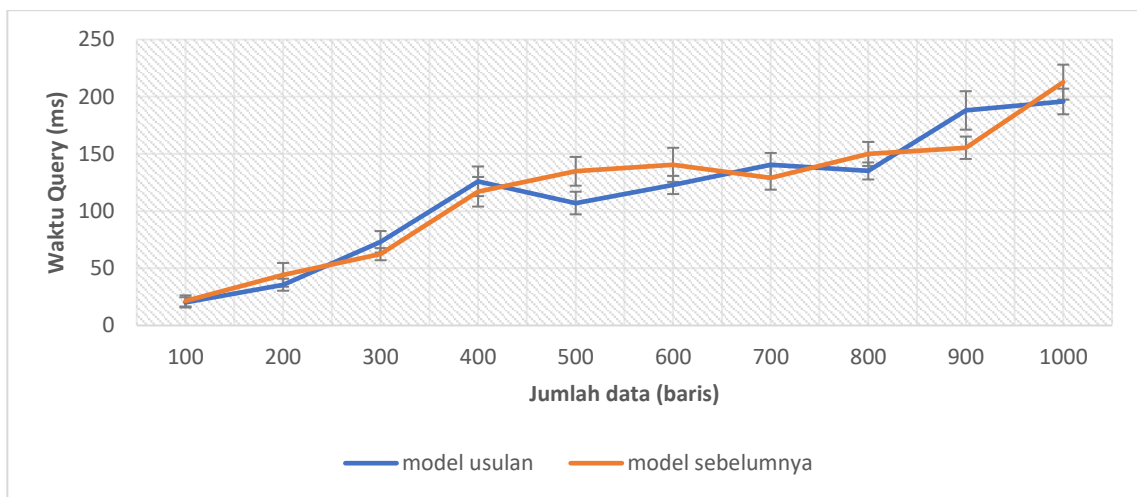
Gambar 15. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 2



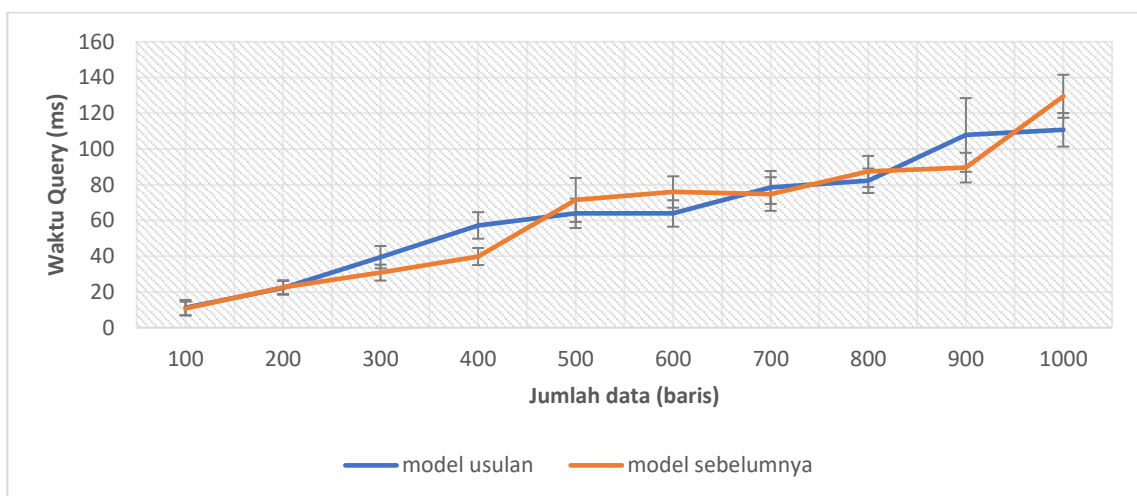
Gambar 16. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 3



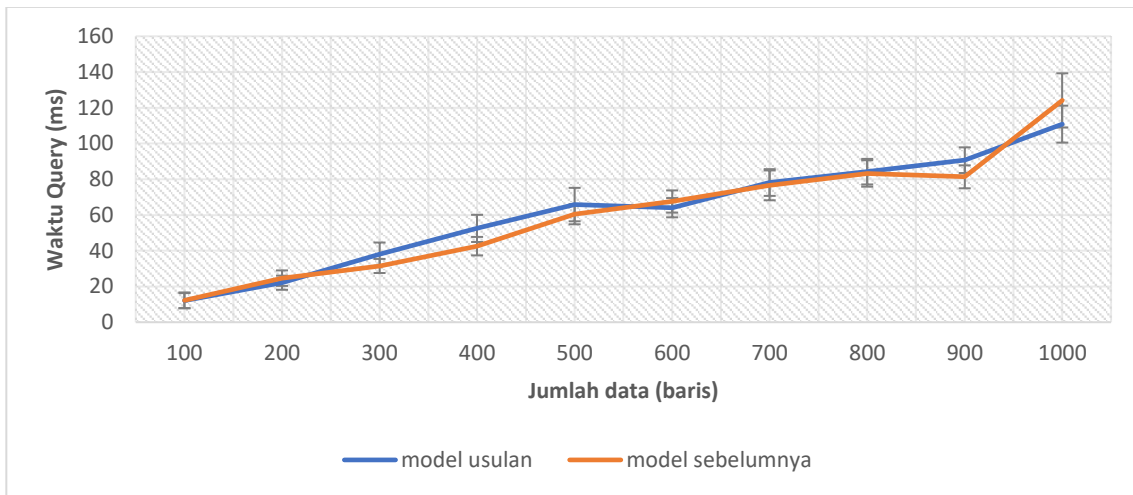
Gambar 17. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 4



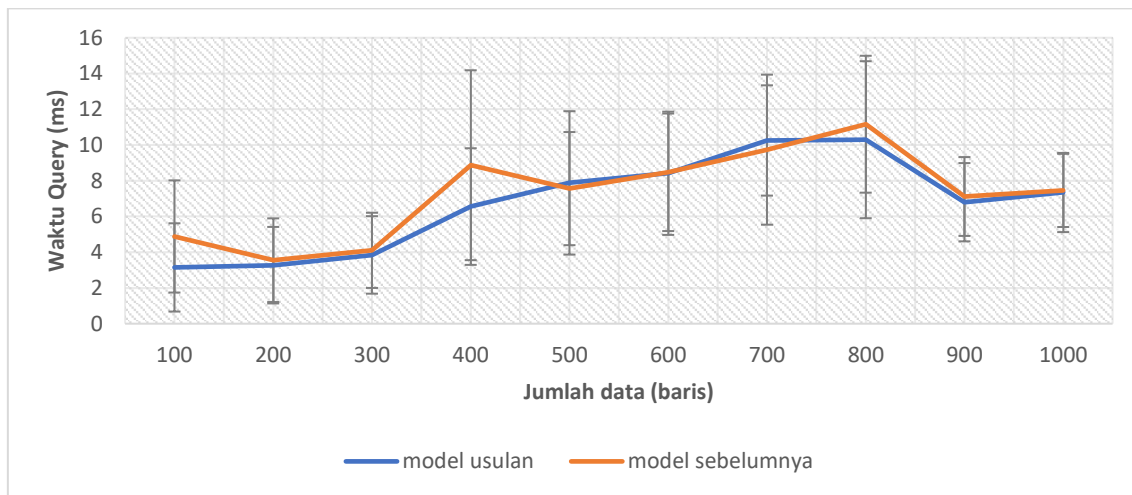
Gambar 18. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 5



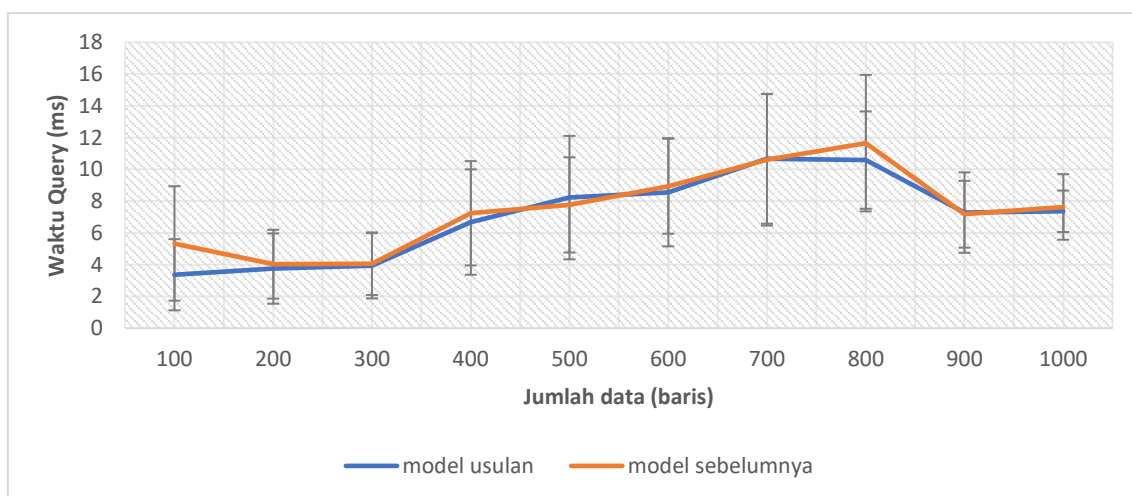
Gambar 19. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 6



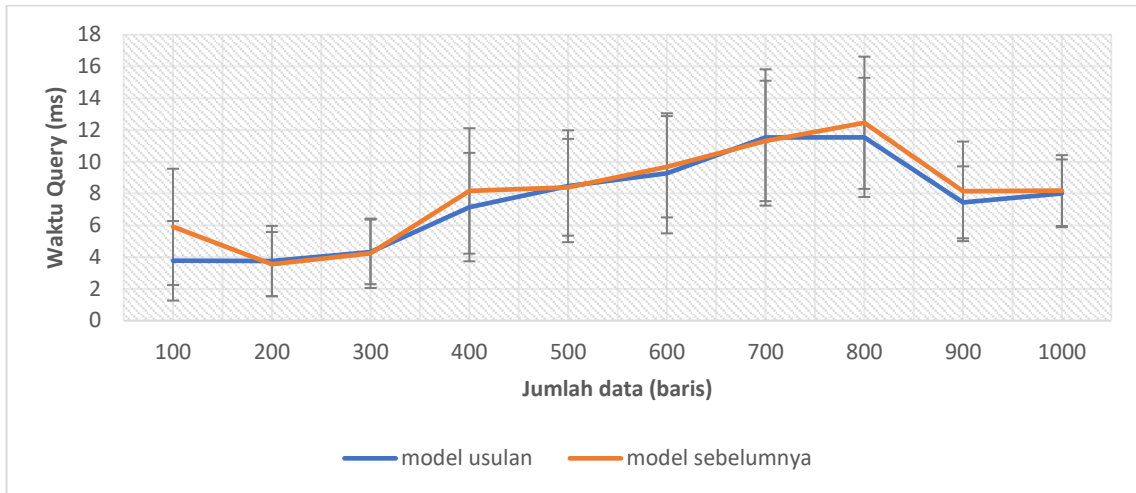
Gambar 20. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 7



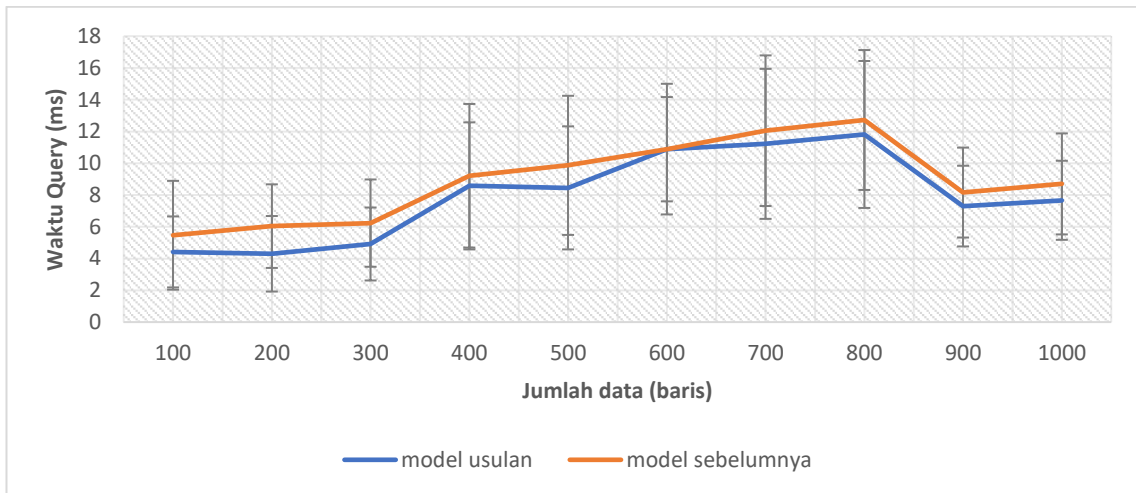
Gambar 21. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 8



Gambar 22. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 9



Gambar 23. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 10



Gambar 24. Perbandingan Waktu Query Model Usulan dan Model Sebelumnya dengan Kasus 11

Petunjuk Penulisan

JURNAL APLIKASI STATISTIKA & KOMPUTASI STATISTIK

Naskah dikirim dalam bentuk *softcopy* ke alamat email pppm@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 1,5, ukuran kertas A4 dan margin 2 cm untuk semua sisi, serta 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*.